Why Did So Many Subprime Borrowers Default During the Crisis: Loose Credit or Plummeting Prices?[*]

Christopher Palmer[†]

University of California at Berkeley

November 18, 2014

Abstract

The foreclosure rate of subprime mortgages increased markedly across 2003–2007 borrower cohorts—subprime mortgages originated in 2006–2007 were roughly three times more likely to default within three years of origination than mortgages originated in 2003–2004. Many have argued that this surge in subprime defaults represents a deterioration in subprime lending standards over time. I quantify the importance of an alternative hypothesis: later cohorts defaulted at higher rates in large part because house price declines left them more likely to have negative equity. Using loan-level data, I find that changing borrower and loan characteristics explain approximately 30% of the difference in cohort default rates, with almost of all of the remaining heterogeneity across cohorts attributable to the price cycle. To account for the endogeneity of prices, I employ a nonlinear instrumental-variables approach that instruments for house price changes with long-run regional variation in house-price cyclicality. Control function results confirm that the relationship between price declines and defaults is causal and explains the majority of the disparity in cohort performance. I conclude that if 2006 borrowers had faced the same prices the average 2003 borrower did, their annual default rate would have dropped from 12% to 5.6%.

Keywords: Mortgage Finance, Subprime Lending, Foreclosure Crisis, Negative Equity

JEL Classification: G01, G21, R31, R38

---

1 Introduction

Subprime residential mortgage loans were ground zero in the Great Recession, comprising over 50% of all 2006–2008 foreclosures despite the fact that only 13% of existing residential mortgages were subprime at the time.[1] The subprime default rate—the number of new subprime foreclosure starts as a fraction of outstanding subprime mortgages—tripled from under 6% in 2005 to 17% in 2009. By 2013, more than one in five subprime loans originated since 1995 had defaulted. While subprime borrowers by definition have been ex-ante judged as having greater default risk than non-subprime mortgages, many have pointed to the disproportionate growth in the share of defaults by subprime borrowers as evidence that the expansion in subprime lending was a major contributing cause to the housing crash of 2007–2009.

Why did the performance of subprime loans decline so sharply? A focal point of the discussion has been the stylized fact that subprime mortgages originated in 2005–2007 performed significantly worse than subprime mortgages originated in 2003–2004.[2] This is visible in the top panel of Figure 1, which uses data from subprime private-label mortgage-backed securities to show this pattern for 2003–2007 borrower cohorts.[3] Each line shows the fraction of borrowers in the indicated cohort that defaulted within a given number of months from origination.[4] The pronounced pattern is that the speed and frequency of default are higher for later cohorts—within any number of months since origination, more recent cohorts have defaulted at a higher rate (with the exception of the 2007 cohort in later years). For example, within two years of origination, approximately 20% of subprime mortgages originated in 2006–2007 had defaulted, in contrast with approximately 5% of 2003-vintage mortgages.

A popular explanation for the heterogeneity in cohort-level default rates over time is that loosen-

---

[1]Statistics derived from the Mortgage Bankers Association National Delinquency Survey. There is no standardized definition of a subprime mortgage, although the term always means a loan deemed to have elevated default risk. Popular classification methods include mortgages originated to borrowers with a credit score below certain thresholds, mortgages with an interest rate that exceeds the comparable Treasury Bill rate by 3%, certain mortgage product types, mortgages made by lenders who self-identify as making predominantly subprime mortgages, and mortgages serviced by firms that specialize in servicing subprime mortgages. For the purposes of this paper, subprime mortgages are defined as those in private-label mortgage-backed securities marketed as subprime, as in Mayer et al. (2009). For an estimate of the effects of foreclosures on the real economy, see Mian et al. (2011).

[2]See JEC (2007), Krugman (2007b), Gerardi et al. (2008), Haughwout et al. (2008), Mayer et al. (2009), Demyanyk and Van Hemert (2010), and Bhardwaj and Sengupta (2011) for examples of contrasting earlier and later borrower cohorts.

[3]This data will be discussed at length in Section 3. The analysis stops in 2007 because by 2008 the subprime securitized market was virtually nonexistent—the number of subprime loans originated in 2008 in the data fell by 99% from the number of 2007 originations.

[4]Following Sherlund (2008) and Mayer et al. (2009), I measure the point in time when a mortgage has defaulted as the first time that its delinquency status is marked as in foreclosure or real-estate owned provided it ultimately terminated without being paid off in full.

ing lending standards led to a change in the composition of subprime borrowers, potentially on both observable and unobservable dimensions (e.g. JEC, 2007 and COP, 2009). Others (e.g. Krugman, 2007a) blame an increase in the popularity of exotic mortgage products (for example, so-called balloon mortgages, which do not fully amortize over the mortgage term, leaving a substantial amount of principal due at maturity). The observed heterogeneity in cohort-level outcomes seen in Figure 1 could be generated by a decrease in the ex-ante creditworthiness of subprime borrowers over time or if the characteristics of originated mortgages became riskier. A third possibility is that price declines in the housing market—national prices declined by 37% between 2005–2009—differentially affected later cohorts, who had accumulated less equity when property values began to plummet. Being underwater—owing more on an asset than its current market value—could be an important friction in credit markets leading to a higher likelihood of default. Borrowers during a period of high price appreciation who have insufficient cash flow to make their mortgage payments can sell their homes or use their equity to refinance into a mortgage with a lower monthly payment. By contrast, if underwater homeowners cannot afford their mortgage payments, their alternatives are limited—lenders are often unwilling to refinance underwater mortgages or allow short sales (where the purchase price is insufficient to cover liens against the property).[5] The pattern of cohort default hazards could therefore come from four sources: price declines, changes in observable borrower characteristics, changes in unobservable borrower characteristics, and changes in mortgage product characteristics.

In this paper, I investigate the relative importance of each of these potential causes of declining cohort outcomes to understand what caused the increase in subprime defaults during the Great Recession. The counterfactual question I ask is whether 2003 borrowers (the best performing cohort) would have defaulted more like 2006 borrowers did if instead they had taken out mortgages in 2006 (when the worst performing cohort did). If so, then it is less plausible that deteriorating lending standards and risky mortgage products were a key driver of the surge in subprime defaults. On the other hand, if 2003 borrowers would have defaulted at a lower rate even after adjusting for observable borrower characteristics, loan characteristics, and market conditions, this would imply important differences in unobserved borrower quality across cohorts.[6]

---

[5]Underwater homeowners may also default strategically to discharge their mortgage debt if they deem the option value of holding onto their property to be low. Bhutta et al. (2010) find that the property value of the median strategically defaulting borrower is less than half of the outstanding principal balance. Genesove and Mayer (1997) show that, all else equal, highly levered sellers also set higher reservation prices.

[6]Note that even absent a significant change in cohort quality, subprime lending could have had a sizable effect on the economy through feedback between subprime defaults and price declines. Isolating the causal effect of prices on defaults is thus an input into the larger question of what was the net impact of subprime lending on the housing

2

To answer these questions, I estimate semiparametric hazard models of default using a panel of subprime loans that combines rich borrower and loan characteristics with monthly updates on loan balances, property values, delinquency statuses, and local price changes. I find that differential exposure to price declines explains 60% of the heterogeneity in cohort default rates. I also estimate that the product characteristics of subprime mortgages—but not the borrower characteristics—play an important role, accounting for 30% of the rise in defaults across cohorts. Conditioning on all three channels (price changes and loan and borrower characteristics) explains almost the entire change in cohort-level default rates, suggesting that the effect of any decline in unobserved borrower quality (e.g. from a deterioration in the accuracy of mortgage applications) was negligible. Returning to the counterfactual question posed above, my results imply that if 2006 borrowers had faced the prices that the average 2003 borrower did (i.e. at the same number of months since origination), 2006 borrowers would have had an annual default rate of 5.6% instead of 12%.[7] Furthermore, I find that if 2003 and 2006 borrowers had taken out identical mortgage products in addition to having faced the same prices, they would have defaulted at nearly identical rates.

House prices are an equilibrium outcome dependent on factors related to default risk. Whatever their source, price declines may have a causal effect on defaults. However, the potential for price changes and defaults to be caused by a third factor may lead to estimating a spurious relationship between price changes and defaults. In other words, some of the sources of price shocks may also have direct effects on the unobserved quality of borrowers and hence on defaults. A prominent hypothesis is that subprime penetration itself may subsequently have caused price declines and defaults, as suggested by Mayer and Sinai (2007), Mian and Sufi (2009), and Pavlov and Wachter (2009). In short, a credit expansion could amplify the price cycle, initially increasing prices from the positive demand shock as the pool of potential buyers grows. However, if the credit expansion involves a decrease in average borrower quality, this process will eventually lead to an increase in defaults, accelerating price declines. Thus, even though individual borrowers are price takers in the housing market, their unobserved quality may be correlated with the magnitude of price declines, resulting in biased estimates of the causal effect of prices on default risk.

The possibility of such a process makes it difficult to determine whether price changes actually cause defaults or if the defaults that are observed simultaneously with price declines are driven by the same latent factors driving prices and would therefore have occurred even absent any price

---

market.

[7] I measure the annual default rate within five years of origination as 12 times the average fraction of existing loans that default each month.

changes. This impediment to estimating the causal effect of prices on defaults is also a challenge in estimating whether there were quality differences across cohorts. If unobserved quality differences affect both defaults and price declines, not taking into account the endogeneity of prices could lead to an underestimate of heterogeneity in cohort quality and an overestimate of the role of prices in affecting defaults.

To isolate the portion of cohort default rates driven by price changes from changes in unobserved borrower quality which also affect prices, I exploit plausibly exogenous long-run variation in metropolitan-area house-price cyclicality. As observed by Sinai (2012), there is persistence in the amplitude of house-price cycles—cities with strong price cycles in the 1980s were more likely to have strong cycles in the 2000s. I use this historical variation in house-price volatility to construct counterfactual price indices, which are unrelated to housing market shocks unique to the 2000s price cycle, e.g. because price volatility in the 1980s occurred well before the widespread adoption of subprime mortgages. Indeed, I show below that my instrument does not predict differential subprime expansion. Nevertheless, I also present evidence that areas that have cyclical housing markets also have cyclical labor markets. To address the possibility that price results could be explained by local labor shocks (an increase in the unemployment rate may cause defaults and depress prices), I verify that my results are robust to controlling for local unemployment rates.[8]

To my knowledge, this paper is the first to instrument for prices to address the joint endogeneity of prices and defaults in estimating the causal effect of price changes on defaults. While many researchers have looked at the relationship between house price appreciation and defaults, none of them have addressed the possible endogeneity of house price changes. For example, the common practice of imputing changes in property values using a metropolitan area home price index, although free from property-specific price shocks, does not address the concern that price changes at the metropolitan area level are themselves the outcome of demand and supply shocks that are likely correlated with unobserved borrower quality. Using a nonlinear instrumental variables approach to account for the endogeneity of covariates in a hazard model setting, I confirm that prices are endogenous, they are an important determinant of default, and they explain over half of the cohort pattern in default rates.

Figure 2 illustrates the differential effect that declining home prices had on origination cohorts by plotting the median mark-to-market combined loan-to-value ratio (CLTV) of each cohort of bor-

---

[8]Mayer (2010), Mian (2010), and Mian and Sufi (2012) argue that price declines first caused unemployment in the recent recession.

rowers over time.[9] The beginning of each line shows the median CLTV at origination for mortgages taken out in January of that cohort's birth year. Thereafter, each line shows the median CLTV of all existing mortgages in the indicated origination cohort.[10] Each cohort's median CLTV began rising in 2007 as prices declined nationwide. However, there are two main differences between early and late cohorts. First, origination CLTVs increased over time—the median 2007 CLTV was 10 percentage points higher than the median 2003 CLTV, lending credence to the argument that underwriting standards deteriorated. Second, earlier cohorts' median CLTVs declined from origination until 2007 as prices rose (increasing the CLTV denominator) and as borrowers made their mortgage payments, reducing their indebtedness (the CLTV numerator), with the former effect dominating because of the low amount of principal paid off early in the mortgage amortization schedule. By contrast, later cohorts had not accumulated any appreciation or paid down any principal, as prices fell almost immediately after their origination dates. By early 2008, more than one-half of borrowers in both the 2006 and 2007 cohorts were underwater, and by early 2009, more than one-half of the 2005 cohort was underwater. Using variation in price changes across cities and cohorts and controlling for CLTV at origination, the empirical specifications below allow me to identify the causal effect of prices on defaults, differentiating between differences in negative equity prevalence across cohorts explained by high CLTVs at origination (a measure of cohort quality) and less opportunity to accumulate equity before price declines begin.

Suggestive evidence that the prevalence of negative equity affected economic outcomes is the bottom panel of Figure 1, which shows the cumulative prepayment probability by cohort—the fraction of each cohort's mortgages that had been paid off within the given number of months since origination.[11] The pattern across cohorts is exactly reversed from the cohort heterogeneity in default rates depicted in the top panel—more recent borrowers prepaid their mortgages much less frequently and at slower rates than borrowers from 2003–2005. Given the evidence that later cohorts were more likely to be underwater, the contrast between the cohort-level trends in defaults and prepayments is consistent with the notion that underwater borrowers in distress default and

---

[9]The combined loan-to-value ratio (CLTV) of a mortgage is the sum of all outstanding principal balances secured by a given property divided by the value of that property. The data used in Figure 2 estimate market values from CoreLogic's Automated Valuation Model, see Section 3 for more details.

[10]Having a high CLTV at origination (equivalent to having a small down payment) is highly correlated with default risk and is routinely factored into the interest rates charged by lenders.

[11]Note that prepayment has a specific meaning in mortgage finance. As the issuer of a callable bond, a mortgage borrower has the prerogative to pay back the debt's principal balance at any time, releasing them of further obligation to the lender. In practice, this is done through refinancing or selling the home and using the proceeds to pay back the lender. See Mayer et al. (2010) for a discussion of mortgage prepayment penalties, an increasingly common feature of subprime mortgages.

above-water borrowers in distress prepay.[12]

The differing experiences of the Pittsburgh and Minneapolis metropolitan areas serve as a motivating case study for the conceptual experiment in which this paper engages using geographic variation in prices. Although they had similar subprime market shares, these two cities had very different price cycles—Pittsburgh did not have much of a cycle, whereas Minneapolis home prices had a price cycle similar to the national average (see top panel of Figure 3).[13] As a consequence, the bottom panel of Figure 3 shows that the fraction of Pittsburgh subprime homeowners that were underwater stayed roughly constant at 30%, while the fraction of Minneapolis subprime homeowners who were underwater increased from under 20% before 2006 to over 35% by the middle of 2008.

The contrast between Pittsburgh and Minneapolis also extends to default rates. The top panel of Figure 4 shows that Pittsburgh cohorts defaulted at very similar rates, with later cohorts actually defaulting *less* than earlier cohorts by the end of the period. By comparison, in Minneapolis, where prices followed a boom-bust pattern, earlier borrower cohorts defaulted at a much lower rate than later cohorts. The bottom panel of Figure 4 shows that approximately 15% of Minneapolis subprime mortgages originated in 2006–2007 had defaulted within 12 months of origination, whereas only 5% of 2003–2004 mortgages had defaulted within the same time frame. The contrasting pattern across cohorts in Pittsburgh and Minneapolis suggests that the relative lack of a price decline and stable prevalence of negative equity in Pittsburgh may explain why default risk appears constant across Pittsburgh cohorts relative to Minneapolis, where an increasing share of underwater borrowers seems to have led to a rapid increase in default rates.

The strategy of this paper is to generalize the Pittsburgh-Minneapolis comparison to a comprehensive national dataset by including loan-level controls for the changing composition of borrowers in each locale and by isolating exogenous variation in each city's price cycle. Intuitively, I compare cohorts in areas with different price cycles (and thus different predicted availability of sell/refinance options for borrowers) to estimate whether they also had different default patterns after adjusting for observable underwriting characteristics.

There is a broad literature on the determinants of mortgage default.[14] A number of studies have examined the proximate causes of the subprime foreclosure crisis in particular (see Keys et al., 2008, Hubbard and Mayer, 2009, Mian and Sufi, 2009, and Dell'Ariccia et al., 2012). Kau

---

[12]Note that this pattern could also be generated by cohort quality if riskier borrowers prepay less frequently, e.g. if they are less likely to trade-up to a more expensive home.

[13]According to Mayer and Pence (2008), 16% and 17% of mortgages originated in 2005 were subprime in Pittsburgh and Minneapolis, respectively.

[14]For example, Deng et al. (2000), Foote et al. (2008), Pennington-Cross and Ho (2010), and Bhutta et al. (2010).

et al. (2011) find that the market was aware of an ongoing decline in subprime borrower quality. Corbae and Quintin (2013) provide a model demonstrating how a period of relaxed underwriting standards could lead to a mass of mortgages originated to borrowers who would subsequently be extraordinarily sensitive to price declines.

Several papers have tried to quantify the relative contributions of underwriting standards and housing market conditions in the increase in the subprime default rate over time (all treating metropolitan area home price changes as exogenous) and have generally found a residual decrease in cohort quality. Sherlund (2008) concludes that leverage is the strongest predictor of increasing default risk and decreasing prepayment risk among subprime loans. Gerardi et al. (2008) use data through 2007 to ask whether lenders, investors, and rating agencies should have known that price declines would induce widespread defaults. Gerardi et al. (2007) examine the importance of negative equity. Krainer and Laderman (2011) examine the correlation between prepayment and default rates and find that declines in prepayment rates are strongly correlated with increases in default rates, particularly among borrowers with low credit scores. Bajari et al. (2008) estimate a dynamic model of default behavior on subprime mortgage data from 20 metropolitan areas and find evidence supporting both lending standards and price declines as drivers of default.

Other papers analyze differences in default or delinquency across cohorts. Mayer et al. (2009) demonstrate heterogeneity in the early default rates of origination cohorts and examine a series of bivariate correlations over time to document that loosening down payment requirements and declining home prices are both highly correlated with increases in early defaults. Bhardwaj and Sengupta (2012) estimate the cohort effects in default and prepayment hazards to be inversely related—later cohorts defaulted relatively more and prepaid relatively less. The most closely related study to this one is Demyanyk and Van Hemert (2011), which explicitly considers vintage effects in borrower quality and finds that prices played a much more important role than observable lending standards in explaining early delinquencies. Using data ending in 2008, they conclude that the bulk of the deterioration in cohort quality was due to unobservables, suggesting that the lending boom coincided with adverse selection among borrowers.

In summary, existing work has focused on whether changing underwriting standards (originated loan characteristics) explain changing default rates or whether prevailing market conditions such as negative equity were acute in areas where many borrowers are defaulting. They all find that a much larger portion of the deterioration in cohort quality is explained by home prices than ex-ante borrower characteristics. In contrast to these papers, with the benefit of several more years of data

on the 2003–2007 subprime borrower cohorts and an instrumental-variables strategy, I am able to make causal inferences about the effect of price changes on default rates.

The paper proceeds as follows. Section 2 discusses the empirical strategy. I describe the data and compare the observable characteristics of borrower cohorts in Section 3. Identification concerns in the context of a hazard model are detailed in Section 4, along with a description of the estimator. After presenting initial descriptive estimates of the determinants of default that drive the cohort pattern, Section 5 presents the instrumental variables strategy and my main results, and Section 6 explores the economic mechanisms through which price declines affect default rates. Using my preferred empirical specification, I estimate cohort-level default rates under several counterfactual scenarios in Section 7. In Section 8, I conclude by summarizing my main findings and briefly discussing policy implications.

2 Empirical Strategy

Many factors determine default risk. Underwriting standards and market conditions, each predictive of future idiosyncratic income shocks and changes in prepayment opportunities, interact to generate defaults. Loose underwriting standards increase default rates because equally sized negative income shocks are more likely to prevent borrowers with high debt-to-income ratios from making mortgage payments and because borrowers with riskier income are more likely to have a negative shock that prevents them from making their mortgage payments. After a period of sustained price growth, younger loans are also relatively more sensitive to price declines because they have not accumulated as much equity and are thus more apt to be underwater and constrained in their ability to sell or refinance their mortgage. If an equal share of each cohort has an income shock that prohibits them from paying back their mortgage, cohorts with positive equity will simply sell their homes or refinance into mortgages with better terms. Later cohorts, on the other hand, have no such option and will default.

The objective of the hazard models presented below is to examine the relative importance of each of these factors by comparing loans with differing underwriting characteristics and in areas with differing price cycles to estimate how much of the heterogeneity in cohort default rates is explainable by each factor. Comparing observationally similar loans (i.e. by controlling for under-writing standards and loan age with a flexible baseline hazard specification) within a geography that were originated at different times allows me to take advantage of temporal variation in house

8

prices within a geographic region. Likewise, comparing observationally similar loans taken out at the same time but in different cities utilizes spatial variation in house prices. To account for the endogeneity of the house price series of each geographic area, I estimate counterfactual price series by mapping each area's 1980–1995 house price volatility onto the most recent price cycle, as discussed in detail in Section 4 below. This setup allows me to decompose observed cohort heterogeneity into its driving factors by successively introducing additional controls that explain away the differences in cohort default rates.

2.1 Hazard Model Specification

I specify the origination-until-default duration as a proportional hazard model with time-varying covariates. Although the data are grouped into monthly observations, the proportional hazards functional form allows estimation of a continuous-time hazard model using discrete data (Prentice and Gloeckler, 1978 and Allison, 1982). Let the latent time-to-default random variable be denoted τ, and let the instantaneous probability (i.e. in continuous-time) of borrower i in cohort c and geography g defaulting at month t given that borrower i has not yet defaulted specified as

$$\lim_{\xi \to 0^+} \frac{\Pr\left(\tau \in (t-\xi, t]\,\middle|\,\tau > t-\xi\right)}{\xi} \quad \equiv \quad \lambda(X_{icg}(t), t) \tag{1}$$

$$= \quad \exp(X'_{icg}(t)\beta)\lambda_0(t) \tag{2}$$

where $\lambda_0(\cdot)$ is the baseline hazard function that depends only on the time since origination t, and $X_{icg}(t)$ is a vector of time-varying covariates that in practice will be measured at discrete monthly intervals. The proportional hazards framework assumes that the conditional default probability depends on the elapsed duration only through a baseline hazard function that is shared by all mortgages. A convenience of this framework is that the coefficient vector β is readily interpretable as measuring the effect of the covariates on the log hazard rate.

Combining a nonparametric baseline hazard function with covariates entering through a parametric linear index function results in a semiparametric model of default. The specification for the covariates is

$$X'_{icg}(t)\beta = \gamma_c + W'_{B,i}\theta_B + W'_{L,i}\theta_L + \mu \cdot \Delta Prices_{icg}(t) + \alpha_g \tag{3}$$

where γ_c and α_g are cohort and geographic fixed effects, respectively; W_B and W_L are vectors of borrower (B) and loan (L) attributes, measured at the time of mortgage origination; and $\Delta Prices_{icg}(t)$

9

is a measure of the change in prices faced by property i at time t.[15] Borrower characteristics include the FICO score (a credit score measuring the quality of the borrower's credit history), debt-to-income (DTI) ratio (calculated using all outstanding debt obligations), an indicator variable for whether the borrower provided full documentation of income during underwriting, and an indicator variable for whether the property was to be occupied as a primary residence. Attributes of the mortgage note include the origination combined loan-to-value ratio (using all open liens on the property for the numerator and the sale price for the denominator), the mortgage interest rate, and indicator variables for adjustable-rate mortgages, cash-out refinance mortgages (when the new mortgage amount exceeds the outstanding principal due on the previous mortgage), mortgages with an interest-only period (when payments do not pay down any principal), balloon mortgages (non-fully amortizing mortgages that require a balloon payment at the end of the term), and mortgages accompanied by additional so-called piggyback mortgages.

The cohort fixed effects γ_c are the parameters of interest. As 2003 is the omitted cohort, the estimated baseline hazard function represents the conditional probability of default for a 2003 mortgage of each given age. The γ_c parameters scale this up or down depending on how cohort c mortgages default over their life-cycle, conditional on X and relative to 2003 mortgages of the same duration. Successively conditioning on geographic fixed effects, borrower characteristics, loan characteristics, and price changes reveals the extent to which each factor explains the systematic variation in default risk across cohorts. The estimated $\hat{\gamma}_c$ without conditioning on any covariates are a measure of the average performance of each cohort. Conditioning on prices, the γ_c are an estimate of the quality of each cohort, where quality is estimated using an ex-post measure (defaults). Conditioning on observable loan and borrower characteristics and prices, the γ_c represents the latent (i.e. unobserved) quality of each cohort. If cohort-level mortgage performance differences were driven by borrower unobservables, or if the explanatory power of the observables declined over time, then this would be captured by the cohort coefficients after controlling for all observables.

3   Data and Descriptive Statistics

In this section I briefly describe the data sources used in my analysis.

---

[15]A natural concern with including fixed effects α_g in a nonlinear panel data model like this is the incidental parameters problem, which arises when the observations per group g is small and the number of groups grows with the sample size such that no progress is made in reducing the variance of the estimated fixed effects. Unlike a panel with fixed effects for each individual, the details of this application suggest this is not a significant worry. The number of observations per geography is already quite large, and as the total number of observations increases, the number of metropolitan areas in the U.S. remains fixed, leading to consistent estimates of α_g.

CoreLogic LoanPerformance (LP) Data. The main data source underlying this paper is the First American CoreLogic LoanPerformance (LP) Asset-Backed Securities database, a loan-level database providing detailed information on mortgages in private-label mortgage-backed securities including static borrower characteristics (DTI, FICO, owner-occupant, etc.), static loan characteristics (LTV, interest rate, purchase mortgage, etc.), and time-varying mortgage attributes updated monthly such as delinquency statuses and outstanding balances.[16] The LP data record monthly loan-level data on most private-label securitized mortgage balances, including an estimated 87% coverage of outstanding subprime securitized balances. Because about 75% of 2001–2007 subprime mortgages were securitized, this results in over 65% coverage of the subprime mortgage market.[17] My estimation sample is formed from a 1% random sample of first-lien subprime mortgages originated in 2003–2007 in the LP database, resulting in a final dataset of over one million loan × month observations.[18]

Table 1 reports descriptive statistics for static (at time of origination) loan-level borrower and mortgage characteristics. On these observable dimensions, it is clear that subprime borrowers comprised a population with high ex-ante default risk.[19] The average subprime borrower in my data had a credit score of 617, slightly above the national 25th percentile FICO score and substantially below the national median score of 720 (Board of Governors of the Federal Reserve System, 2007). Among borrowers who reported their income on their mortgage application, the average back-end debt-to-income ratio, which combines monthly debt payments made to service all open property liens, was almost 40%, well above standard affordable housing thresholds. More than half of the loans in my estimation sample were for cash-out refinances, where the borrower is obtaining the new mortgage for an amount higher than the outstanding balance of the prior mortgage. As of April 2013, when my data end, 24% of the mortgages in my sample have defaulted and 50% have been paid off, leaving 26% of the loans in the data still outstanding.

---

[16]Using LP data is standard in the economics literature for microdata-based analysis of subprime and near-prime loan performance. See Sherlund (2008), Mayer et al. (2009), Demyanyk and Van Hemert (2011), and Krainer and Lederman (2011) for examples. See GAO (2010) for a more complete discussion of the LP database and comparison with other loan-level data sources.

[17]See Mayer and Pence (2008) for a description of the relative representativeness of subprime data sources. Foote et al. (2009) suggest that non-securitized subprime mortgages are less risky than securitized ones.

[18]As mentioned above, for my purposes a subprime loan is one that is in a mortgage-backed security that was marketed at issuance as subprime. I additionally drop mortgages originated for less than $10,000 and non-standard property types such as manufactured housing following Sherlund (2008).

[19]One measure of the elevated default risk inherent to subprime mortgages Gerardi et al. (2007), who find that homeownership experiences begun with a mortgage from a lender on the Department of Housing and Urban Development subprime lender list have a six times greater default hazard than ownership experiences that start with a prime mortgage.

Table 2 presents descriptive statistics by origination cohort. The distribution of many borrower characteristics is stable across cohorts. Average FICO scores, DTI ratios, combined loan-to-value ratios (measured using all concurrent mortgages and the sale price of the home, both at the time of origination), documentation status, and the fraction of loans that were owner-occupied or were taken out as part of a cash-out refinance are roughly constant across cohorts.[20] While there is substantial evidence that, pooling prime, near-prime, and subprime mortgages, borrower characteristics were deteriorating across cohorts (see JEC, 2007), the lack of a noticeable decrease in borrower observables in my data is consistent with observations from Gerardi et al. (2008) and Demyanyk and Van Hemert (2011) who argue that the declines within the population of subprime borrowers were too small to account for the heterogeneity in performance across cohorts.[21] Among mortgage product characteristics, however, there are important differences across cohorts, including a marked increase in prevalence of interest-only loans, mortgages with balloon payments, and mortgages accompanied by additional liens on the property. This finding of relatively stable borrower characteristics and large changes in certain mortgage characteristics is consistent with the findings of Mayer et al. (2009).

Specifications which directly examine the effects of negative equity make use of a novel feature of the LP dataset: contemporaneous combined loan-to-value ratios (CLTVs), which are a measure of the total amount of debt secured against a property relative to its market value. To calculate the CLTV numerator, CoreLogic uses public records filings on additional liens on the property to estimate the total debt secured against the property at origination. For the denominator, CoreLogic has an automated valuation model (popular in the mortgage lending industry) that uses the characteristics of a property combined with recent sales of comparable properties in the area and monthly home price indices to impute a value for each property in each month.

CoreLogic Home Price Index. For regional measures of home prices, I use the CoreLogic monthly Home Price Index (HPI) at the Core Based Statistical Area (CBSA) level.[22] These indices follow the Case-Shiller weighted repeat-sales methodology to construct a measure of quality-adjusted market prices from January 1976 to April 2013. They are available for several property categories—I

---

[20]Note that the at-origination CLTVs reported here use the sale price of the home for its value, whereas the contemporaneous (mark-to-market) CLTVs in Figure 2 use estimated market values. If the divergence between these two measures over time is an important predictor of default, it will affect the magnitude of the estimated cohort main effects, which capture all unobserved factors changing across cohorts.

[21]Still, the nationwide decline in underwriting standards was driven in part by the subprime expansion: Even though the composition of the subprime borrower population was relatively stable over time, subprime borrowers represented a growing share of *overall* mortgage borrowers.

[22]There are 955 Core Based Statistical Areas in the United States, each of which is either a Metropolitan Statistical Area or a Micropolitan Statistical Area (a group of one or more counties with an urban core of 10,000–50,000 residents).

use the single family combined index, which pools all single family structure types (condominiums, detached houses, etc.) and sale types (i.e. does not exclude distressed sales). Each CBSA's time series is normalized to 100 in January 2000.

The CoreLogic indices have distinct advantages over other widely used home price indices. The extensive geographic coverage (over 900 CBSAs) greatly exceeds the Case-Shiller index, which is only available for twenty metropolitan areas and the FHFA indices, which cover roughly 300 metropolitan areas. Unlike the FHFA home price series, CoreLogic HPIs are available for all residential property types, not just conforming loans purchased by the GSEs. Finally, its historical coverage—dating back to 1976—predates the availability of deed-based data sources such as DataQuick that allow researchers to construct their own price indices but generally start only as early as 1988. I match loans to CBSAs using each loan's zip code, as provided by LP, and a 2008 crosswalk between zip codes and CBSAs available from the U.S. Census Bureau.[23]

Other Regional Data. For specifications that examine the importance of local labor market fluctuations, I use Metropolitan Statistical Area and Micropolitan Statistical Area unemployment rates from the Bureau of Labor Statistics (BLS) Local Area Unemployment Statistics series.[24] I also use publicly available Home Mortgage Disclosure Act (HMDA) data to calculate the subprime market share in a given CBSA $\times$ year by merging the lender IDs in the HMDA data with the Department of Housing and Urban Development subprime lender list as in Mian and Sufi (2009).[25] HMDA data discloses the census tract of each loan, which I allocate proportionally to CBSAs using a crosswalk from tracts to zip codes and then from zip codes to CBSAs.

4 Estimation and Identification

4.1 Estimation

Arranging the data into a monthly panel with a dependent variable $default_{icgt}$ equal to unity if existing mortgage i defaulted in month t, the likelihood $h(t)$ of observing failure for a given monthly observation must take into account the sample selection process. Namely, loans are not observed after they have defaulted, so the likelihood of sampling a given observation is a discrete hazard, which conditions on failure not having yet occurred. Suppressing dependence on X, the discrete

---

[23]Available at http://www.census.gov/population/metro/data/other.html.

[24]Available at http://www.bls.gov/lau/home.htm.

[25]Using the HUD subprime lenders list to mark mortgages as subprime results in both false positives and false negatives: lenders who self-designate as predominantly subprime certainly issue prime mortgages as well, and non-subprime-identifying mortgage lenders also issue subprime mortgages. See Mayer and Pence (2008).

hazard is

$$
\begin{aligned}
h(t) &\equiv \Pr(default_{icgt} = 1) \\
&= \Pr(\tau \in (t-1,\ t] \,|\, \tau > t-1) \\
&= \int_{t-1}^{t} f(\tau)d\tau / S(t-1) \\
&= (F(t) - F(t-1))/S(t-1) \\
&= 1 - S(t)/S(t-1)
\end{aligned}
$$

where $f(\cdot)$ and $F(\cdot)$ are the density and cumulative density of τ, the random variable representing mortgage duration until failure, and $S(\cdot) = 1 - F(\cdot)$ is the survivor function, the unconditional probability that observed mortgage duration exceeds the given amount of time. Using the familiar identity that $S(t) = \exp(-\Lambda(t))$, where $\Lambda(\cdot)$ is the integrated hazard function $\Lambda(t) = \int_0^t \lambda(\tau)d\tau$, I can rewrite the likelihood of observing failure for a given observation to be

$$
\begin{aligned}
h(t|X) &= 1 - \exp(-\Lambda(t) + \Lambda(t-1)) \\
&= 1 - \exp\left(-\int_{t-1}^{t} \exp(X(\tau)'\beta)\lambda_0(\tau)d\tau\right)
\end{aligned}
$$

where the last line used the specification of $\lambda(\cdot)$ in equation (2). If time-varying covariates are constant within each discrete time period (for example if the observed value of X_t represents the average of $X(\tau)$ for $\tau \in (t-1,\ t]$),

$$
h(t|X) = 1 - \exp\left(-\exp(X_t'\beta)(\Lambda_0(t) - \Lambda_0(t-1))\right). \tag{4}
$$

where $\Lambda_0(\cdot)$ is the integrated baseline hazard $\Lambda_0(t) = \int_0^t \lambda_0(\tau)d\tau$.

Incorporating this likelihood of observing $default_{icgt} = 1$, each month $\times$ loan observation's contribution to the overall log-likelihood is

$$
\ell_{icgt} = default_{icgt} \cdot \log(h(t|X_{icgt})) + (1 - default_{icgt})\log(1 - h(t|X_{icgt})). \tag{5}
$$

I can then estimate the hazard model parameters of equation (2) by Quasi-Maximum Likelihood (MLE) in a Generalized Linear Model framework where the link function $G(\cdot)$ satisfying $h(t) = G^{-1}(X_t'\beta + \psi_t)$ is the complementary log-log function

$$
G(h(t)) = \log(-\log(1 - h(t))) = X_t'\beta + \underbrace{\log(\Lambda_0(t) - \Lambda_0(t-1))}_{\psi_t}.
$$

Estimating a full set of dummies ψ_t allows for the baseline hazard to be fully nonparametric à la Han

and Hausman (1990).[26] The estimates of the baseline hazard function represent the average value of the continuous-time baseline hazard function $\lambda_0(\cdot)$ over each discrete interval $\bar{\lambda}_{0t} = \int_{t-1}^{t} \lambda_0(\tau)d\tau$ and are obtained as $\hat{\bar{\lambda}}_{0t} = \exp(\hat{\psi}_t)$. Under the usual MLE regularity conditions, estimates of β and ψ will be consistent and asymptotically normal.

4.2 Identification

The proportional hazard model is identified—implying that the population objective function is uniquely maximized at the true parameter values—under the assumptions that 1) conditional on current covariates, past and future covariates do not enter the hazard (often termed strict exogeneity), and 2) any sample attrition is unrelated to the covariates (Wooldridge, 2007).[27] Stated in terms of the conditional distribution $F(\cdot|\cdot)$ of failure times τ, the strict exogeneity and non-informative censoring assumptions are met provided

$$F\left(\tau|\tau > t-1, \{X_{icgs}, c_{is}\}_{s=1}^{T}\right) = F(\tau|\tau > t-1, X_{icgt})$$

where c_{is} is an indicator for whether loan i was censored at time s. In principle, if lags or leads of the covariates enter into λ, the strict exogeneity condition can be satisfied by including them as explanatory variables in the vector X_{icgt}.

An important form of censoring in mortgage data arises from borrowers paying back their mortgages in full. Mortgages that have been prepaid are treated as censored because all that can be learned about their latent time until termination by default is that it is at least as long as the observed elapsed time until prepayment. Technically, any such hazard model with multiple failure types is a competing risks model, which can be generalized to accommodate the potential dependence of one risk on shocks to another. Under the assumption there is no unobserved individual heterogeneity in the default hazard (or that unobserved heterogeneity in the default and prepayment hazards are independent at the individual level), competing risks models can be estimated as separable hazard models with observations representing other failure types treated as censored.[28] As in Gerardi et al. (2008), Sherlund (2008), Foote et al. (2010), and Demyanyk and Van Hemert (2011), I adopt this approach and focus on estimation of the default hazard.[29] I also verify the

---

[26]Alternatively, ψ_t can be thought of as estimating a piecewise-constant baseline hazard function. As discussed above in the context of the geographic fixed effects, the incidental parameters problem is not a concern here since increases in sample size (the number of loans) would not increase the number of ψ needing to be estimated.

[27]The linear-index functional form assumption that the effect of covariates on the hazard is linear in logs is not necessary for identification and is made for the sake of parsimony and convenience in interpreting the coefficients.

[28]See Heckman and Honoré (1989) for a full discussion of identification in competing risks models.

[29]The most well-known example of allowing for correlated default and prepayment unobserved heterogeneity is

robustness of my main results to allowing for unobserved heterogeneity in the default hazard.

Turning to causality, the key identifying assumption for the estimated coefficient μ in equation (3) to be interpretable as the causal effect of the decline in property values is that fluctuations in home prices and unobserved shocks to default risk are independent.[30] To illustrate how the exogeneity of X affects estimates of β in a hazard model setting, consider the case of time-invariant covariates and no censoring. In this simplified setting, the exogeneity condition necessary for the maximum likelihood estimates of the hazard model parameters to represent causal effects is that the probability of failure (conditional on reaching a given period) is correctly specified in (2) and (3). Again, letting τ be the random variable denoting the mortgage duration until failure, the formal condition is

$$\lim_{\xi \to 0^+} E\left[\frac{1\left(\tau \in (t - \xi, t]\right)}{\xi} - \lambda(X_{icgt}, t)\Big| X, \tau > t - \xi\right] = 0 \tag{6}$$

where $1(\cdot)$ is the indicator function. Analogous to omitted variables bias in a linear regression, this condition would be violated if there were an omitted factor ω which affects default rates and is not independent of X. In this case, misspecification leads to violation of the exogeneity assumption because ω affects failure, is not in λ, and survives conditioning on X. To see this, suppose that the true instantaneous probability of default conditional on $\tau > t - \xi$ is not $\lambda(X, t)$ but is

$$\tilde{\lambda}(X, \omega, t) = \exp(X\beta + \omega)\tilde{\lambda}_0(t),$$

where both X and ω may depend on t. Then

$$\lim_{\xi \to 0^+} E\left[\frac{1\left(t - \xi < \tau \leq t]\right)}{\xi} - \lambda(X, t)\Big| X, \tau > t - \xi\right] = E\left[\tilde{\lambda}(X, \omega, t)\big| X\right] - \lambda(X, t)$$
$$= E\left[e^\omega \exp(X\beta)\tilde{\lambda}_0(t)\big| X\right] - \lambda(X, t).$$

If ω and X are independent, then the exogeneity condition becomes

$$E\left[e^\omega \exp(X\beta)\tilde{\lambda}_0(t)\big| X\right] - \lambda(X, t) = \exp(X\beta)E\left[e^\omega\right]\tilde{\lambda}_0(t) - \exp(X\beta)\lambda_0(t).$$

Thus, the presence of independent ω simply scales the estimate of the baseline hazard function. In other words, the baseline hazard function estimated without controlling for ω will be estimating $E\left[e^\omega\right]\tilde{\lambda}_0(t)$—but the estimation of the slope coefficients will be unaffected and the exogeneity condition of equation (6) will hold in expectation. However, if ω and X are not independent, then the

Deng et al. (2000), who jointly estimate a competing risks model of mortgage termination using the mass-points estimator of McCall (1996).

[30]This condition is stronger than price and default shocks being uncorrelated and is required in non-additive models. See Imbens (2007) for a discussion.

omission of ω leads to a violation of equation (6), and estimated β will not represent the marginal effect of X on default, as discussed in Section 5.2 below.[31]

In the general case, even independent unobserved heterogeneity will affect the conditional distribution of $\tau|X$ (and hence the estimated coefficients), a common obstacle in nonlinear panel models. Lancaster (1979) introduced the Mixed Proportional Hazard (MPH) model where the heterogeneity enters in multiplicatively (additively in logs).[32] Conditional on unobserved heterogeneity ε, the hazard function becomes

$$\lambda(t|X_{icgt}, \varepsilon_i) = \exp(X'_{icgt}\beta + \varepsilon_i)\lambda_0(t). \tag{7}$$

The literature on unobserved heterogeneity in duration models has broadly found that ignoring unobserved heterogeneity biases estimated coefficients down in magnitude. Intuitively, the presence of ε induces survivorship bias—loans with low draws of ε last longer and are thus overrepresented in the sample relative to their observables. Individuals whose observable characteristics put them at a high ex-ante risk of default and yet have lengthy durations are likely observed in the sample because they have low unobserved individual-specific default risk (high latent quality). The negative correlation between X and ε induced by the sample selection process can prevent consistent estimation of β.

Equation (7) pins down the conditional distribution F of latent failure times τ to be

$$F(\tau|X_{icgt}, \varepsilon_i) = 1 - \exp\left(-\Lambda((t|X_{icgt}, \varepsilon_i)\right)$$

where $\Lambda(\cdot|X, \varepsilon)$ is the integrated hazard. Specifying the distribution of ε to have cumulative distribution function $G(\cdot)$, the distribution $\tilde{F}(\tau|X_{icgt})$ of $\tau|X$ is then obtained by integrating out ε:

$$\tilde{F}(\tau|X_{icgt}) = \int_{-\infty}^{\infty} F(\tau|X_{icgt}, \varepsilon_i)dG(\varepsilon_i).$$

Finally, the modified likelihood $\tilde{h}(t|X)$ of observing failure at time $\tau \in (t-1, t]$ is

$$\tilde{h}(t|X) = 1 - \tilde{S}(t|X)/\tilde{S}(t-1|X) \tag{8}$$

where the new survivor function is denoted $\tilde{S}(\cdot|X) = 1 - \tilde{F}(\cdot|X)$. Estimation then proceeds by replacing $h(\cdot|X)$ with $\tilde{h}(\cdot|X)$ in the log-likelihood expression of equation (5). After presenting

---

[31]Estimating a proportional hazard model with no censoring and time-invariant covariates is equivalent to a linear regression of log duration on the covariates (Wooldridge, 2007). This illustrates why this special case permits unobserved heterogeneity provided it is independent of the covariates; in a linear model, additive unobservables affect the consistency of the parameter estimates only if they are correlated with the covariates.

[32]Elbers and Ridder (1982) showed that the MPH model is identified provided there is at least minimal variation in the regressors.

my main results, I verify that my results are robust to the presence of independent unobserved heterogeneity by specifying $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ so that $G(\varepsilon) = \Phi(\varepsilon/\sigma)$, where $\Phi(\cdot)$ is the standard normal cumulative density function.[33]

4.3   Isolating Long-Run Variation in Housing Price Cycles

One example of an omitted factor that may be correlated with X is the expansion of subprime credit, which may initially increase prices as a positive shock to the demand for owner-occupied housing, as suggested by Mayer and Sinai (2007), Mian and Sufi (2009), and Pavlov and Wachter (2009). If the credit expansion leads to a decrease in the quality of the marginal borrower, prices will eventually fall as these riskier borrowers default, depressing prices both from a positive shock to the supply of owner-occupied housing on the market and from negative foreclosure externalities (see Hartley, 2010 and Campbell et al., 2011).[34] Thus, the expansion of subprime credit may be an omitted variable that directly affects both defaults (by decreasing the quality of the marginal subprime borrower) and prices, potentially leading to a spurious estimated relationship between prices and defaults. A related worry from the perspective of the exogeneity condition in equation (6) is that areas with the strongest price declines are also likely the areas hit hardest by the recession. If a negative employment shock simultaneously causes both defaults and price declines, then local labor market strength may be an important omitted variable that biases the estimates towards finding an effect of prices on default. Below, I discuss how I account for each of these potential biases.

To address these endogeneity concerns, I develop an instrument that isolates the long-run component of each Core Based Statistical Area's (CBSA) price cycle and is arguably independent of contemporaneous shocks to prices or default rates, e.g. from credit or labor market fluctuations. The CoreLogic repeat-sales price index for each CBSA, discussed in greater detail above, provide a measure of the relative level of nominal house prices in a given CBSA $\times$ month, denoted here as HPI_{gt}. Sinai (2012) notes that a similar set of metropolitan areas had large 1980s and 2000s price cycles. Using this persistence, I determine the portion of a CBSA's price cycle that is predictable using only the historical cyclicality of that city. First, I form a summary measure σ_g^P quantifying

---

[33]There is a large literature on the relative merits of parametric assumptions on the baseline hazard function and the unobserved heterogeneity distribution. See Lancaster (1979), Heckman and Singer (1984), Han and Hausman (1990), Meyer (1992), Horowitz (1999), and Hausman and Woutersen (2012).

[34]Dagher and Fu (2011) provide an example of the mechanism behind such an expansion: counties that had significant entry of non-bank mortgage lenders had stronger growth in credit and prices, as well as stronger subsequent increases in defaults and decreases in prices. Brueckner et al. (2012) offer a model of how price increases could fuel lender expectations and further credit expansion. Berger and Udell (2004) also discuss empirical evidence of underwriting standards deteriorating during a credit expansion.

the long-run cyclicality of CBSA g defined as the standard deviation of monthly changes in the CoreLogic repeat sales home price index from 1980-1995

$$\sigma_g^P \equiv \left(\frac{1}{T-1} \sum_{t \in \mathcal{T}} (\Delta HPI_{gt} - \overline{\Delta HPI}_g)^2 \right)^{1/2} \tag{9}$$

where $T = 180$ is the number of months over which the standard deviation is calculated; $\mathcal{T}$ is the set of months from January 1980 to December 1995, inclusive; $\Delta HPI_{gt} = HPI_{gt} - HPI_{gt-1}$; and $\overline{\Delta HPI}_g$ is the average value of ΔHPI_{gt} for CBSA g and $t \in \mathcal{T}$.[35] Figure 5 shows the average value of the CoreLogic repeat sales home price index by quartile of σ^P. The persistence in price volatility isolated by the first stage is visible: the average price cycle in the late 2000s was much more pronounced for CBSAs that had stronger price cycles in the 1980s, that is, higher quartiles of σ^P have monotonically stronger price cycles.

5 Results

5.1 Results Treating Price Changes as Exogenous

Table 3 reports estimates of equation (2) using the estimator described above, treating price changes as exogenous to offer initial estimates of the relationship between price changes, underwriting standards, and cohort-level differences in default rates. I cluster all standard errors at the CBSA level to account for area-specific shocks to the default rate in inference. All specifications include nonparametric controls for the baseline hazard function.[36] Column 1 includes only cohort fixed effects to quantify the pattern of declining cohort-level performance from Figure 1 in a hazard-model framework. These coefficients can be interpreted as the change in the log hazard rate and imply, for example, that subprime loans in the 2007 cohort had a default hazard 73 log points greater than the 2003 cohort (the omitted category). These unadjusted cohort coefficients are large and precisely estimated, implying that the probability of a 2005–2007 cohort mortgage defaulting in any given month conditional on the mortgage having survived to that month is more than twice as high as 2003 cohort mortgages. Column 2 adds fixed effects for each CBSA in the sample (570 fixed ef-

---

[35]I calculate the standard deviation of the first differences in the HPI variable to emphasize the importance of the (low-frequency) price cycle. CBSAs with high variance of HPI in levels (as opposed to high ΔHPI) could simply be areas that had sustained price growth or high-frequency volatility.

[36]The baseline hazard controls consist of an indicator variable for each possible value of loan age from 1–70 months, with the final indicator variable also turned on for all values of loan age exceeding 70 months. The estimated baseline hazard functions resemble the hump-shaped baseline hazards of Deng et al. (2000) and are available from the author on request.

fects) to verify that cohort differences are not driven by the geographic composition of later cohorts. Conditioning on CBSA fixed effects does not materially affect the estimated differences in cohort default hazards.

Columns 3 and 4 add borrower characteristics and loan characteristics, respectively, as detailed in Section 3. The coefficients on these credit risk factors all have intuitive signs. Borrowers had higher default rates if they lacked full income documentation, were not owner-occupants, or had lower FICO scores and higher DTI ratios. Mortgages defaulted more frequently if they were non-fixed rate mortgages, had higher CLTVs or interest rates, or were accompanied by additional liens. Column 3, which includes only borrower characteristics, shows that the adjusted default hazard of earlier cohorts is higher than in column 1, suggesting that, relative to 2003 borrowers, 2004 and 2005 subprime borrowers underperformed relative to what would be expected based on their individual attributes. For 2006–2007 cohorts, the differential default hazard is lower than in column 1, although the average decrease between column 1 cohort effects and column 3 cohort effects is approximately zero. The inability of borrower characteristics to substantively explain the cohort-level differences is not surprising given the summary statistics reviewed above showing that the mean observable attributes of borrowers are not changing much across cohorts.[37] The results of column 4 tell a different story: including controls for loan characteristics and not borrower characteristics explains on average 24% of the unadjusted cohort effects estimated in column 1. This suggests that the loan characteristics that were changing across cohorts were an important driver of defaults. Conditioning on both borrower and loan characteristics together in column 5 reduces the residual cohort heterogeneity (i.e. the column 4 coefficients relative to the column 1 coefficients) by an average of 29%.

To get a sense of which covariates are most important in explaining the cohort pattern, I estimated the specification of column 5, leaving out one characteristic at a time. Three characteristics stand out as contributing substantially the attenuation of the estimated cohort effects: the balloon and interest-only dummies and the loan interest rate. As the interest rate should represent everything that the market knew about the riskiness of the loan, its importance reenforces that priced observables are important in predicting the cohort-level default pattern. The importance of the balloon and interest-only indicators is consistent with Table 2, which showed that balloon mortgages and interest-only mortgages were the two product characteristics that changed the most

---

[37]While individual borrower characteristics do not explain much of the differences in default rates across cohorts, they are individually strong predictors of default, as evidenced by the large increase in the log likelihood value between columns 2 and 3.

across cohorts and thus had the strongest potential to explain cohort-level defaults.

Column 6 drops all borrower- and loan-level covariates and instead controls for the 12-month change in log of the CoreLogic repeat-sales Home Price Index (HPI), defined at the CBSA-level as

$$\Delta \log(HPI_{icgt}) \equiv \log(HPI_{icgt}) - \log(HPI_{icgt-12}) \tag{10}$$

where HPI_{icgt} is the value of the CoreLogic repeat-sales price index for CBSA g in the calendar month corresponding to loan i having a duration of t.[38] This variable is a strong predictor of default. The coefficient on the 12-month change in log HPI implies that properties experiencing the 75th percentile 12-month price change (+5%) would have a 33% lower hazard than properties exposed to the 25th percentile 12-month change in prices (–5%), corresponding to an approximately one percentage point decrease in the annual default rate. Controlling for the 12-month change in prices, the cohort effects in column 6 are lower than the estimates in column 5, showing that price changes in the most recent 12 months seem to be more closely related to observed cohort heterogeneity than borrower and loan characteristics. The residual differences in default rates across cohorts decrease on average by 50% (depending on the cohort) relative to the baseline cohort coefficients in column 1.

Controlling for both borrower and loan characteristics and price changes leaves little cohort-level heterogeneity unexplained. The estimates in column 7 of the latent quality of each cohort (i.e. the portion of cohort outcomes not attributable to price changes or individual-level controls) are statistically insignificant with the exception of the 2005 cohort. While statistically significant, more than 70% of the unadjusted estimate of the difference between the 2003 and 2005 cohorts (column 1) is explained by prices and observables.

These results illustrate that observable loan characteristics and prices play important roles in explaining the heterogeneity in default rates across origination cohorts, together explaining on average 95% of the cohort disparities in column 1.[39] In particular, places where price declines are greater have higher default rates, and the incidence of these price declines is disproportionately borne by later cohorts. I now turn towards developing causal estimates of the impact of prices on default behavior.

---

[38]I index HPI by i as well to emphasize that in my notation t refers to event time (i.e. loan age). Even though HPI only varies by CBSA × calendar month, for example, not all six-month old ($t = 6$) mortgages in CBSA g have the same HPI value.

[39]The additional explanatory power gained from controlling for prices and characteristics simultaneously suggests that there are important interactions between prices and loan and borrower characteristics. One implication of the proportional-hazard framework is that interactions between the covariates is implicit: the cross-partial of the hazard function with respect to two covariates is the hazard function times the product of the two coefficients on the covariates. For example, this multiplicative relationship between the covariates allows for price declines to have larger effects for riskier borrowers.

5.1.1 Unobserved Heterogeneity

This section examines the robustness of the above results to misspecification from ignoring independent unobserved heterogeneity ε by allowing the true hazard model to be specified as in (7). The results of maximizing the sample log-likelihood function described by (5), replacing $h(t|X)$ with $\tilde{h}(t|X)$ defined in equation (8) and modeling $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, are presented in Table 4. There are two important caveats in comparing these results to the results of Table 3. Because of the computational burden of maximizing the likelihood while integrating out the unobserved Gaussian heterogeneity, columns 1–4 do not include geographic fixed effects or cluster standard errors by CBSA as in the rest of the paper. Column 5 includes state fixed effects to test how sensitive the point estimates are to controlling for constant differences across regions.

Column 1 shows that the unadjusted differences in default rates across cohorts is even more pronounced when accounting for independent unobserved heterogeneity than the baseline results of column 1 of Table 3. I account for this by comparing the adjusted cohort coefficients in columns 2–5 to column 1 of Table 4. Including borrower and loan characteristics in column 2 explains 32% of cohort heterogeneity—the average decrease in the estimated cohort dummies. Controlling instead for 12-month price changes in column 3 reduces the residual difference in the default hazard across cohorts by an average of 68%. Conditioning on both price changes and loan and borrower characteristics in column 4 explains 92% of the cohort differentials in column 1. The total explanatory power of prices and observables is attenuated somewhat by including state fixed effects in column 5, where the combination of prices and observables explains 81% of the cohort pattern in column 1. Still, only the 2005 cohort is statistically significant at the 95% confidence level, and these standard errors are likely a lower bound because they do not allow for spatial correlations in default risk. Taking columns 4 and 5 together, as before, the 2005 cohort is the only borrower cohort to have a default hazard that is statistically distinguishable from the 2003 cohort hazard after adjusting for prices and loan and borrower observables, although these covariates explain 73% (column 5) to 81% (column 4) of the 2005 cohort coefficient in column 1. I conclude that the qualitative pattern of Table 3 is robust to allowing for independent unobserved heterogeneity: prices explain over 60% of cohort heterogeneity in default risk and combined with borrower and loan characteristics explain approximately 90% of the increase in defaults across cohorts.

5.2 Nonlinear Instrumental Variables Estimation

As discussed above, the interpretation of these results as causal requires the strong assumption that changes in the average default risk of a given area are not the cause of local price changes unless they are captured by loan and borrower covariates. Because defaults themselves cause price declines, this assumption is likely to be violated by any shock to area default risk. The demand shock resulting from the credit expansion may initially increase prices, and eventually a higher share of riskier borrowers may exacerbate price declines. In this way, if price changes are endogenous to subprime penetration and subprime growth reduces unobserved borrower quality, then the estimation would misattribute much of the increase in defaults to price changes instead of to differences in unobserved cohort quality. A second way that price declines may be endogenous to other factors that also affect default risk is from fluctuations in local labor market conditions. Adverse local labor shocks may simultaneously decrease prices (negative demand shock for owner-occupied housing) and increase defaults (negative income shock to existing mortgage borrowers).

The potential for changes in local house prices to themselves be a function of contemporaneous shocks to the default hazard through subprime lending or employment shocks necessitates instrumenting for prices. To instrument in this nonlinear setting, I use the control function approach (see Heckman and Robb, 1985). This estimator involves conditioning on a consistent estimate of the endogeneity in the endogenous explanatory variable and in a linear model is equivalent to two-stage least squares.[40]

To see why the control function approach solves the endogeneity problem, suppose again that there exists an omitted variable ω in the default hazard equation, which is not independent of X. Labeling the true hazard function $\tilde{\lambda}(\cdot)$, if

$$\tilde{\lambda}(X, \omega, t) = \exp(X\beta + \omega)\lambda_0(t) = e^\omega \exp(X\beta)\lambda_0(t).$$

If I do not control for ω in estimating this model, the resulting β coefficients will be estimating a different object than the marginal effect of X on the log hazard. Formally, the exogeneity condition introduced in equation (6) above now fails:

$$
\begin{aligned}
E\left[default_t - \lambda(X, t) \big| X, \tau > t\right] &= E\left[\tilde{\lambda}(X, \omega, t) \big| X\right] - \lambda(X, t) \\
&= E\left[e^\omega \big| X\right] \lambda(X, t) - \lambda(X, t)
\end{aligned}
$$

---

[40]Unlike a linear model, consistency of the control function approach in a nonlinear model relies on the instrument and the endogenous portion of the endogenous explanatory variable being independent (as opposed to just uncorrelated).

$$= [\exp(X\beta + f(X)) - \exp(X\beta)] \lambda_0(t)$$

$$\neq 0$$

where $E(e^\omega|X) \equiv f(X)$ because X and ω are not independent. Thus, under misspecification, the coefficients on X will not converge to the marginal effect of X on the log hazard and instead combine both the direct effect of X on default and the indirect effect of ω on default after projecting onto X.

Conditioning on an estimate of the endogenous component of X solves this problem. Let the right-hand side endogenous variable be specified as

$$\Delta Prices = Z_1\Pi_1 + Z_2\Pi_2 + v$$

where the endogeneity problem arises because v and ω are not independent. The key identifying assumption is that the instruments Z_1 and included right-hand side controls Z_2 (the elements of X apart from $\Delta Prices$) are independent from v and ω. Conditioning on v then satisfies the exclusion restriction

$$
\begin{aligned}
E\left[default_t - \lambda(X,v,t)\big|X,v,\tau > t\right] &= E\left[\tilde{\lambda}(X,\omega,t)\big|X,v\right] - \lambda(X,t,v) \\
&= E\left[e^\omega|v\right]\lambda(X,t,v) - \lambda(X,t,v) \\
&= (\exp(g(v)) - \exp(\rho_1 + \rho_2 v))\exp(X\beta)\lambda_0(t)
\end{aligned}
$$

where $g(v) \equiv E(e^\omega|v)$. If the conditional expectation $E(e^\omega|v) = \exp(\rho_1 + \rho_2 v)$, then this condition will hold, and controlling for a consistent estimate of v will be sufficient to allow estimation of the partial effect of X on the log hazard. This will be satisfied exactly if ω conditional on v is distributed normally: if $\omega|v \sim \mathcal{N}(\rho_2 v, 2\rho_1)$ then $e^\omega|v$ is distributed log normally with mean $E(e^\omega|v) = \exp(\rho_1 + \rho_2 v)$. If the conditional distribution of ω given v is non-normal, then controlling linearly for v in the hazard model relies on the quality of the linear model as a first-order approximation to the conditional mean function. As a robustness check, below I consider third- and fifth-order series approximations to the log of the conditional expectation function, e.g. $\log\left(E(e^\omega|v)\right) = \sum_{k=0}^{5}\rho_k v^k$ and find that the results are insensitive to this flexibility.

5.2.1 First Stage

The instrument set for the price change variable is the long-run cyclicality measure σ_g^P interacted with calendar-month indicator variables. The first stage for the 12-month price change is then

$$\Delta\log(HPI_{icgt}) = \sum_s \pi_s \sigma_g^P \cdot 1(s = t + t_0(i)) + Z'_{2,icgt}\pi_2 + v_{icgt} \tag{11}$$

where Z_{2icgt} contains the same covariates as equation (3) above—cohort effects, geographic fixed effects, loan and borrower characteristics, and the nonparametric baseline hazard function to ensure that predicted values from equation (11) are orthogonal to the other controls in equation (2). The function $t_0(i)$ evaluates to the calendar time of loan i's origination date, and the π_s coefficients are turned on when the observation on loan i at t months after origination corresponds to calendar month s.

Table 5 reports the results from estimating equation (11) by OLS with standard errors clustered at the CBSA level. Column 1 includes just the instrument set and no other controls. The statistical relationship between actual price changes and the interactions between the cyclicality measure and calendar time is strong—the instruments explain 50% of the variation in twelve-month CBSA-level house price changes. Adding controls for the baseline hazard and CBSA fixed effects in column 2 improves the overall fit slightly (R^2 increases to .56). Including loan and borrower characteristics in column 3 does not affect the partial F-statistic, which tests the joint hypothesis that all of the coefficients on the instrument set are zero, suggesting that weak instruments are not a problem in this setting. The cohort coefficients in columns 2 and 3 illustrate that later cohorts were exposed to stronger price declines than earlier cohorts, in part by virtue of selection—younger loans are statistically more likely than older loans to not have terminated.

To provide intuition for how this instrument operates, I compute counterfactual price indices by regressing log home price indices on geographic fixed effects and an interaction of σ_g^P with calendar-month indicators as follows

$$\log(HPI_{gt}) = \alpha_g + \sum_s \pi_s \sigma_g^P \cdot 1(s = t) + u_{gt} \tag{12}$$

where HPI_{gt} is the value of the CoreLogic home price index in CBSA g in calendar month t. The estimated $\hat{\pi}_s$ shift the baseline log HPI of each CBSA (α_g) according to the cross-sectional relationship each calendar month between prices and 1980s price volatility.[41] Predicted values $\widehat{\log HPI}_{gt}$ from this regression provide an alternative time series of home prices in geography g based on the quasi-fixed tendency of home prices in geography g to cycle up and down.

Figure 6 shows the actual log home price series for 2003–2013 (left-hand panel) along with predicted values from equation (12) (right-hand panel). The left-hand panel shows that the actual

---

[41]It is worth pointing out that equation (12) does not control for main effects for each date. While this loads much of the national month-to-month variation in house prices onto the π_t, date effects are the very object the hazard model seeks to explain. As they aren't instruments and they don't belong in the second-stage, I purposefully omit them here.

HPI series are characterized by idiosyncratic deviations from the national trend, i.e. price shocks that potentially arise from such factors as local credit expansions and local labor market fluctuations that may also independently affect default rates. It is precisely the effects of these types of shocks that the instrument is designed to abstract from. Because nothing in equation (12) allows for differential price trends across CBSAs, the predicted time series in the right-hand panel all change in the same direction each month, differing only in the magnitude of the price change depending on their historical price volatility. To the extent that the actual price paths reflect time-varying local housing market changes, each line on the right is an estimate of the counterfactual price path that might occur absent local price shocks that are potentially driven by factors that also affect local default risk. Intuitively, my empirical strategy instruments for the actual price series on the left with the predicted price series on the right.

5.2.2 Exclusion Restriction

The necessary exclusion restriction for the IV results to be unbiased estimates of the causal effect of price changes is the independence of the size of a CBSA's 1980s price cycle (σ_g^P) from any other factors that affect default (besides prices). Note that with CBSA fixed effects, it is not a threat to identification if cyclical areas are fundamentally different from acyclical areas in some time-invariant way (e.g. an inherently risky area may always have both higher defaults and larger price swings). However, this exclusion restriction would be violated if pro-cyclical areas (high σ_g^P) have pro-cyclical trends in prices and the credit risk of borrowers. For example, if high-σ_g^P areas had more rapid subprime growth, then σ_g^P may proxy for changes in unobserved borrower quality in CBSA g. Similarly, if high-σ_g^P areas have greater unemployment rate fluctuations, these adverse shocks to local aggregate demand could increase defaults (through an income shock) and decrease prices (through a demand shock).

Figures 7 and 8 offer graphical evidence that subprime shares and unemployment rates—adjusting both for CBSA fixed effects—did not vary systematically with σ_g^P. The relevant period is different for each endogeneity concern. Figure 7 plots the annual adjusted subprime share of HMDA-covered mortgages originated in 2003–2007 by quartiles of σ_g^P. There is no apparent relationship between σ_g^P and subprime originations—places with historically large price cycles do not seem to have been any more prone to subprime credit expansion.[42] Figure 8 shows that the top

---

[42]Note that the same fact is not true about the relationship between subprime originations and the size of the late 2000s price cycle—areas that originated the highest share of subprime mortgages indeed had stronger (contemporaneous) price cycles, further evidence of the need for instrumenting.

quartile of σ_g^P had around a 1 percentage point *lower* unemployment rate in recession than the bottom quartile.

Regression versions of these tests tell a somewhat different story. I test whether there is a first stage for the annual subprime share of all residential mortgage originations and the monthly unemployment rates by re-estimating equation (11), replacing the dependent variable with the subprime share of mortgages originated in each cohort and with the monthly unemployment rate in each CBSA. Consistent with Figure 7, the pattern of estimated coefficients $\hat{\pi}_t$ for the subprime share first stage is nearly completely flat and statistically insignificant, showing that loans from areas with higher historical price volatility were no more likely to have been originated during a relatively large subprime credit expansion. However, unlike in Figure 8, the estimated $\hat{\pi}_t$ in the unemployment regression mimic the national trends in the unemployment rate, with historically cyclical areas having differentially lower unemployment rates leading up to the recession and more quickly rising unemployment rates thereafter. This illustrates that areas with historically cyclical housing markets also have cyclical labor markets and that national labor market changes load onto the instruments. From this analysis, I conclude that σ_g^P successfully allows isolation of the effect of prices on defaults in a world where price declines and borrower quality are not jointly determined but that instrumental variables estimates are likely confounded by changing labor market conditions. An important caveat is that housing market changes can also affect labor markets (see Mian and Sufi, 2012). To the extent that the observed correlation between my instrument and labor market outcomes is an effect of the price cycle and not vice versa, then the instrument captures the total causal effect of price changes. However, because of the difficulty in ascertaining which caused which, I treat the relationship between the instrument and unemployment as a threat to validity. To account for this unemployment channel, I present additional control function specifications below that also control for the unemployment rate, thereby isolating the variation in prices that is not correlated with local labor market shocks or local subprime expansion.[43]

---

[43]The relationship between house price cyclicality and labor market cyclicality hints at the economics behind why some areas may be more cyclical than others. Areas with high housing supply elasticity, potentially arising from geographic constraints, land-use regulations, or credit market regulations, could be pro-cyclical in both markets. Similarly, areas with an industry mix that makes them particularly sensitive to recessions or commodity price shocks may experience coincident fluctuations in housing and labor.

5.2.3 Control Function Results

Table 6 employs a nonlinear IV control function approach, which accounts for the endogeneity of price to the credit expansion by controlling for the first-stage residuals $\hat{v}_{icgt}$ in the default hazard

$$X'_{icgt}\beta = \gamma_c + W'_{B,i}\theta_B + W'_{L,i}\theta_L + \mu \cdot \Delta Prices_{icgt} + \kappa \hat{v}_{icgt} + \alpha_g \tag{13}$$

where $\hat{v} = \Delta \log(HPI) - \Delta \widehat{\log HPI}$ and $\Delta \widehat{\log HPI}$ is fitted from equation (11). To account for the generated regressor problem in inference (Pagan, 1984 and Murphy and Topel, 1985), I also report bootstrapped standard errors in brackets below clustered standard errors. The generated regressor problem arises because v depends on an unknown parameter vector π, as seen in equation (11). Consistently estimating π in a first stage to generate $\hat{v}$ does not affect the consistency of parameters estimated in (13). However, by treating $\hat{v}$ as fixed, i.e failing to account for the correlation between the estimation error in $\hat{\pi}$ and the error in estimating β, the usual asymptotic standard errors will generally be understated unless $\kappa = 0$. The block bootstrap solves this by mimicking the data-generating process. In this setting, individual mortgages are resampled with replacement instead of month $\times$ loan observations being drawn with replacement as would be the approach of standard nonparametric bootstrap. The two stages (estimating $\hat{v}$ from (11) and estimating equation (13)) are then run on each bootstrapped sample and the resulting bootstrap standard errors are the empirical standard deviation of each element of β across 200 bootstrap replications.

Column 1 of Table 6 repeats column 6 of Table 3, controlling for the 12-month change in prices and not conditioning on borrower or loan observables W_B and W_L. Column 2 additionally controls for the residuals $\hat{v}_{icgt}$, estimated from OLS on equation (11) (omitting loan and borrower characteristics in the construction of the residuals). The coefficient on the price change variable is still large and significant—borrowers experiencing a 1% price decline over the previous year have a 4.4% higher conditional probability of default. The adjusted cohort differences are smaller in column 2 than column 1, meaning that after accounting for endogeneity, the role of prices in explaining the default pattern is larger. Comparing column 2 to the benchmark differences in cohort performance measured in column 1 of Table 3, controlling and instrumenting for prices without controlling for borrower or loan characteristics explains 60% of the difference in unadjusted cohort outcomes. The statistical significance of the coefficient κ on the residual is equivalent to a Hausman test for the endogeneity of price changes, similar to a Rivers and Vuong (1988) test for endogeneity in a probit model, confirming that price changes are endogenous.

To address the correlation between the instrument and local labor market shocks, column 3

also controls for the monthly CBSA unemployment rate, measured in percentage points.[44] Conditional on the covariates in the column 3 specification, a one percentage point increase in the local unemployment rate is associated with a decrease in the default hazard by 2%. The counterintuitive sign on the monthly unemployment rate and the increase in the magnitude of the coefficient on prices from –4.4 to –4.5 suggests that price results are not driven by correlation between price shocks and local labor shocks. The estimated differences in cohort quality in column 3 do not differ substantively from column 2.

The bootstrapped standard errors in columns 2 and 3 (in brackets) are in general much larger than the standard errors clustered at the CBSA level (in parentheses), representing a high degree of variability in the estimated residuals $\hat{v}$ across bootstrap samples using the control function approach. However, the relative stability of the coefficient magnitudes suggest that the patterns described above hold at least qualitatively. Further, because the asymptotic standard errors are correct under the null hypothesis $H_0 : \kappa = 0$, the conventional t-statistic on the fitted residuals is still a valid test of exogeneity.

The next three columns additionally control for borrower and loan characteristics. The estimated cohort effects in these specifications capture the latent quality of each cohort, i.e. the heterogeneity in cohort-level default rates not explained by ex-ante observable quality or price changes. Column 4 repeats column 7 of Table 3 for convenience, controlling for price changes in addition to all of the other controls. Column 5 reports control function estimates of this specification. The coefficient on the price change variable increases in magnitude from –3.9 to –4.6. The coefficient on the endogenous portion of the 12-month change in house prices is again positive and significant. Importantly, I cannot reject that each of the cohort latent quality measures is statistically indistinguishable from zero with the exception of the 2005 cohort, as before. Moreover, the estimated cohort effects in column 4 are each smaller than those in column 3 which treat prices as exogenous. Column 6 again controls for the monthly unemployment rate. The magnitude of unemployment on default is almost identical as in column 3, suggesting that the unemployment rate does not interact meaningfully with loan and borrower characteristics. The price effects—both the main effect and the residuals— are strengthened by the inclusion of the unemployment rate control, although this difference is not statistically significant. Each of the cohort effects is attenuated slightly from column 5. The specifications in columns 5 and 6 both explain 95% of the unadjusted differences in cohort default

---

[44]The sample size decreases slightly in specifications controlling for unemployment rate because of one CBSA for which BLS does not estimate monthly unemployment rates.

rates in column 1 of Table 3.

Interestingly, the bootstrapped standard errors in columns 5 and 6 are much more similar to their asymptotic counterparts than the bootstrapped standard errors of columns 2 and 3. Unlike columns 2 and 3, the results of columns 5 and 6 are robust to bootstrapping the standard errors. This suggests that much of the instability of the bootstrap estimates in columns 2 and 3 is driven by not controlling for loan and borrower characteristics, which explain a substantial amount of individual heterogeneity in default risk.

A consistent pattern in Table 6 is that instrumenting (columns 2, 3, 5, and 6) increases the magnitude of the estimated effect of price changes relative to not instrumenting (columns 1 and 4). An explanation for this is the positive sign on the estimated coefficient $\hat{\kappa}$ on the residuals. While the partial effect of a price shock v on the log of the default hazard—equal to $\mu + \kappa$ because the residuals enter into the $X\beta$ index through both $\Delta Prices$ and v—is strongly negative in each specification, the effect of an exogenous change in prices captured by μ alone is much greater. This is consistent with some degree of treatment effect heterogeneity—if price declines arising from shocks that are correlated with default risk (e.g. credit market changes) have a weaker effect on defaults than price declines induced by, for example, national price declines unrelated to local credit market fluctuations, then isolating the exogenous variation in prices would increase the estimated price effect.

Table 7 addresses the possibility that the conditional distribution of the endogeneity is mis-specified. As mentioned above, controlling for v linearly in $X\beta$ relies on the assumption that the omitted default risk factors ω are distributed normally condition on v. In general, if $\omega | v \nsim \mathcal{N}$ then $E(e^\omega | v) = g(v) \neq \rho_1 + \rho_2 v$. In this case, the specification of $X\beta$ needs to be augmented to include a consistent estimate of $\log(g(v))$, which I approximate using third- and fifth-order polynomials in the fitted residuals, e.g. $\log\left(g(v_{icgt})\right) = \sum_{k=0}^{5} \rho_k \hat{v}_{icgt}^k$. Columns 1–3 do not control for borrower or loan characteristics. Column 1 is repeated from column 1 of Table 6 for convenience. Column 2 adds the residuals squared and the residuals cubed. These coefficients are strongly significant, and a likelihood ratio test for the hypothesis that $\rho_2 = \rho_3 = 0$ rejects, pointing to likely non-normality of the unobserved heterogeneity that is correlated with price shocks. However, the slope coefficients are relatively unaffected from the additional flexibility in the estimate of $\log\left(E(e^\omega | v)\right)$. Column 3 adds fourth- and fifth-order terms, which again do not noticeably affect the estimated effect of prices or differences in the latent quality of cohorts. The estimated coefficients $\hat{\rho}$ on the powers of the residuals in column 3 are very imprecise, and a likelihood ratio test fails to reject that $\rho_4 = \rho_5 = 0$.

Columns 4–6 in Table 7 repeat the specifications in columns 1–3, additionally controlling for borrower and loan characteristics. The same patterns are apparent: powers of the residuals are jointly significant, rejecting the exogeneity of price changes, and the estimated effects of the covariates are relatively unchanged.

The control function results of Tables 5 and 6 are consistent with the results of Table 3, providing evidence that there is a large causal effect of price declines on defaults. Even after accounting for the endogeneity of the effect of prices on default risk and controlling for local labor market conditions, there is little evidence that unobserved borrower quality declined across 2003–2007 cohorts. Comparing the asymptotic and bootstrapped standard errors, that pattern of Table 6 holds that the bootstrapped standard errors are greatly affected by the inclusion of micro-level covariates as controls. The bootstrapped standard errors in columns 1–3 are often an order of magnitude larger than the corresponding asymptotic ones, while the bootstrapped standard errors of columns 4–6 are an average of only 29% higher than the asymptotic standard errors.

In summary, this section was concerned with determining how much of the pattern across origination cohorts in default rates was due to differences in the observed characteristics of mortgage borrowers in each cohort—both the creditworthiness of the individual borrowers and the characteristics of their mortgages—and differences in their exposure to price declines. The results confirm that prices and mortgage characteristics are both are important, with price changes causally explaining at least 60% of the increase in cohort default rates.

6   Mechanisms

I now turn towards identifying the causal mechanisms through which prices affect default rates by testing for negative equity having a causal impact on defaults and whether this explains cohort heterogeneity. The intuition offered above centers around the differential effect of price declines on later cohorts in pushing them underwater, as seen in Figure 2. Mortgage borrowers who are underwater have elevated default risk. Distressed borrowers (i.e. borrowers unable to make their monthly mortgage payments) who have positive equity have two main alternatives to default. First, if interest rates have gone down or if borrowers qualify for a lower interest rate because they have more equity from paid down mortgage principal and accumulated price appreciation, they can refinance into a mortgage with a lower monthly payment, using the new mortgage to repay the

original one.[45] Second, they can sell their home and use the proceeds to pay off their outstanding mortgage debt and move into a more affordable housing situation. Neither of these options is readily available to distressed borrowers who are underwater. Lenders are normally unwilling to originate a refinance mortgage to someone who has zero equity, let alone negative equity. Selling a house secured by a mortgage in a negative equity position (known as a short sale) requires either coming up with sufficient cash to pay the shortfall between the sale price and the outstanding debt or working with the lender to secure forgiveness of the remaining debt. By definition, distressed borrowers are unlikely to have ample savings, making the former unlikely. Lenders are also wary of agreeing to short sales, partly because of asymmetric information about the borrower's current and future finances. An additional source of elevated default risk comes from the possibility that underwater borrowers will default strategically.[46]

Empirically testing that the reason price declines explain the bulk of cohort heterogeneity is through the prevalence of negative equity presents several practical challenges. First, the extent to which borrowers are current with their monthly payments is related to their unobserved quality. I instrument for the actual balance of the mortgage with the scheduled balance calculated using the origination interest rate as if the borrower had paid back a 30-year fixed-rate mortgage on schedule. Second, constructing a measure of negative equity status requires knowing the current market value of the home, an unknown (and endogenous) quantity that must be estimated by the borrower as well as the econometrician. CoreLogic provides such a measure using their Automated Valuation Model that imputes property values in each month for each subprime mortgage in the data. As this estimated value is partly a function of nearby market prices and therefore affected by CBSA-level shocks, I instrument for this valuation using the origination loan amount and counterfactual price indices computed using the historical volatility instruments. Third, because the prepayment obstacles faced by borrowers depend on the total debt of all loans secured against their home, measuring negative equity necessitates knowing updated information about additional liens. Not observing updated information on the outstanding balance of additional liens, I assume that all second mortgages have not been paid down. Although this introduces additional measurement error into the estimated balances, which are already affected by local public records access policies,

---

[45]Relatedly, a home equity line of credit can also be used to borrow additional funds secured by unrealized capital gains. These funds can be used to temporarily make mortgage payments.

[46]Strategic default is when a borrower who has available cash flow to make mortgage payments defaults anyway, exercising a put option on the property. This is optimal if the option value of holding onto the property (i.e. expected future price appreciation) is lower than value of discharging the debt, net any cost of defaulting (see Foote et al., 2008). In other words, borrowers may find it advantageous to default if they do not expect future prices to rise quickly enough.

instrumenting for outstanding balances using scheduled balances solves this problem.

I define the variable $Underwater_{it}$ for whether the current CLTV of a loan, estimated by Core-Logic based on the outstanding debt owed to all outstanding liens and contemporaneous market conditions, is greater than 100%. I first present estimates that do not account for the endogeneity of CLTV. Table 8 contains default hazard specifications of the form above, replacing $\Delta Prices$ with functions of CLTV

$$X'_{icgt}\beta = \gamma_c + W'_{B,i}\theta_B + W'_{L,i}\theta_L + \eta' q(CLTV_{it}) + \alpha_g. \tag{14}$$

Controlling for $q(CLTV) = 1(CLTV > 1) = Underwater$ in addition to loan and borrower characteristics and CBSA fixed effects in column 1 of Table 8 shows that underwater mortgages have more than double the conditional default probability of mortgages that are not underwater. There is substantial unexplained cohort heterogeneity in column 1—even after adjusting for location, mortgage age, borrower and loan characteristics, and the estimated negative equity status, differences in cohort default rates relative to the 2003 cohort are all positive and significant, with the exception of the 2007 cohort. Compared with column 7 of Table 3, the underwater indicator variable explains much less cohort heterogeneity than the 12-month change in prices. Columns 2 of Table 8 tests whether this is driven by the functional form restriction on $q(\cdot)$ by controlling for a linear spline in the current CLTV that allows for a location and scale shift in the effect of $CLTV$ in several bins:

$$q(CLTV_{it}) = \sum_{j=1}^{J} 1(CLTV_{it} \in \mathcal{C}_j) \times (a_j + b_j CLTV_{it}) \tag{15}$$

where j indexes the set $\mathcal{C}$ consisting of J CLTV intervals $\{[0, 80), [80, 85), [85, 90), ..., [150, \infty)\}$. Adding flexibility in the specification of the leverage function $q(\cdot)$ further decreases the adjusted differences across cohorts but only explains on average an additional 8% of the differences in the latent quality of cohorts. The specification in column 2 explains 62% of the cohort-level differences in default.

To see whether prices still explain cohort heterogeneity even conditional on underwater, i.e. to test whether the effect of prices is driven entirely by negative equity, column 3 additionally controls for the twelve-month change in log HPI. Adding in the price change variable in addition to the linear spline controls significantly affects the estimated cohort heterogeneity relative to column 2 but also relative to column 7 of Table 3, which is identical to column 3 except for the inclusion of $q(CLTV)$. This suggests that CLTVs and prices interact in explaining defaults. Controlling for both price changes and current CLTVs reduces the 2006 and 2007 cohort differences to be strongly

negative—controlling for a flexible function of their relative equity, the price changes they faced, and loan and borrower characteristics, 2006–2007 borrowers defaulted less than would be expected. The estimated latent quality of the 2004–2005 cohorts positive and significant, with the 2005 estimate smaller and the 2004 result larger than the results in Table 3 that do not control for current CLTVs. The coefficient on the price change variable is large and significant.

There is a large relationship between defaults and negative equity and evidence that prices also affect defaults in other ways than through negative equity. Still, caution is required interpreting these results because mark-to-market leverage (CLTV) could be correlated with unobserved borrower quality. I now discuss an instrumental-variables strategy to account for this endogeneity.

6.1 Instrumenting for Loan-to-Value Ratios

The main obstacle in interpreting the results in columns 1–3 of Table 8 is the endogeneity of CLTVs, which are the ratio of loan principal balances and property values. To the extent that borrowers whose unobserved quality is low (high) pay back their mortgages more slowly (rapidly), loan balances (and hence CLTVs) will be determined in part by unobserved borrower quality. Similarly, borrowers with lower unobserved quality may take out mortgages with slow amortization schedules that leave them more likely to be underwater. To address the endogeneity of CLTV numerators, I calculate the scheduled loan principal amount at each month since origination if borrower had taken out a 30-year fixed interest rate loan with same origination interest rate and purchase price and was current on all payments time.[47] Using the amortization formula,

$$Scheduled\ Principal_{it} = M_i \left((1 + r_i)^t - \frac{(1 + r_i)^{360} \left((1 + r_i)^t - 1 \right)}{(1 + r_i)^{360} - 1} \right)$$

where M_i is the purchase price of property i, t is the loan age in months, and r_i is the origination interest rate divided by 12.

To account for the endogeneity in purchase prices, I compute what the purchase price of the home would have been if the borrower had taken out only a first-mortgage for the same dollar amount at the conforming loan limit (80% of purchase price). In logs, using this predicted purchase price is equivalent to using the log of the origination amount as an instrument. Finally, with the predicted home price indices, I can calculate an alternative measure of the change in a property's

---

[47]Cunningham and Reed (2013) refer to this as a synthetic mortgage IV strategy.

value since origination using the predicted HPI series

$$Appr\widehat{eciation}_{igt} = \log \widehat{HPI}_{igt} - \log \widehat{HPI}_{ig1}$$

where $\log \widehat{HPI}$ are the predicted values from estimating equation (12).

Before presenting first stage results, in FIgure 9 I illustrate graphically the statistical relationship between each of the three instruments and the corresponding component of CLTVs. Diagonal lines depict the fitted bivariate linear regression line. Panel I plots actual log principal balances versus scheduled log principal balances. The fit is very strong and the slope of the bivariate regression line is close to 1, showing the tight relationship between traditional amortization schedules and loan balances. The most noticeable deviation is the presence of many outliers well below the regression line, representing people that paid their mortgages back faster than scheduled. Instrumenting will address the possibility that their faster payback is a signal of these borrowers' unobserved (high) quality. Panel II plots actual log sale prices against log origination amounts. The average relationship between origination balances and actual sale prices is not far off from a setting where all borrowers took out mortgages at 80% of the sale price of the home, in which case there would be a perfect fit between log origination amount and log sale price with an intercept of log(1.25) and a slope of 1. The most obvious outliers are those well above the regression line—borrowers who took out mortgages with much lower leverage (i.e. through a larger downpayment in the case of sales or from accumulated equity in the case of refinances). Using log origination amounts as an instrument to explain CLTVs will account for any correlation between actual sale prices, initial leverage, and unobserved borrower quality. Panel III plots assessed property values against counterfactual property values

$$\widehat{Value} = 1.25 \times Origination\ Amount \times \exp(Appr\widehat{eciation})$$

to show the predictive power of the generated instrument $Appr\widehat{eciation}$. The workhorse behind this relationship is the long-run price cyclicality instrument σ_g^P used to predict HPI values and subsequently impute appreciation-since-origination and corresponding counterfactual property values. There is a clear positive relationship between counterfactual property values and assessed values. Positive deviations from the regression line represent homes in areas and months with much higher prices than would be predicted based on the 1980s price cycle of that city. Negative deviations represent homes where price declines have been more acute than expected a priori. Instrumenting for actual assessed values will address the potential for these price changes to be correlated with

unobserved borrower default risk.

The first stage for CLTV is a linear regression of CLTV on the scheduled loan balance, the loan origination amount, predicted appreciation using the counterfactual price series, and the usual controls Z_2

$$CLTV_{icgt} = Z'_{1,igt}\Upsilon_1 + Z'_{2,icgt}\Upsilon_2 + \nu_{icgt} \tag{16}$$

where the instrument set consists of

$$Z_{1,igt} = \left(\begin{array}{ccc} \log(Scheduled\,Principal_{it}) & \log(Origination\,Amount_i) & \widehat{Appreciation}_{igt} \end{array} \right).$$

Table 9 reports the results of estimating equation (16) by OLS with clustered standard errors. Note that missing data—loans for which CoreLogic has not estimated a contemporaneous CLTV in a given month—reduces the sample size of specifications involving CLTV from 1.2 to 1.0 million monthly loan observations. Column 1 reports results of regressing $CLTV$ on Z_1 without controlling for Z_2. The relationship between each of the instruments and CLTV values is large and very precisely estimated. Mortgages with higher origination amounts (positive predictors of sale prices) have lower CLTVs. Mortgages with higher scheduled principal balances have higher CLTVs. Mortgages with higher predicted appreciation have lower CLTVs. Adding cohort indicator variables, baseline hazard controls, and CBSA fixed effects in column 2 strengthens the estimated effect of origination amounts and scheduled principal and attenuates the effect of predicted appreciation on the CoreLogic contemporaneous CLTVs. The cohort pattern confirms the trends in median CLTVs plotted in Figure 2, showing that later cohorts have much higher CLTVs. Successively controlling for borrower and loan characteristics in column 3 and price changes in column 4 continues the trend. The instruments are still powerful predictors of CLTVs. Column 5 additionally controls for the monthly CBSA unemployment rate. Local labor market conditions are clearly correlated with CLTVs: the coefficient on the unemployment rate suggests that the equity share of property values in areas with high unemployment rate is lower. Controlling for the unemployment rate, the predicted appreciation instrument is no longer significant. Still, the partial F-statistic for the joint significance of the instruments is above 200 in every column.

Columns 4–6 of Table 8 report the results of estimating the default hazard function after incorporating $\hat{\nu}_{icgt}$ from equation (16) into the linear index $X\beta$ in equation (14).[48] Column 4 includes

---

[48]Imbens and Wooldridge (2007) discuss the control function approach when the estimating equation contains several non-linear functions of the right-hand side endogenous variable. Under the assumption that the unobserved component of default risk is independent of the instruments (the control function exclusion restriction), controlling for the fitted residuals of CLTV is sufficient to instrument for any function of CLTV.

the underwater indicator variable as a parsimonious summary of the causal influence of negative equity on default conditional on the CLTV residuals, price changes, and price change residuals. Columns 5 and 6 instead control for a linear spline in q. The estimated effect of prices is large and significant across all specifications, showing an elasticity of default with respect to price declines of –3 to –5, meaning that for a fixed CLTV, a 1% price decline increases the default hazard by 3–5%. The effect of being underwater on default is still significant but greatly attenuated from column 1, suggesting that holding prices fixed, a mortgage being underwater *causes* the default hazard to be 33% higher (28 log points) than that of above-water mortgages. This suggests that some of the performance differences across cohorts that columns 1–3 attributed to negative equity were actually unobserved differences in borrower quality across cohorts that affected both defaults and equity. Indeed, the CLTV residuals are significant in columns 4–6, rejecting the null hypothesis that CLTVs are exogenous.

Comparing columns 3 and 5, the estimated cohort differences after controlling and instrumenting for price changes and mark-to-market leverage are slightly smaller than the corresponding estimates column 3 that do not account for the endogeneity of prices or CLTVs. Continuing a trend in my findings, the specification in column 5 is more successful at explaining the default rates of later cohorts than earlier cohorts, suggesting that that negative equity was a more important factor in late-cohort defaults than early cohort defaults. While highly predictive of individual defaults, the smaller effect of CLTV controls on earlier cohort default rates is consistent with earlier cohorts' CLTVs not having increased as much (see Figure 2).

Because local labor market fluctuations are not excludable from my instrument, I control directly for the unemployment rate in column 6 of Table 8. As in Table 6, conditional on all of the other controls, mortgages in cities with increased unemployment rates are slightly less likely to default—a one percentage point increase in the local unemployment rate decreases the default hazard by 5%. Accounting for local labor market fluctuations does not materially affect the estimated coefficients on prices or CLTV residuals. However, including the unemployment rate decreases the measure of the difference in latent quality between the 2003 and 2004–2005 cohorts enough to be statistically insignificant.

Taken together, the results of Table 8 provide several explanations about the mechanisms through which price declines cause defaults, decomposing cohort-level differences in default rates into four factors: borrower and loan characteristics, price declines, and local economic conditions. Negative equity is a prominent channel and explains much of the the relationship between cohort

default rates and price changes, especially among later borrower cohorts. Another important factor is unemployment, which may cause and be caused by price declines (Mian and Sufi, 2012). Nevertheless, prices affect default risk in other ways besides their effect on default through equity and their correlation with local economic conditions. The role of price expectations is one likely explanation for prices having such a strong relationship with default even conditional on negative equity. If buyers' expectations of future prices are correlated with recent price changes, even above-water borrowers wishing to sell in areas experiencing recent price declines may be unable to in the face of a thin market of patient buyers. Still, there is strong evidence that negative equity is responsible for much of the effect of prices on defaults and that the differential prevalence of negative equity across cohorts explains a significant portion of the observed increase in cohort-level default rates.

7   Estimating Counterfactual Default Rates

Using the control function specification estimated in column 5 of Table 6 as my preferred specification, I calculate average default rates for each cohort using counterfactual explanatory variables as an estimate of the impact of the price and mortgage characteristics channels. Using the estimated coefficients, predicted values $\hat{h}$ are an estimate of the probability each loan defaulted for each month it existed. By equation (4),

$$\hat{h}_{icgt} = 1 - \exp(-\exp(X'_{icgt}\hat{\beta} + \hat{\psi}_t)$$

where $\hat{\psi}_t$ are nonparametric estimates of the log baseline hazard function between time $t-1$ and t as discussed in Section 4.1. I aggregate these individual default probabilities to calculate cohort-level average default rates, which I annualize multiplying by twelve. The predicted average annual default rate for cohort c is then defined as

$$\widehat{\text{Default Rate}}_c = \frac{12}{N_c} \times \sum_{t \leq 60} \hat{h}_{icgt}$$

where N_c is the number of cohort-c monthly loan observations in the sample of loans within five years of origination. I limit the sample to observations on loans within five years of origination to facilitate comparisons across cohorts. Because I define default to occur the first month that a mortgage is marked as in foreclosure or real-estate owned, this rate is similar to the average number of foreclosure starts in each month divided by the number of loans that were extant during that month.

Table 10 shows the counterfactual default rates for eight scenarios, each representing a differ-

ent combination of counterfactual price paths and loan characteristics. The first row reports the actual default rates for each cohort. The actual spread between the default rates of 2003 and 2006 mortgages was 8.2 percentage points. The model's predicted default rates using observed covariates (not shown) match the actual default rates to 2–3 decimal places, suggesting that this parsimonious model fits the data quite well. Rows 2 and 3 estimate the average default rates that would have prevailed if all mortgages had the characteristics reported in Table 2 of the average 2003 (row 2) or 2006 (row 3) mortgage. Default rates would have been lower if the characteristics of mortgages had not changed over time, especially for later borrower cohorts. If all borrowers had taken out the average 2006 mortgage, row 3 shows that default rates would have been roughly one percentage point higher for 2003–2005 cohorts and lower for 2006–2007 cohorts. The spread between the 2003 and 2006 cohorts is cut in half by fixing mortgage characteristics. However, even if the composition of mortgage products did not change from 2003–2006—a conceptual upper bound on the effect of stricter mortgage regulation, the 2006 cohort would have still defaulted 3.7 percentage points more frequently than the 2003 cohort.[49]

The remaining rows experiment with counterfactual price paths. Rows 4–6 use actual individual loan characteristics and three alternative price scenarios. Row 4 assigns each loan to have the average price change that 2003-cohort loans faced at the same number of months since origination. Row 5 does the same exercise using the prices to which 2006-cohort loans were exposed, and row 6 looks at the effect of flat prices—0% price growth over the life of the mortgage. As expected, mortgages from every cohort would have defaulted much less if they had experienced several years of rapid price appreciation, as did 2003-cohort mortgages. Ceteris paribus, if 2006-cohort mortgages had faced the same prices that the average 2003 mortgage did, their default rate would have been 5.6% instead of 12%. Similarly, if 2003-cohort mortgages had faced the prices that the average 2006 mortgage faced, their default rate would have been 8.5% instead of 4.2%. The counterfactual default rates for the scenario in which there is no price growth is predictably in between the 2003 and 2006 price scenarios. The spread between the 2003 and 2006 cohorts seen in row 1 is mostly gone in rows 4–6, showing that if they had faced the same prices, the 2006 cohort default rate would have been at most 2.5 percentage points higher than the 2003 default rate.

The final two rows report default rates under the counterfactual of constant prices and mortgage characteristics. The combination of fixed prices and mortgage characteristics explains the entire

---

[49]Note that this statement assumes that holding mortgage product characteristics fixed would not have affected aggregate prices.

difference in the unadjusted cohort default rates of column 1, with the 2006 cohort predicted to *outperform* the 2003 cohort if both had faced the same (zero) price growth and had taken out mortgages with the characteristics of either the average 2003-cohort mortgage (row 7) or the average 2006-cohort mortgage (row 8). As a measure of the latent quality of each of these cohorts, rows 7 and 8 suggest that there were no important declines in unobserved borrower quality across subprime cohorts. Using the zero price growth scenarios as a benchmark, it seems that the low and high actual default rates experienced by the 2003 and 2006 cohorts, respectively, were not particularly representative of the relative quality of these cohorts. Intuitively, this makes sense—by historical standards, neither the 2003 nor 2006 price paths seem to have been particularly normal.

8 Conclusion

There has been an active debate about the surge in the subprime default rate in the mid- to late-2000s, with blame being placed on risky mortgage products, risky borrowers, and price declines. The accompanying analysis has focused on contrasting the relative performance of late and early cohorts to tease out these stories. Diverse views of the cause of this deterioration in cohort-level mortgage outcomes have motivated strong opinions about the appropriate regulator response to the subprime default crisis. Advocates of stricter mortgage lending regulation argue that the cohort pattern represents a deterioration in underwriting standards over time, i.e. the lending of riskier mortgage products to riskier borrowers, and that these looser standards were the main precipitating factor in the crash.

This paper demonstrates why the cohort comparison is potentially misleading: cohorts may differ not only in their composition (loan and borrower characteristics) but crucially in the degree to which they were affected by price fluctuations. I ascertain the relative contribution of each of these factors by combining observable loan and borrower characteristics with data on price changes in a model that explains 95% of the heterogeneity in cohort performance. Decomposing the observed deterioration in subprime loan performance, I find that the differential impact of the price cycle on later cohorts explains 60% of the rapid rise in default rates across subprime borrower cohorts. Loan characteristics, especially whether the mortgage had an interest-only period or was not fully amortizing, are important as well and explain 30% of the observed default rate differences across cohorts. Changing borrower characteristics, on the other hand, had little detectable effect on cohort outcomes. While quite predictive of individual default, borrower characteristics simply did not

change enough across cohorts to explain the increase in defaults.

The results of this paper suggest a scope for underwriting standards to ex-ante affect mortgage outcomes and for ex-post programs such as principal reduction and loan modifications that reduce the frictions associated with being underwater.[50] Such policies may interact: all else equal, mortgages with lower origination CLTVs are less sensitive to price declines. Nevertheless, the view that borrower quality declined across subprime cohorts on unobservable dimensions is inconsistent with the results of this paper. I find that if 2006 borrowers had faced the prices that the average 2003 borrower did, 2006 borrowers would have had an annual default rate of 5.6% instead of 12%. I conclude that a 2003 borrower taking out the average 2006 mortgage in 2006 would be no less likely to default than a 2006 borrower in the same circumstances.

---

[50]Of course, the presence of ex-post remedies may induce moral hazard. See Mayer and Hubbard (2009), Wheaton (2010), Feldstein (2011), and the enacted Home Affordable Modification Program for examples of loan modification programs and proposals, many designed to preserve incentives for responsible borrowing and maintenance.

References

Allison, Paul D, "Discrete-time methods for the analysis of event histories," *Sociological methodology*, 1982, *13* (1), 61–98.

Bajari, Patrick, Chenghuan Sean Chu, and Minjung Park, "An Empirical Model of Subprime Mortgage Default From 2000 to 2007," December 2008. National Bureau of Economic Research Working Paper #14625.

Berger, Allen N and Gregory F Udell, "The institutional memory hypothesis and the procyclicality of bank lending behavior," *Journal of Financial Intermediation*, 2004, *13* (4), 458–495.

Bhardwaj, Geetesh and Rajdeep Sengupta, "Subprime Loan Quality," 2011. Federal Reserve Bank of St. Louis Working Paper # 2008-036E.

___ **and** ___ , "Subprime mortgage design," *Journal of Banking & Finance*, 2012, *36* (5), 1503–1519.

Bhutta, Neil, Jane Dokko, and Hui Shan, "The Depth of Negative Equity and Mortgage Default Decisions," 2010. Finance and Economics Discussion Series Working Paper #2010-35.

Board of Governors of the Federal Reserve System, "Report to the Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit," Technical Report, http://www.federalreserve.gov/boarddocs/RptCongress/creditscore/creditscore.pdf August 2007.

Brueckner, Jan K., Paul S. Calem, and Leonard I. Nakamura, "Subprime mortgages and the housing bubble," *Journal of Urban Economics*, 2012, *71* (2), 230 – 243.

Campbell, John, Stefano Giglio, and Parag Pathak, "Forced Sales and House Prices," *American Economic Review*, 2011, *101(5)*, 2108–2131.

(COP) Congressional Oversight Panel of the Troubled Assets Recovery Program, "Foreclosure Crisis: Working toward a Solution," Technical Report, http://cop.senate.gov/reports/library/report-030609-cop.cfm COP Report, March 6, 2009.

Corbae, Dean and Erwan Quintin, "Leverage and the Foreclosure Crisis," August 2013. National Bureau of Economic Research Working Paper #19323.

Cunningham, Chris and Robert R Reed, "Negative equity and wages," *Regional Science and Urban Economics*, 2013, *43* (6), 817–825.

Dagher, Jihad and Ning Fu, "What Fuels the Boom Drives the Bust: Regulation and the Mortgage Crisis," 2011. IMF Working Paper WP/11/215.

Dell'Ariccia, Giovanni, Deniz Igan, and Luc Laeven, "Credit booms and lending standards: Evidence from the subprime mortgage market," *Journal of Money, Credit and Banking*, 2012, *44* (2-3), 367–384.

Demyanyk, Yuliya and Otto Van Hemert, "Understanding the subprime mortgage crisis," *Review of Financial Studies*, 2011, *24* (6), 1848–1880.

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

Elbers, Chris and Geert Ridder, "True and spurious duration dependence: The identifiability of the proportional hazard model," *The Review of Economic Studies*, 1982, *49* (3), 403–409.

Feldstein, Martin S., "How to Stop the Drop in Home Values," *The New York Times*, October 12, 2011.

Foote, Christopher L, Kristopher S Gerardi, and Paul S Willen, "Negative equity and foreclosure: Theory and evidence," *Journal of Urban Economics*, 2008, *64* (2), 234–245.

—— , —— , **Lorenz Goette, and Paul S Willen**, "Reducing Foreclosures: No Easy Answers," in Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, eds., *NBER Macroeconomics Annual 2009, Volume 24*, University of Chicago Press, April 2010, pp. 89–138.

Genesove, David and Christopher J Mayer, "Equity and Time to Sale in the Real Estate Market," *The American Economic Review*, 1997, pp. 255–269.

Gerardi, Kristopher, Adam Shapiro, and Paul Willen, "Subprime outcomes: Risky mortgages, homeownership experiences, and foreclosures," 2007. FRB of Boston Working Paper No. 07-15.

—— , **Andreas Lehnert, Shane M Sherlund, and Paul Willen**, "Making sense of the subprime crisis," *Brookings Papers on Economic Activity*, 2008, (2), 69–159.

Government Accountability Office, "Nonprime Mortgages: Analysis of Loan Performance, Factors Associated with Defaults, and Data Sources," Technical Report 2010.

Han, Aaron and Jerry A Hausman, "Flexible parametric estimation of duration and competing risk models," *Journal of Applied Econometrics*, 1990, *5* (1), 1–28.

Hartley, Dan, "The Effect of Foreclosures on Nearby Housing Prices: Supply or Disamenity?," 2010. FRB of Cleveland Working Paper No. 10-11.

Haughwout, Andrew, Richard Peach, and Joseph Tracy, "Juvenile delinquent mortgages: Bad credit or bad economy?," *Journal of Urban Economics*, 2008, *64* (2), 246 – 257.

Hausman, Jerry A and Tiemen M Woutersen, "Estimating a semi-parametric duration model without specifying heterogeneity," 2010. Working Paper, The Johns Hopkins University, Department of Economics.

Heckman, James J and Bo E Honoré, "The identifiability of the competing risks model," *Biometrika*, 1989, *76* (2), 325–330.

—— **and Burton Singer**, "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica: Journal of the Econometric Society*, 1984, pp. 271–320.

—— **and Richard Robb**, "Alternative methods for evaluating the impact of interventions: An overview," *Journal of Econometrics*, 1985, *30* (1), 239–267.

Horowitz, Joel L, "Semiparametric estimation of a proportional hazard model with unobserved heterogeneity," *Econometrica*, 1999, *67* (5), 1001–1028.

Hubbard, R Glenn and Christopher J Mayer, "The Mortgage Market Meltdown and House Prices," *The B.E. Journal of Economic Analysis & Policy*, 2009, *9* (3).

Imbens, Guido and Jeffrey M Wooldridge, "Control Function and Related Methods," What's New in Econometrics, Summer 2007.

Imbens, Guido W, "Nonadditive Models with Endogenous Regressors," in R. Blundell, W. Newey, and T. Persson, eds., *Advances in Economics and Econometrics*, Vol. III, Cambridge University Press, 2007, chapter 2, pp. 17–46.

(JEC) Joint Economic Committee, "The Subprime Lending Crisis: The Economic Impact on Wealth, Property Values and Tax Revenues, and How We Got Here," Technical Report October 2007.

Kau, James B., Donald C. Keenan, Constantine Lyubimov, and V. Carlos Slawson, "Subprime mortgage default," *Journal of Urban Economics*, 2011, *70* (2–3), 75 – 87.

Keys, Benjamin J, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, "Did securitization lead to lax screening? Evidence from subprime loans," *The Quarterly Journal of Economics*, 2010, *125* (1), 307–362.

Krainer, John and Elizabeth Laderman, "Prepayment and delinquency in the mortgage crisis period," September 2011. Federal Reserve Bank of San Francisco Working Paper 2011-25.

Krugman, Paul, "A Catastrophe Foretold," *The New York Times*, October 26, 2007.

——— , "Some Housing Pictures," *The New York Times*, October 27, 2007.

Lancaster, Tony, "Generalised residuals and heterogeneous duration models: With applications to the Weilbull model," *Journal of Econometrics*, 1985, *28* (1), 155–169.

Mayer, Christopher J., "Comment on "Reducing Foreclosures: New Easy Answers"," in Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, eds., *NBER Macroeconomics Annual 2009, Volume 24*, University of Chicago Press, April 2010, pp. 139–148.

——— **and Karen Pence**, "Subprime mortgages: what, where, and to whom?," 2008. National Bureau of Economic Research Working Paper #National Bureau of Economic Research Working Paper #14083.

——— **and Todd M. Sinai**, "U.S. House Price Dynamics and Behavioral Finance," in Christopher L Foote, Lorenz Goette, and Stephan Meier, eds., *Policymaking Insights from Behavioral Economics*, Boston: Federal Reserve Bank of Boston, 2009, pp. 259–297.

——— **, Karen Pence, and Shane M Sherlund**, "The rise in mortgage defaults," *Journal of Economic Perspectives*, 2009, *23* (1), 27–50.

——— **, Tomasz Piskorski, and Alexei Tchistyi**, "The Inefficiency of Refinancing: Why Prepayment Penalties Are Good for Risky Borrowers," December 2010. National Bureau of Economic Research Working Paper #16586.

McCall, Brian P, "Unemployment insurance rules, joblessness, and part-time work," *Econometrica*, 1996, pp. 647–682.

Meyer, Bruce D, "Unemployment Insurance and Unemployment Spells," *Econometrica*, 1990, *58* (4), 757–782.

Mian, Atif, "Comment on "Reducing Foreclosures: No Easy Answers"," in Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, eds., *NBER Macroeconomics Annual 2009, Volume 24*, University of Chicago Press, April 2010, pp. 149–156.

⎯⎯ , **Amir Sufi, and Francesco Trebbi**, "Foreclosures, House Prices, and the Real Economy," January 2011. National Bureau of Economic Research Working Paper #16685.

⎯⎯ **and** ⎯⎯ , "The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis," *The Quarterly Journal of Economics*, 2009, *124* (4), 1449–1496.

⎯⎯ **and** ⎯⎯ , "What explains high unemployment? The aggregate demand channel," 2012. National Bureau of Economic Research Working Paper #17830.

Murphy, Kevin M. and Robert H. Topel, "Estimation and Inference in Two-Step Econometric Models," *Journal of Business & Economic Statistics*, 1985, *3* (4), 370–79.

Pagan, Adrian, "Econometric issues in the analysis of regressions with generated regressors," *International Economic Review*, 1984, *25* (1), 221–247.

Pavlov, Andrey and Susan Wachter, "Subprime Lending and House Price Volatility," 2009. U of Penn, Inst for Law & Econ Research Paper 08-33.

Pennington-Cross, Anthony and Giang Ho, "The Termination of Subprime Hybrid and Fixed-Rate Mortgages," *Real Estate Economics*, 2010, *38* (3), 399–426.

Prentice, Ross L and Lynn A Gloeckler, "Regression analysis of grouped survival data with application to breast cancer data," *Biometrics*, 1978, pp. 57–67.

Rivers, Douglas and Quang H Vuong, "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 1988, *39* (3), 347–366.

Sherlund, Shane M, "The Past, Present, and Future of Subprime Mortgages," 2008. Finance and Economics Discussion Series Working Paper #2008-63.

Sinai, Todd M., "House Price Moments in Boom-Bust Cycles," May 2012. National Bureau of Economic Research Working Paper #18059.

Wheaton, William C., "A Proposal to Repair US Residential Mortgages." Unpublished mimeo, MIT.

Wooldridge, Jeffrey M, *Econometric analysis of cross section and panel data*, The MIT press, 2002.

Figure 1. Cumulative Default Probability by Origination Cohort

I. Cumulative Default Probability

II. Cumulative Prepayment Probability
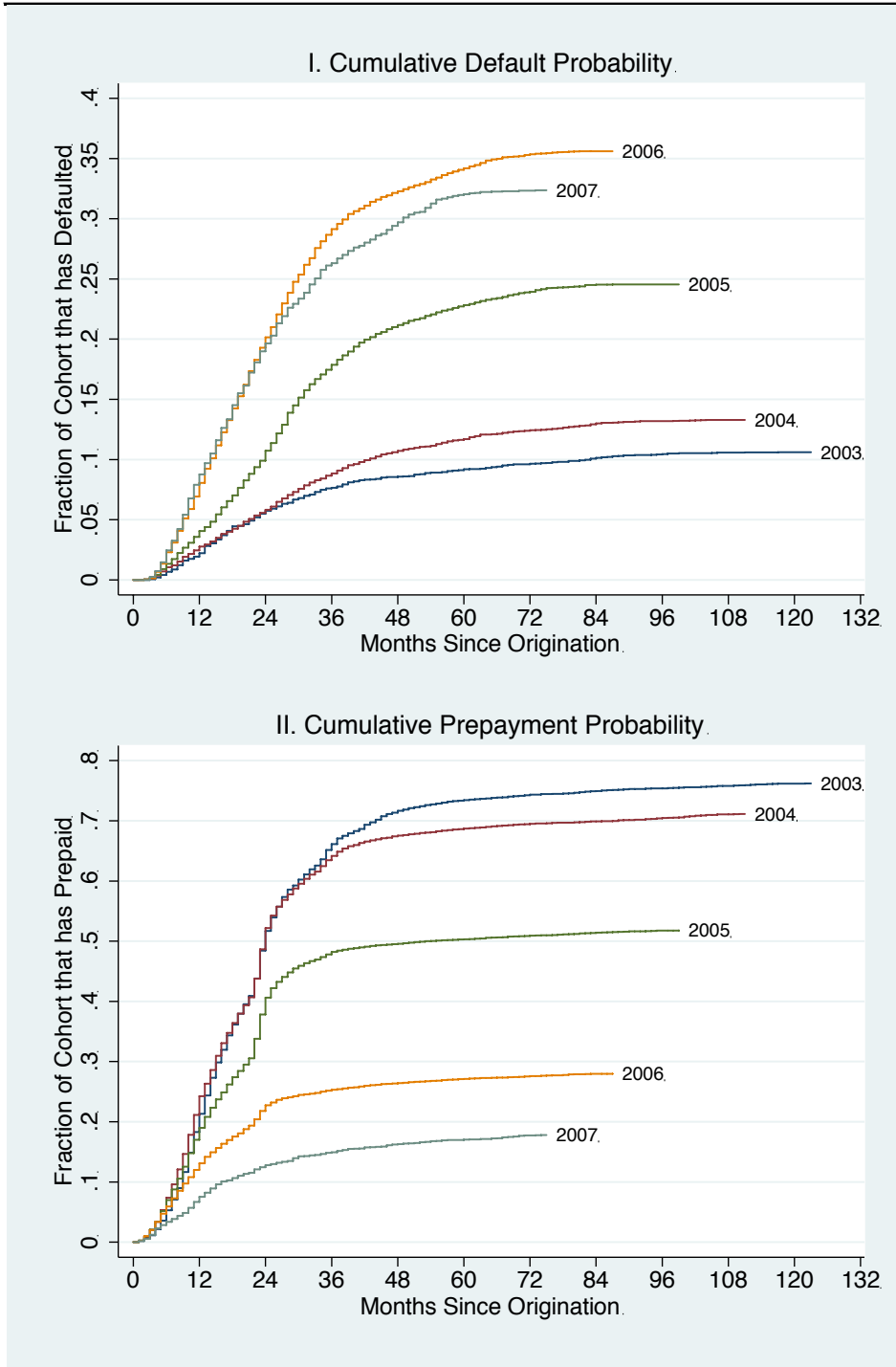
Notes: Figure plots the fraction of each cohort that has terminated by default (left panel) or prepayment (right panel) within a given number of months since origination. Default is measured as the first time that a loan's delinquency status is marked as in foreclosure or real-estate owned provided it ultimately terminated without being paid off in full. Prepayment means repayment in full, i.e. through refinancing or selling.

Figure 2. Median Combined Loan-to-Value Ratio Over Time by Cohort
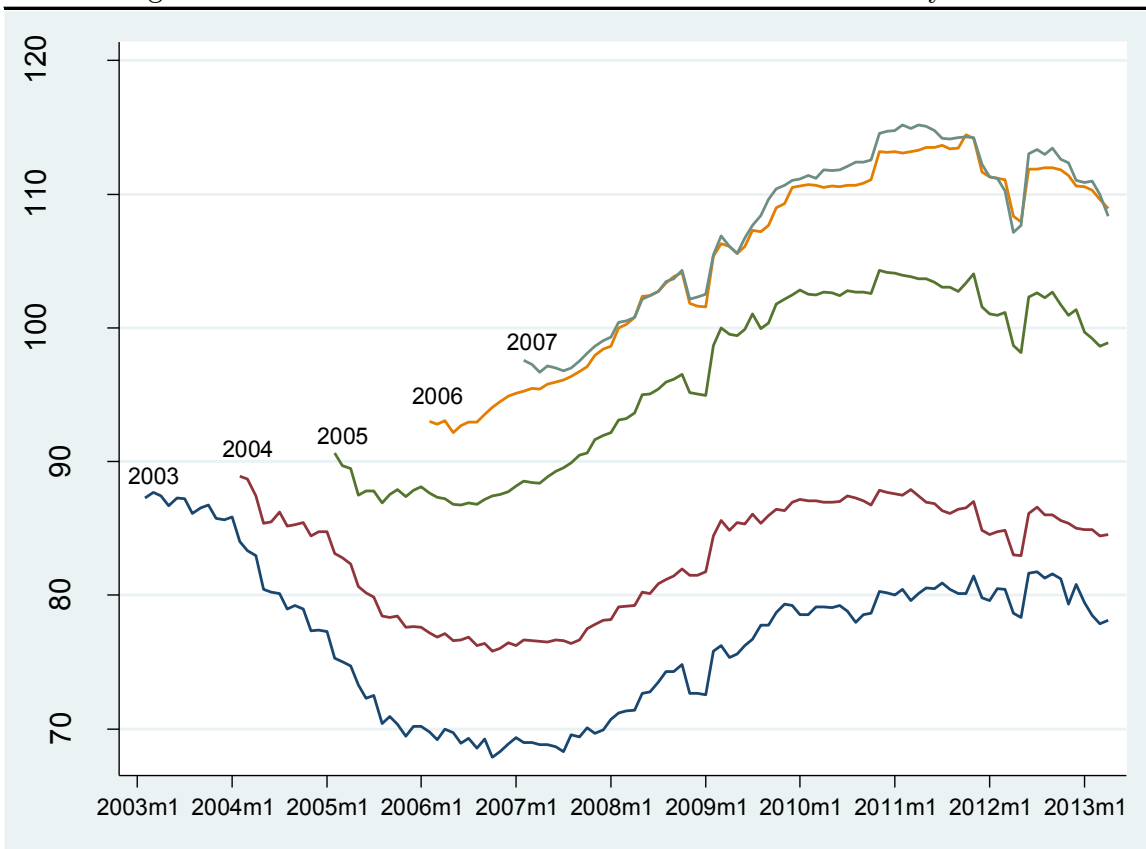
Notes: Figure shows the median current combined loan-to-value ratio (CLTV) of subprime borrowers for existing subprime mortgages in each cohort in each calendar month in percentage points. Current CLTVs are calculated by LoanPerformance as the total outstanding principal on a loan divided by an automated assessing model's estimate of the market value of each home.

Figure 3. Prices and Negative Equity Prevalence:
Pittsburgh vs. Minneapolis
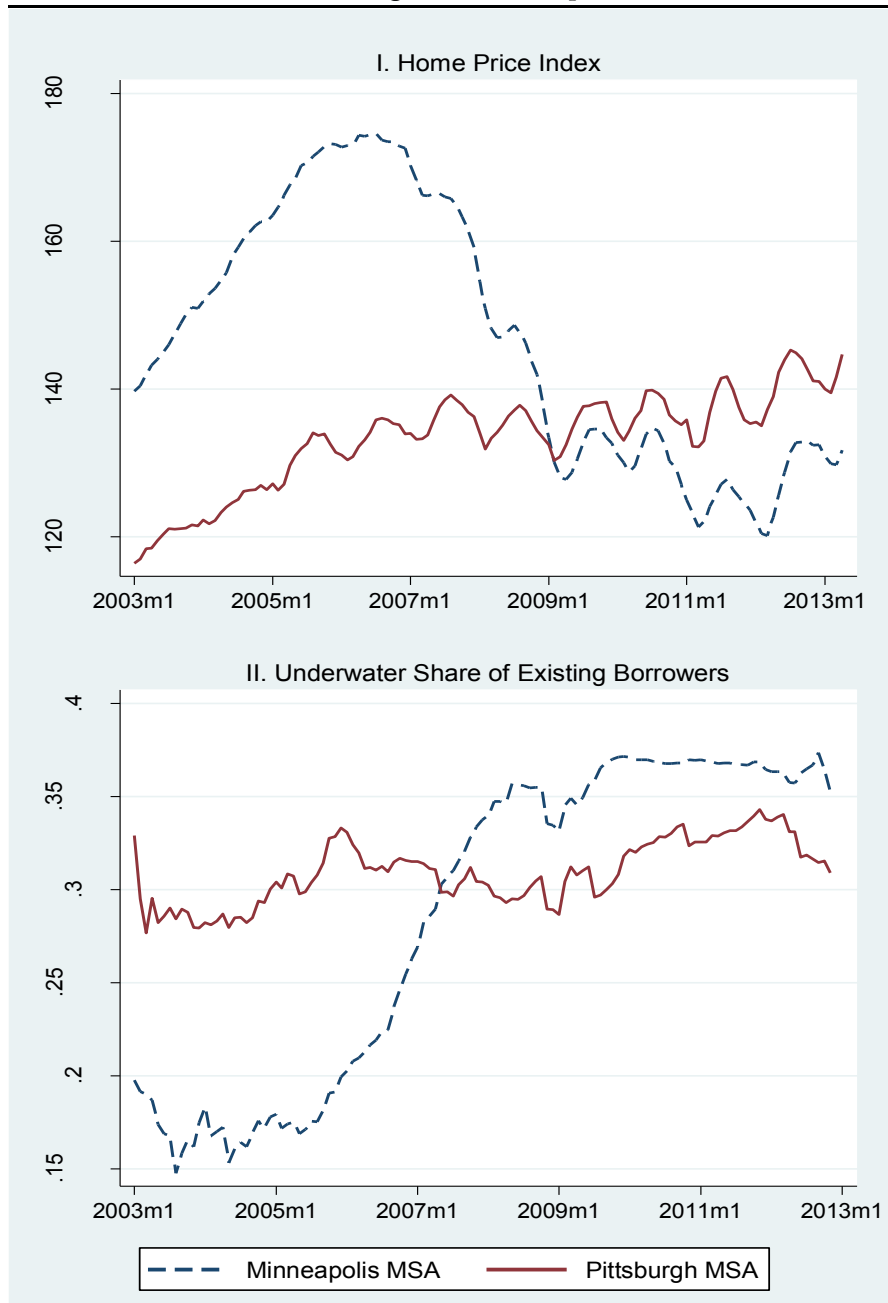
Notes: Top panel shows the CoreLogic repeat-sales home price index
for the Minneapolis and Pittsburgh Metropolitan Statistical Areas
(MSA). Both series have been normalized to 100 in January 2000.
Bottom panel shows shows the fraction of all outstanding subprime
borrowers that were underwater in each calendar month in the
indicated MSA. Underwater is determined by the current combined
loan-to-value ratio (CLTV) for a loan being above 100%. CLTVs
are calculated by CoreLogic as the total outstanding principal on a
loan divided by an automated assessing model's estimate of the

Figure 4. Cumulative Default Probabilities:
Pittsburgh vs. Minneapolis

I. Pittsburgh
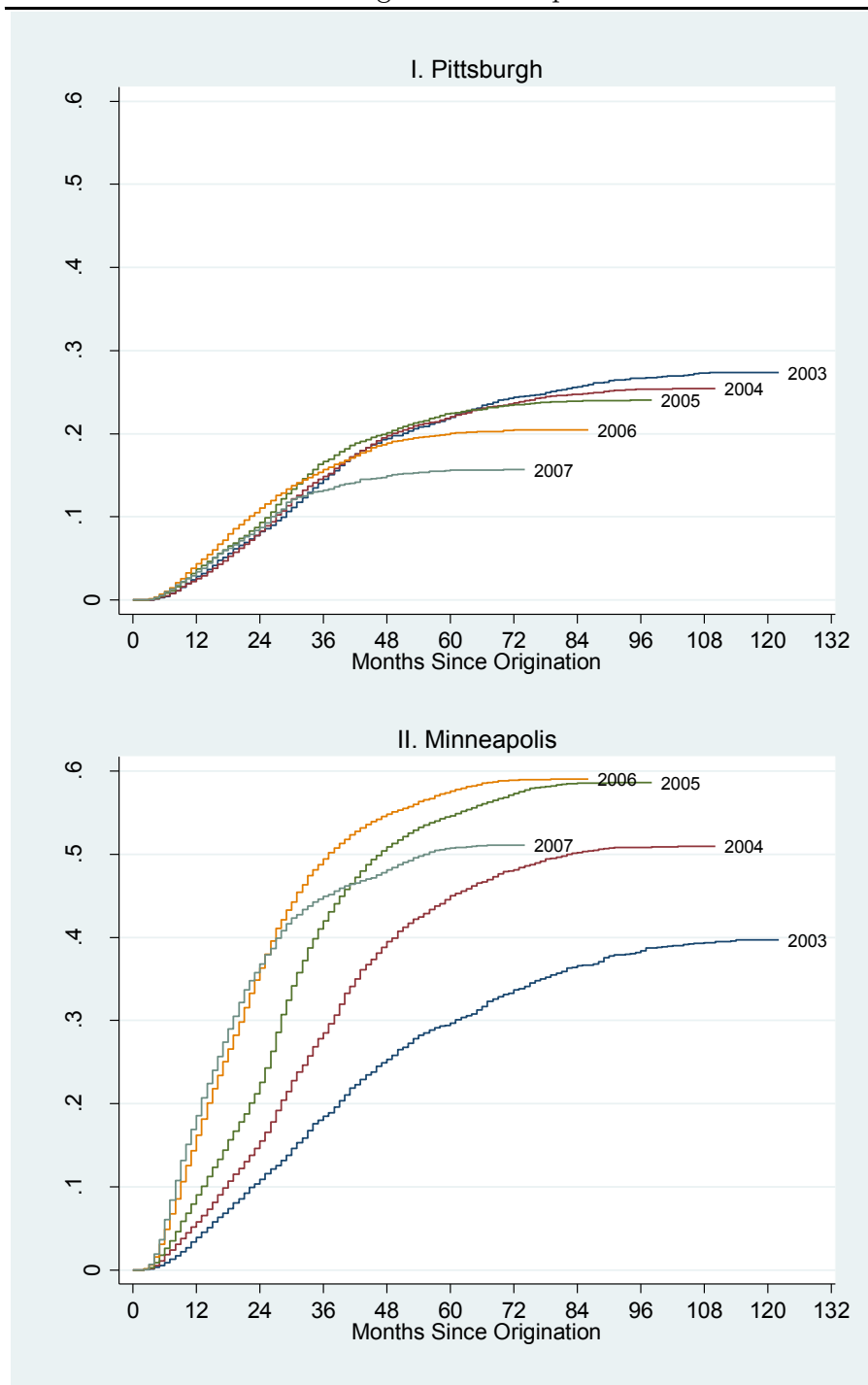
II. Minneapolis

Notes: Graphs show the cumualtive default probability of each cohort in the Pittsburgh and Minneapolis CBSAs, respectively. Each line shows the fraction of that cohort that had defaulted within a given number of months since origination. See Figure 1 notes for more details.

Figure 5. Persistence of House Price Cyclicality:
Average Home Price Index by Quartile of σ^P
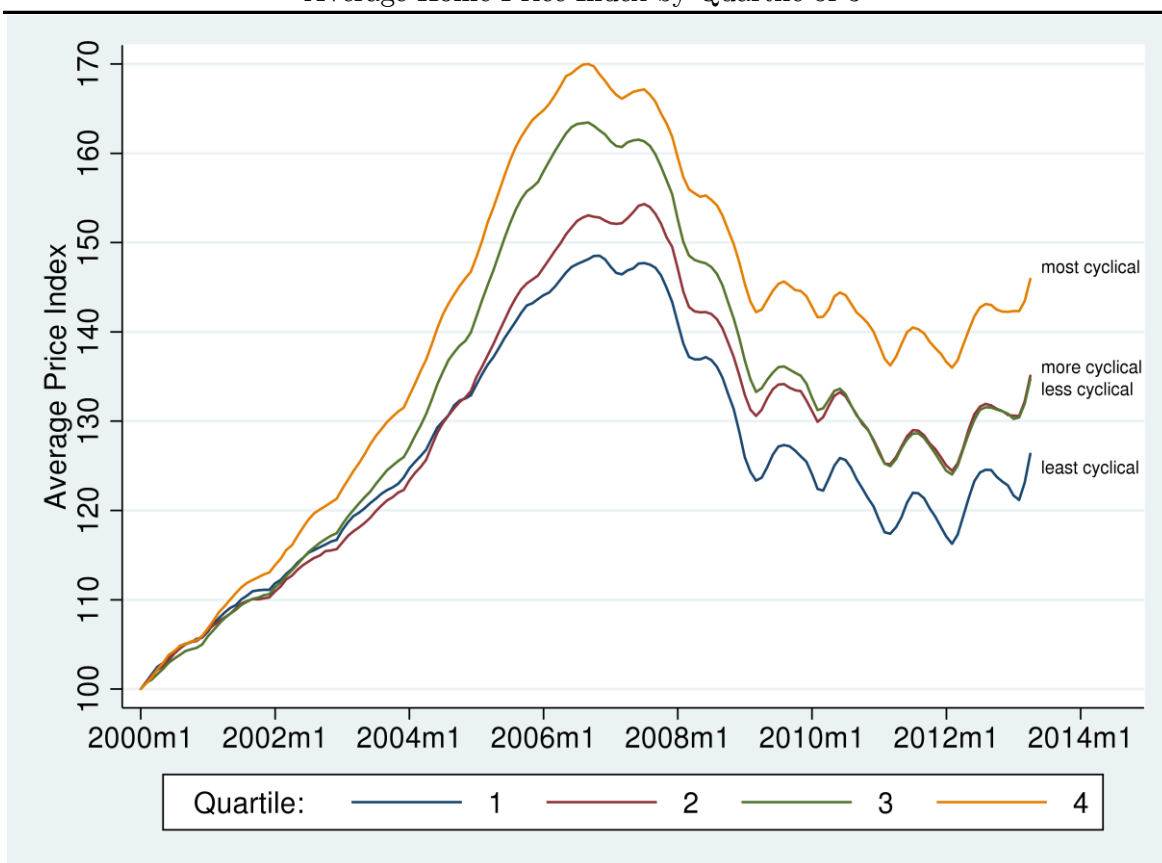
Notes: Figure plots month average HPI values by cyclicality quartile. Cyclicality is measured as the standard deviation of one month changes to the log home price index from 1980-1995, as defined in equation (5) in the text. Each series has been normalized to 100 in January 2000.

Figure 6. Observed and Predicted Home Price Indices
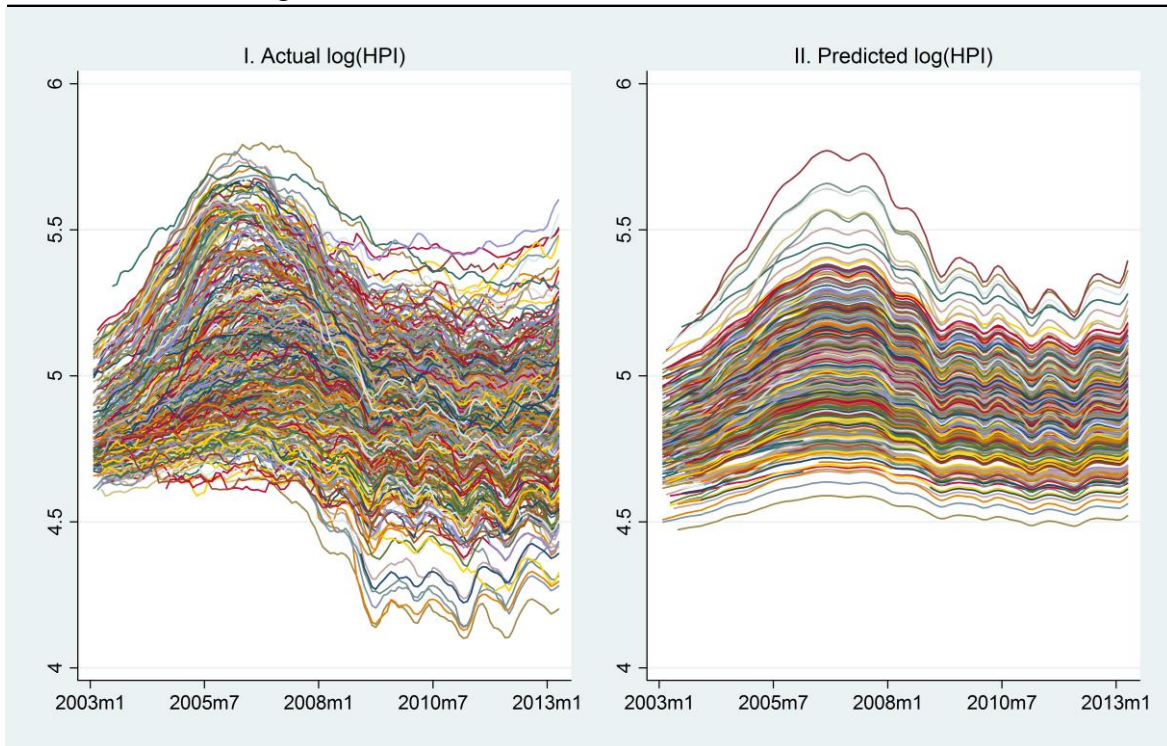
Notes: Figure plots observed log home price indices and predicted indices using long-run variation in the price cycle. The right-hand panel lines show the predicted values from a first stage regression of log(HPI) on CBSA fixed effects and the instrument set, as specified in equation (8) in the text.

Figure 7. Subprime Market Share by Long-Run Price Cyclicality Quartile
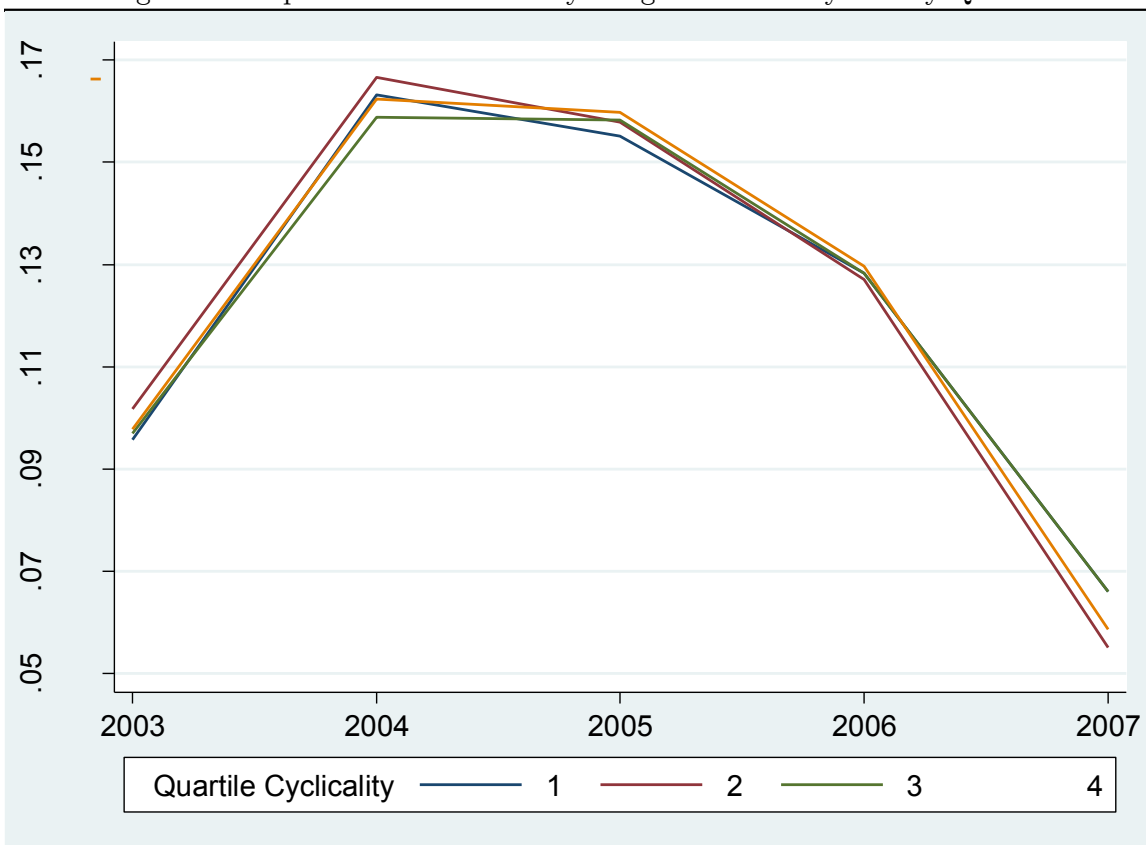
Notes: Figure shows average subprime market share by quartile of the price cyclicality measure defined by equation (5). Subprime market shares are calcualted using HMDA data as the fraction of mortgages originated in a given year that were made by a lender on the HUD subprime lender's list in any year and adjusted for CBSA fixed effects.

Figure 8. Unemployment Rates by Long-Run Price Cyclicality Quartile
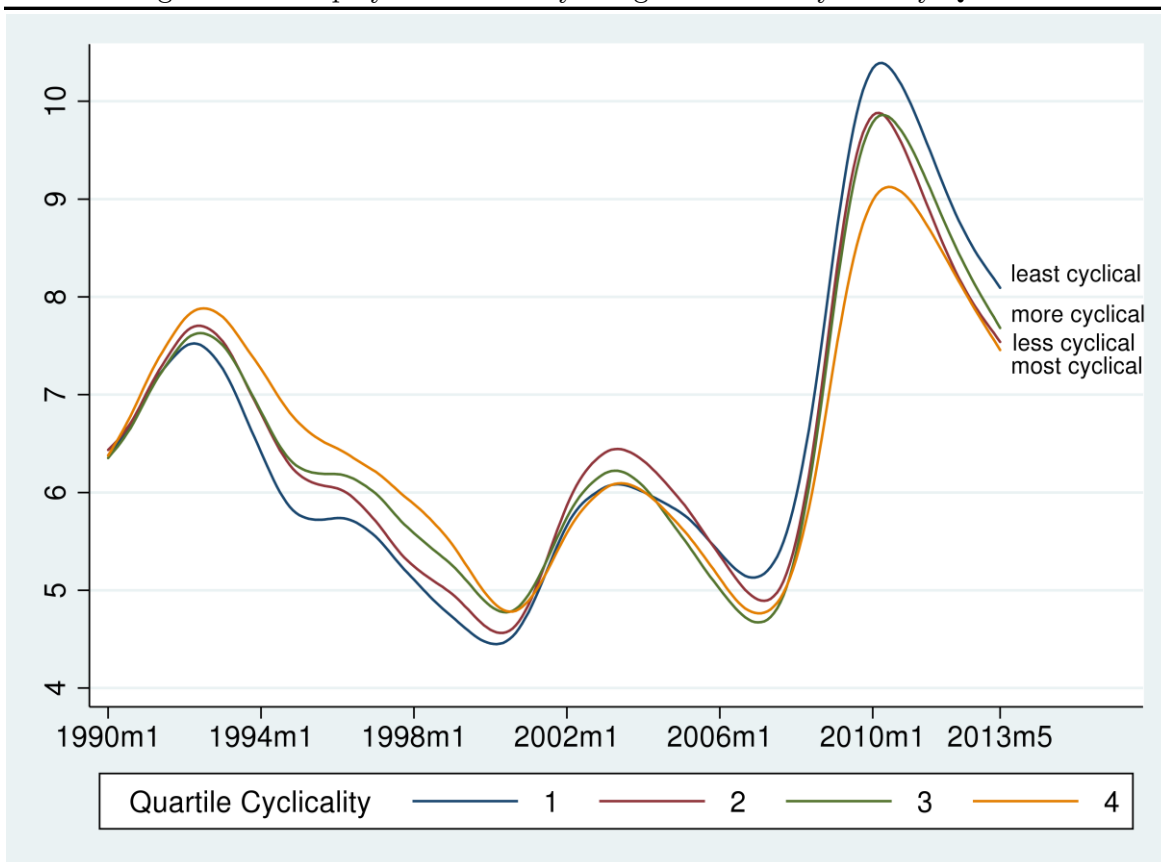
Notes: Figure shows average filtered unemployment rates by quartile of the price cyclicality measure defined by equation (5). Unemployment rates are obtained from the Bureau of Labor Statistics Local Area Unemployment Series and are adjusted for CBSA fixed effects and then filtered with a HP filter with lambda = 1,600.

Figure 9. First-Stage Plots for Combined Loan-to-Value
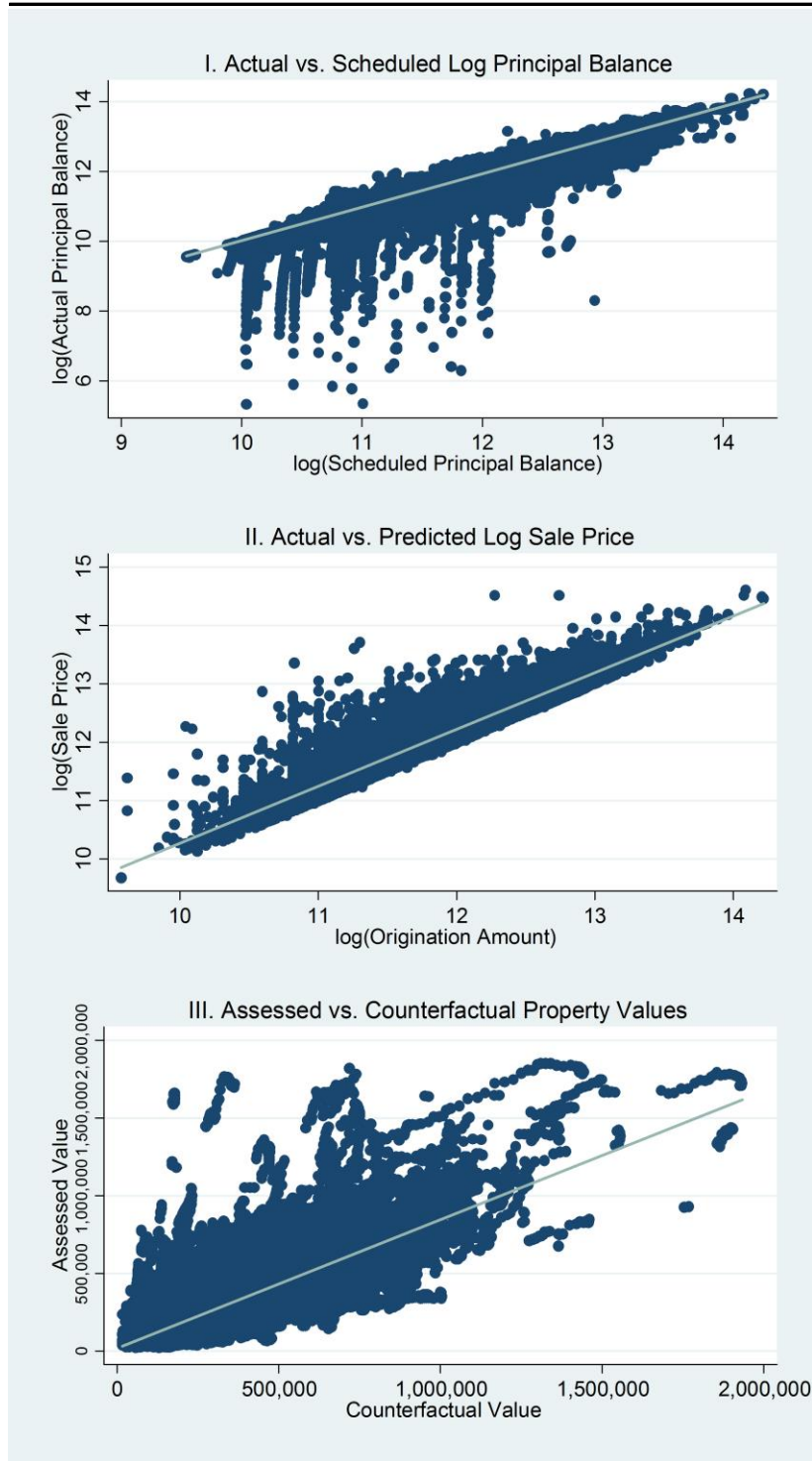
Notes: Panel I plots actual log principal balances versus log balances corresponding to the 30-year fixed-rate mortgage amoritization schedule. Panel II plots log sale prices against log origination amounts. Panel III plots property values against counterfactual values, imputed using home price indices predicted using long-run local variation in home-price cyclicality. Diagonal lines show the fitted bivariate linear regression line.

Table 1. Summary Statistics

	mean	sd	min	max
Default	0.24	0.42	0	1
Prepaid	0.50	0.50	0	1
Censored	0.26	0.44	0	1
2004 Cohort	0.23	0.42	0	1
2005 Cohort	0.29	0.46	0	1
2006 Cohort	0.26	0.44	0	1
2007 Cohort	0.08	0.28	0	1
FICO Score	617.26	59.12	432	881
Debt-to-Income (non-missing)	0.40	0.10	0	0.9
DTI missing	0.26	0.44	0	1
Combined LTV	0.85	0.14	0	1.57
Interest Rate	7.27	1.33	1	13
Full Documentation	0.68	0.46	0	1
Owner Occupied	0.92	0.28	0	1
Cash-out Refi	0.54	0.50	0	1
Adjustable Rate	0.54	0.50	0	1
Interest-only	0.13	0.34	0	1
Balloon	0.09	0.28	0	1
Has 2nd Lien	0.20	0.40	0	1
		N = 32,172		

Notes: Default, prepaid, and censored are indicator variables for a mortgage's termination type. The remaining characeristics are measured at time of orignation. Full documentation, owner occupied, cash-out refinance, adjustable rate, interest-only, balloon mortgage, and has second lien are all indicator variables for the given characteristic. See Section 3 in the text for more details.

Table 2. Summary Statistics by Cohort

	Cohort				
	2003	2004	2005	2006	2007
Default	0.11	0.13	0.25	0.36	0.32
	(0.31)	(0.34)	(0.43)	(0.48)	(0.47)
Prepaid	0.76	0.71	0.52	0.28	0.18
	(0.43)	(0.45)	(0.50)	(0.45)	(0.38)
FICO Score	617.00	618.15	618.59	616.08	614.33
	(61.85)	(61.15)	(59.68)	(56.48)	(54.72)
Debt-to-Income (non-missing)	0.39	0.39	0.40	0.41	0.41
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
DTI missing	0.26	0.23	0.32	0.21	0.24
	(0.44)	(0.42)	(0.47)	(0.41)	(0.43)
Combined LTV	0.83	0.84	0.86	0.86	0.84
	(0.13)	(0.13)	(0.14)	(0.14)	(0.15)
Interest Rate	7.23	6.78	6.89	7.89	8.07
	(1.29)	(1.21)	(1.18)	(1.25)	(1.39)
Full Documentation	0.71	0.70	0.68	0.67	0.68
	(0.45)	(0.46)	(0.47)	(0.47)	(0.46)
Owner Occupied	0.91	0.91	0.92	0.92	0.91
	(0.28)	(0.28)	(0.28)	(0.27)	(0.29)
Cash-out Refi	0.57	0.57	0.53	0.51	0.58
	(0.50)	(0.49)	(0.50)	(0.50)	(0.49)
Adjustable Rate	0.61	0.63	0.57	0.45	0.34
	(0.49)	(0.48)	(0.50)	(0.50)	(0.47)
Interest-only	0.03	0.11	0.21	0.13	0.09
	(0.16)	(0.31)	(0.41)	(0.33)	(0.29)
Balloon	0.01	0.00	0.02	0.21	0.28
	(0.10)	(0.04)	(0.15)	(0.41)	(0.45)
Has 2nd Lien	0.07	0.15	0.24	0.28	0.16
	(0.25)	(0.36)	(0.42)	(0.45)	(0.37)
Observations	4407	7251	9444	8336	2734

Notes: Table reports means and standard deviations in parentheses of individual loan characteristics by borrower cohort. See notes to Table 1 for further details.

Table 3. Effects of Loan Characteristics and Prices:
Default Hazard Model Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2004 Cohort	0.217***	0.223***	0.290***	0.234***	0.188***	0.137**	0.094
	(0.072)	(0.071)	(0.073)	(0.069)	(0.069)	(0.068)	(0.066)
2005 Cohort	0.717***	0.709***	0.747***	0.564***	0.519***	0.407***	0.190***
	(0.100)	(0.097)	(0.097)	(0.089)	(0.087)	(0.075)	(0.068)
2006 Cohort	0.954***	0.984***	0.820***	0.556***	0.579***	0.470***	0.045
	(0.130)	(0.129)	(0.128)	(0.118)	(0.121)	(0.093)	(0.086)
2007 Cohort	0.734***	0.800***	0.613***	0.424***	0.466***	0.235***	-0.107
	(0.120)	(0.116)	(0.120)	(0.112)	(0.116)	(0.083)	(0.084)
12-month Δlog(HPI)						-3.685***	-3.857***
						(0.131)	(0.152)
CBSA FE	n	y	y	y	y	y	y
Borrower Characteris	n	n	y	n	y	n	y
Loan Characteristics	n	n	n	y	y	n	y
Observations	1,224,716	1,224,716	1,224,716	1,224,716	1,224,716	1,224,716	1,224,716
Log likelihood	-44,335	-43,574	-42,642	-43,186	-42,498	-43,142	-42,033

Notes: Table reports maximum-likelihood estimates of the default hazard model given in equations (2) and (3) in the text. All specifications include indicator variables for each value of loan age as a non-parametric baseline hazard. Standard errors in parentheses are clustered at the CBSA level.

Table 4. Default Hazard Model Estimates: Allowing for Unobserved Heterogeneity

	(1)	(2)	(3)	(4)	(5)
2004 Cohort	0.254***	0.188**	0.130**	0.078	0.113*
	(0.073)	(0.082)	(0.056)	(0.063)	(0.065)
2005 Cohort	0.924***	0.674***	0.372***	0.172***	0.250***
	(0.077)	(0.078)	(0.052)	(0.059)	(0.061)
2006 Cohort	1.361***	0.865***	0.380***	0.006	0.117*
	(0.096)	(0.082)	(0.052)	(0.061)	(0.064)
2007 Cohort	1.079***	0.682***	0.110*	-0.186***	-0.039
	(0.097)	(0.093)	(0.060)	(0.070)	(0.074)
12-month Δlog(HPI)			-4.063***	-4.743***	-4.139***
			(0.103)	(0.128)	(0.140)
Borrower Characteristics	n	y	n	y	y
Loan Characteristics	n	y	n	y	y
State Fixed Effects	n	n	n	n	y
Observations	1,224,716	1,224,716	1,224,716	1,224,716	1,224,716
Log likelihood	-44,295	-43,140	-43,631	-42,476	-42,284

Notes: Table reports maximum-likelihood estimates of the default hazard model given in equations (2) and (7) in the text. All specifications include indicator variables for each value of loan age as a non-parametric baseline hazard. Standard errors in parentheses are homoskedastic MLE standard errors.

Table 5. Effect of Long-Run Cyclicality on Price Changes:
First-Stage Results

	(1)	(2)	(3)
2004 Cohort		-0.011***	-0.012***
		(0.003)	(0.003)
2005 Cohort		-0.031***	-0.031***
		(0.007)	(0.007)
2006 Cohort		-0.053***	-0.051***
		(0.011)	(0.011)
2007 Cohort		-0.064***	-0.059***
		(0.013)	(0.012)
Baseline hazard	n	y	y
CBSA FE	n	y	y
Borrower covariates	n	n	y
Loan covariates	n	n	y
Observations	1,224,716	1,224,716	1,224,716
R-squared	0.497	0.559	0.562
Partial F-stat	49.04	31.23	30.97

Notes: Table estimates first stage specifications detailed by equation (11) by OLS. Dependent variable is the 12-month change in the log house price index. The instruments are calendar month indicator variables interacted with the historical cyclicality measure defined by equation (9) in the text. Standard errors are clustered by CBSA.

Table 6. Effect of Accounting for Endogeneity of Prices:
Control-Function Estimates of Default Hazard

	(1)	(2)	(3)	(4)	(5)	(6)
2004 Cohort	0.137**	0.127*	0.123*	0.094	0.083	0.078
	(0.068)	(0.068)	(0.068)	(0.066)	(0.066)	(0.066)
		[0.172]	[0.158]		[0.105]	[0.106]
2005 Cohort	0.407***	0.362***	0.357***	0.190***	0.142**	0.134**
	(0.075)	(0.076)	(0.075)	(0.068)	(0.068)	(0.066)
		[0.195]	[0.178]		[0.104]	[0.098]
2006 Cohort	0.470***	0.393***	0.403***	0.045	-0.034	-0.028
	(0.093)	(0.095)	(0.093)	(0.086)	(0.089)	(0.086)
		[0.242]	[0.212]		[0.079]	[0.074]
2007 Cohort	0.235***	0.147*	0.177**	-0.107	-0.195**	-0.170*
	(0.083)	(0.088)	(0.087)	(0.084)	(0.089)	(0.087)
		[0.256]	[0.203]		[0.084]	[0.076]
12-month $\Delta\log$(HPI)	-3.685***	-4.356***	-4.479***	-3.857***	-4.576***	-4.722***
	(0.131)	(0.352)	(0.325)	(0.152)	(0.362)	(0.335)
		[0.872]	[0.817]		[0.413]	[0.387]
$\Delta\log$(HPI) Fitted Residuals		0.941**	1.138***		1.004**	1.236***
		(0.431)	(0.413)		(0.448)	(0.433)
		[0.918]	[0.851]		[0.463]	[0.431]
Unemployment Rate			-0.020**			-0.021**
			(0.008)			(0.009)
			[0.022]			[0.011]
CBSA FE	y	y	y	y	y	y
Borrower Characteristics	n	n	n	y	y	y
Loan Characteristics	n	n	n	y	y	y
Observations	1,224,716	1,224,716	1,223,448	1,224,716	1,224,716	1,223,448
Log likelihood	-42,305	-43,138	-43,103	-42,033	-42,029	-41,993

Notes: Table reports maximum-likelihood control-function estimates of the default hazard model given in equations (2) and (7) in the text. Fitted residuals are estimated from a linear first stage regression of the 12-month change in the log price index on the instruments and remaining controls. All specifications include indicator variables for each value of loan age as a non-parametric baseline hazard. Standard errors in parentheses are clustered at the CBSA level. Standard errors in brackets are from 200 block bootstrap replications.

Table 7. Effect of Allowing a Flexible Endogeneity Distribution:
Nonparametric Control-Function Estimates of Default Hazard

	(1)	(2)	(3)	(4)	(5)	(6)
2004 Cohort	0.127*	0.120*	0.120*	0.083	0.077	0.077
	(0.068)	(0.067)	(0.067)	(0.066)	(0.065)	(0.065)
	[0.172]	[0.166]	[0.166]	[0.105]	[0.105]	[0.105]
2005 Cohort	0.362***	0.337***	0.339***	0.142**	0.117*	0.120*
	(0.076)	(0.070)	(0.071)	(0.068)	(0.064)	(0.064)
	[0.195]	[0.196]	[0.198]	[0.104]	[0.108]	[0.109]
2006 Cohort	0.393***	0.358***	0.359***	-0.034	-0.066	-0.065
	(0.095)	(0.089)	(0.089)	(0.089)	(0.086)	(0.087)
	[0.242]	[0.242]	[0.244]	[0.079]	[0.081]	[0.082]
2007 Cohort	0.147*	0.122	0.121	-0.195**	-0.219**	-0.219**
	(0.088)	(0.087)	(0.087)	(0.089)	(0.091)	(0.092)
	[0.256]	[0.257]	[0.259]	[0.084]	[0.084]	[0.084]
12-month Δlog(HPI)	-4.356***	-4.737***	-4.658***	-4.576***	-4.944***	-4.877***
	(0.352)	(0.402)	(0.403)	(0.362)	(0.415)	(0.415)
	[0.872]	[0.946]	[0.970]	[0.413]	[0.459]	[0.465]
Δlog(HPI) Fitted Residuals	0.941**	0.263	0.615	1.004**	0.413	0.699
	(0.431)	(0.494)	(0.552)	(0.448)	(0.494)	(0.567)
	[0.918]	[1.045]	[1.093]	[0.463]	[0.543]	[0.558]
(Δlog(HPI) Fitted Residuals)^2		-4.513***	-1.215		-4.841***	-1.764
		(1.580)	(4.205)		(1.791)	(4.238)
		[2.710]	[6.134]		[1.854]	[4.071]
(Δlog(HPI) Fitted Residuals)^3		23.690**	-7.211		19.707**	-4.513
		(9.852)	(32.970)		(8.849)	(28.217)
		[10.317]	[28.550]		[8.665]	[24.850]
(Δlog(HPI) Fitted Residuals)^4			-75.177			-71.583
			(65.940)			(64.402)
			[93.464]			[77.720]
(Δlog(HPI) Fitted Residuals)^5			282.743			195.488
			(446.678)			(381.055)
			[394.651]			[356.268]
CBSA FE	y	y	y	y	y	y
Borrower Characteristics	n	n	n	y	y	y
Loan Characteristics	n	n	n	y	y	y
Observations	1,224,716	1,224,716	1,224,716	1,224,716	1,223,448	1,224,716
Log likelihood	-43,138	-43,114	-43,111	-42,029	-42,007	-42,006

Notes: See Table 5 notes.

Table 8. Effect of Current Combined Loan-to-Value Ratio on Default Hazard:
Control Function Results

	(1)	(2)	(3)	(4)	(5)	(6)
2004 Cohort	0.202***	0.188***	0.127**	0.116*	0.110*	0.092
	(0.066)	(0.064)	(0.063)	(0.064)	(0.064)	(0.063)
2005 Cohort	0.424***	0.348***	0.138**	0.126*	0.133**	0.095
	(0.075)	(0.068)	(0.063)	(0.066)	(0.064)	(0.063)
2006 Cohort	0.317***	0.159*	-0.185**	-0.162**	-0.125	-0.146*
	(0.096)	(0.083)	(0.076)	(0.078)	(0.079)	(0.077)
2007 Cohort	0.143	-0.055	-0.417***	-0.372***	-0.292***	-0.279***
	(0.096)	(0.084)	(0.078)	(0.081)	(0.089)	(0.085)
Underwater	0.683***			0.284***		
	(0.060)			(0.052)		
12-month Δlog(HPI)			-3.221***	-4.693***	-4.718***	-4.814***
			(0.237)	(0.333)	(0.382)	(0.371)
CLTV Fitted Residuals				0.007***	0.013***	0.011***
				(0.000)	-0.001	(0.001)
Δlog(HPI) Fitted Residuals				1.579***	2.252***	2.170***
				(0.406)	(0.500)	(0.495)
Unemployment Rate						-0.050***
						(0.013)
CLTV Linear Spline	n	y	y	n	y	y
Observations	1,037,581	1,037,581	1,037,581	1,036,611	1,037,581	1,036,611
Log likelihood	-35,935	-35,723	-35,477	-35,444	-35,379	-35,329

Notes: Table reports maximum-likelihood estimates of the default hazard model given in equations (2) and (14) in the text. Current combined loan-to-value ratios (CLTVs) are calculated by LoanPerformance as the total outstanding principal on a loan divided by an automated assessing model's estimate of the market value of each home. Underwater is an indicator for CLTV>1. The linear spline is defined by equation (15) in the text. All specifications include individual loan and borrower characteristics, CBSA fixed effects, and indicator variables for each value of loan age as a non-parametric baseline hazard function. Standard errors in parentheses are clustered at the CBSA level.

Table 9. First-Stage Results for Combined Loan-to-Value Ratios

	(1)	(2)	(3)	(4)	(5)
log(Origination Amount)	-0.640***	-0.724***	-0.950***	-0.958***	-0.968***
	(0.047)	(0.043)	(0.047)	(0.046)	(0.046)
log(Prinicipal Balance)	0.787***	0.917***	1.063***	1.066***	1.076***
	(0.044)	(0.042)	(0.047)	(0.046)	(0.046)
Predicted Appreciation	-1.265***	-0.627***	-0.610***	-0.602***	-0.103
	(0.162)	(0.137)	(0.138)	(0.137)	(0.100)
2004 Cohort		-0.001	0.003	0.001	0.019***
		(0.008)	(0.007)	(0.007)	(0.007)
2005 Cohort		0.030**	0.038***	0.024**	0.072***
		(0.013)	(0.013)	(0.012)	(0.013)
2006 Cohort		0.091***	0.110***	0.084***	0.146***
		(0.021)	(0.022)	(0.020)	(0.023)
2007 Cohort		0.156***	0.173***	0.145***	0.190***
		(0.028)	(0.028)	(0.026)	(0.027)
12-month Δlog(HPI)				-0.326***	-0.204***
				(0.024)	(0.032)
Unemployment Rate					0.052***
					(0.006)
Baseline hazard	n	y	y	y	y
CBSA FE	n	y	y	y	y
Borrower covariates	n	n	y	y	y
Loan covariates	n	n	y	y	y
Observations	1,037,581	1,037,581	1,037,581	1,037,581	1,036,611
R-squared	0.242	0.355	0.423	0.428	0.462
Partial F-stat	239.10	331.70	232.29	231.66	221.20

Notes: Table estimates first stage specifications detailed by equation (16) by OLS. Dependent variable is current combined loan-to-value ratio, calculated by CoreLogic as the total outstanding principal on a loan divided by an automated assessing model's estimate of the market value of each home. Standard errors are clustered by CBSA.

Table 10. Counterfactual Annual Default Rates by Cohort

| | Counterfactual Scenario | | Default Rate by Cohort | | | | | |
	Prices	Loan Characteristics	2003	2004	2005	2006	2007	Overall
(1)	Actual	Actual	4.2%	5.3%	9.2%	12.0%	9.8%	8.7%
(2)	Actual	2003	3.7%	4.6%	7.0%	7.4%	6.5%	6.2%
(3)	Actual	2006	5.3%	6.6%	9.9%	10.5%	9.2%	8.8%
(4)	2003	Actual	4.1%	4.6%	5.6%	5.6%	4.4%	5.1%
(5)	2006	Actual	8.5%	9.4%	11.3%	11.0%	8.4%	10.2%
(6)	No price change	Actual	6.3%	6.9%	8.2%	7.9%	6.0%	7.4%
(7)	No price change	2003	5.3%	5.7%	6.0%	4.7%	3.9%	5.3%
(8)	No price change	2006	7.6%	8.1%	8.5%	6.8%	5.6%	7.5%
		Observations	115,567	193,554	281,346	285,277	106,764	982,508

Notes: Table reports estimated annual default rates under the indicated counterfactual scenarios for prices and loan characteristics. Annual default rates are defined as 12 times the average fraction of loans that default in each month, measured over all existing loans within five years of origination. Scenarios using actual characteristics retain observed covariates. Scenarios using a given year's prices replace all price changes with the average price changes faced by the given year's borrowers at each value of loan age. Scenarios using a given year's loan characteristics assign all loans the average characteristics from the indicated cohort.