BANK FOR INTERNATIONAL SETTLEMENTS

BIS Working Papers

No 875

The Impact of credit risk mispricing on mortgage lending during the subprime boom

by James A Kahn and Benjamin S Kay

Monetary and Economic Department

August 2020

JEL classification: G21, E44, E32.

Keywords: financial crisis, mortgage insurance, housing finance, default risk.

This paper was produced as part of the BIS Consultative Council for the Americas (CCA) research conference on "Macro models and micro data" hosted by the Central Bank of Argentina, Buenos Aires, 23–24 May 2019. BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS.

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

The Impact of Credit Risk Mispricing on Mortgage Lending during the Subprime Boom

James A. Kahn and Benjamin S. Kay*

November 29, 2019

Abstract

We provide new evidence that credit supply shifts contributed to the U.S. subprime mortgage boom and bust. We collect original data on both government and private mortgage insurance premiums from 1999-2016, and document that prior to 2008, premiums did not vary across loans with widely different observable characteristics that we show were predictors of default risk. Then, using a set of post-crisis insurance premiums to fit a model of default behavior, and allowing for time-varying expectations about house price appreciation, we quantify the mispricing of default risk in premiums prior to 2008. We show that the flat premium structure, which necessarily resulted in safer mortgages cross-subsidizing riskier ones, produced substantial adverse selection. Government insurance maintained an flatter premium structure even post-crisis, and consequently also suffered from adverse selection. But after 2008 the government reduced its exposure to default risk through a combination of higher premiums and rationing at the extensive margin.

Keywords: Financial Crisis, Mortgage Insurance, Housing Finance, Default Risk
JEL Codes: G21 (Banks • Depository Institutions • Micro Finance Institutions • Mortgages), E44 (Financial Markets and the Macroeconomy), E32 (Business Fluctuations • Cycles)

*Kahn: Yeshiva University (james.kahn@yu.edu); Kay: Board of Governors of the Federal Reserve System (benjamin.s.kay@frb.gov). We thank Robert Garrison II and Jared Rutner for excellent research assistance, and the Office of Financial Research for support of this work. CoreLogic Loan—Level Market Analytics (LLM 2.0) was obtained while at the OFR. We also thank Larry Cordell, Narayana Kocherlakota, Douglas McManus, Tess Scharlemann, and other participants at presentations at the University of Connecticut, Office of Financial Research, Federal Reserve Board, the University of Rochester, and the NYC Real Estate Conference, for helpful comments. Disclaimer: Views expressed in this paper are those of the authors and not necessarily of the Office of Financial Research. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of anyone else associated with the Federal Reserve System. An earlier version of this paper had the title "Mispricing Mortgage Credit Risk: Evidence from Insurance Premiums, 1999-2016."

Was the subprime lending boom of the early 2000s the consequence only of increased optimism on the part of borrowers and lenders regarding house price appreciation? Or was it also the result of a pure supply shift, an increase in the quantity of loans in the direction of greater risk? While the two hypotheses are not mutually exclusive, they are distinct. The optimism story (see, for example, Adelino et al. (2016), Brueckner et al. (2012)) holds that market participants believed that there was a reduction in the quantity of default risk: the collateral was safer than before, economic conditions appeared robust, and securitization facilitated diversification. There was no change in the price of a given level of credit risk. Yes, ex post these beliefs proved incorrect, but this is only really knowable with hindsight. The supply shift hypothesis (Mian and Sufi (2009)), by contrast, is that lending shifted in the direction of greater quantities of risk at the same or lower price. If mortgage demand curves slope downward, the supply shift hypothesis implies that the price of credit risk declined and the quantity increased. The optimism hypothesis implies that the perceived quantity of credit risk (given observable characteristics) declined, but not its price. [1]

In practice, it is difficult to distinguish variation in the pricing of credit risk from changes in the quantity of credit risk. In particular, mortgage interest rates are an amalgam of many difficult-to-quantify or unreported factors: interest rate risk, prepayment risk, prepaid interest (i.e. "points"), and details of any mortgage insurance. Consequently, interest rate spreads on mortgages are not pure indicators of credit risk. In the face of these measurement difficulties, much of the research in this area has focused instead on quantities, in particular the numbers or dollar value of high-risk mortgages (e.g. Foote et al. (2016), Mian and Sufi (2009), and Ambrose and Diop (2014)).[2]

In comparison to mortgage interest rates, premiums on private mortgage insurance (PMI) provide a market based measure of default risk largely uncontaminated by interest rates, prepayment, and other factors irrelevant to credit risk. Mortgage insurance is an important but often over-

[1]Kaplan et al. (2017) use the term "credit conditions" for the supply side of the mortgage market, and focus on constraints such as loan-to-value limits and fees, in addition to a "spread" in the mortgage rate. Apart from these details, they treat credit conditions as distinct and independent from beliefs. In our approach, beliefs and credit conditions are intertwined. More optimistic beliefs will endogenously lead to more relaxed credit conditions.

[2] Justiniano et al. (2016)) is an exception.

looked feature of mortgage lending in the United States, United Kingdom, Hong Kong, Australia, and Canada. According to Urban Institute (2017), in 2016 roughly 65 percent of purchase mortgages in the United States were (privately or publicly) insured. Moreover, since insured mortgages include virtually all mortgages with LTV above 80 percent, they play an even larger role in the market for risky mortgages. Moreover, mortgage insurance does not just shift a large portion of the default risk to the the insurer, it reverses the typical copayment pattern of standard insurance: insurers bear the losses from default up to the coverage limit, and only when losses exceed the insurance coverage does the holder of the mortgage suffer any losses.[3] Since coinsurance is a mechanism to balance risk-sharing with incentives to avoid risky choices (Doherty and Smetters (2005)), this structure should give mortgage insurers the incentive to take primary responsibility for risk mitigation.

Given that one of the key alleged causes of the 2008 financial crisis is the misapprehension of risks in mortgages (United States Financial Crisis Inquiry Commission (2011)), and mortgage insurers were major underwriters of mortgage risk,[4] their behavior during the period leading up to 2008 has been surprisingly neglected.[5] This partly reflects a data gap: Commercial mortgage performance data like LPS and CoreLogic, and the portfolios published by Fannie Mae and Freddie Mac, do not provide data on insurance premiums.

This paper makes the following contributions: First, to fill the data gap, we collect original data on mortgage insurance premiums from 1999-2016. This details the evolution of PMI offerings in their scope as well as in their price. We also assemble data on Federal Housing Authority (FHA) premiums during the same time period, and devise adjustments to make them comparable to PMI premiums. This work is described in Section 1. Second, to characterize the overall pricing

[3]In effect, it puts the mortgage holder more in the position of a typical insurer, and the insurer more like the insured, with the coverage representing the copayment.

[4]In our 2005 data, 27 percent of single family purchase mortgage loans had an LTV above 80, making them a candidate for mortgage insurance. Since second liens (with a first lien having an 80 LTV) are a substitute for mortgage insurance, the population of potentially insurable loans is even larger. In fact, though, the cost difference between second liens and mortgage insurance was miniscule.

[5]Epperson et al. (1985) includes quantitative analysis of PMI pricing, but in a market environment very different from that of 2000-2008. In addition, we recently learned about related work of Bhutta and Keys (2017), who argue, consistent with our view, that the mortgage insurers passively accommodated the shift to riskier products prior to 2008.

of mortgage insurance, as borrowers substitute among loan types, we construct chain-weighted price indexes of insurance products in four risk categories. These indexes, described in Section 2, reveal broad changes in the pricing of default risk over time. Unfortunately, the indices cannot distinguish between changes in the underlying credit risk from changes in the accuracy of risk pricing.

To address this last distinction, in Section 3 we fit a parametric model of default behavior to PMI prices in 2013. This quantifies default risk conditional on borrower's equity, the distribution of house price changes, and borrower credit worthiness. With 2013 PMI premiums as our benchmark, but allowing for differing expectations about house price appreciation, we are able to judge the accuracy of premiums in 2005, arguably the peak of the boom. In so doing, we infer a pattern of pricing (and mispricing), consistent with movements along downward-sloping demand curves, that explains much of the unusually large market share of risky products during the boom, as well as the large movements between private and government insurance.

1 A Closer Look at Premiums

The two categories of residential mortgage insurance in the United States are PMI and government mortgage insurance. Both are important. Most US home buyers who obtain a government sponsored enterprise (GSE) mortgage with a down payment of less than 20 percent of the purchase price are required to purchase PMI, which protects the holder against losses on the covered portion of the loan. Government insurance, such as that offered by the FHA or Department of Veterans' Affairs (VA), represents an alternative to PMI, but has typically been priced to attract borrowers with lower down payments and credit scores.[6]

From 1998 to 2007 PMI was the dominant product with about 65% market share of insured loans. From 2008-2018Q1 government insurance has dominated with about 70% market share (Goodman et al. (2018), p. 32.). In Section 1.1 we detail the pricing of PMI. The primary focus

[6] PMI typically covers between 12 and 35 percent depending on the loan-to-value ratio. FHA insurance offers 100% coverage, while VA coverage is 25%.

of this paper is on private insurance because the pricing of PMI is a market price and therefore an informative equilibrium outcome. The share of PMI in new mortgage issuance has varied widely, depending on market conditions, but in recent years has been on the order of half the insured market. Section 1.2 then examines the pricing of government mortgage insurance.

We also make use of mortgage origination data from CoreLogic Loan Level Market Analytics (LLMA 2.0). This is a database with observations on over 15 million mortgages during the period 1999-2014, including the borrower's FICO score, LTV, and documentation level, as well as the mortgage interest rate and whether the loan is insured. We limit our analysis to 30-year fixed rate, owner-occupied, single-family mortgages. This allows us to corroborate our assessment of product availability during this time period and to obtain mortgage quantities by product.

1.1 Private Mortgage Insurance

This section details our original data on private mortgage insurance premiums from 1999-2016. These premiums provide a detailed history of how risk was (or was not) priced during this turbulent period. There is a dramatic change in the pricing structure of premiums during our sample. Before 2008, for prime mortgages with full documentation that were always insurable, the principal risk priced by private mortgage insurers was leverage (as measured by the loan-to-value ratio, hereafter LTV). It is notable that before 2008 there was no pricing of credit risk for FICO scores 640 and higher.[7] After 2008 PMI pricing on prime loans varied substantially by FICO score. Figure I illustrates the representative case of PMI premiums on $\leq$ 90 LTV, $\geq$660 FICO, full documentation mortgages during the 1999-2016 period. PMI rates fan out by FICO scores only starting in 2008. Prior to 2008, we see that only LTV risk was differentially priced.

These products were insurable throughout our sample. This was not the case for riskier products with lower documentation, lower FICO scores, or higher LTVs. During this period, riskier products saw major changes in availability. Figure II shows that from 2000 to 2005 mortgage insur-

[7]Mortgages with LTV exceeding 97 percent and a FICO score below 660 were charged a higher premium, but nearly all of high-LTV borrowers opted for government insurance. There were some other small risk based adjustments but not for FICO scores.

ers nearly doubled their product offerings (leaving aside briefly the precise definition of a product) and almost all the new products were higher risk. Post-crisis fewer products are available than in 2000, and almost all the eliminated products were higher risk ones.[8]

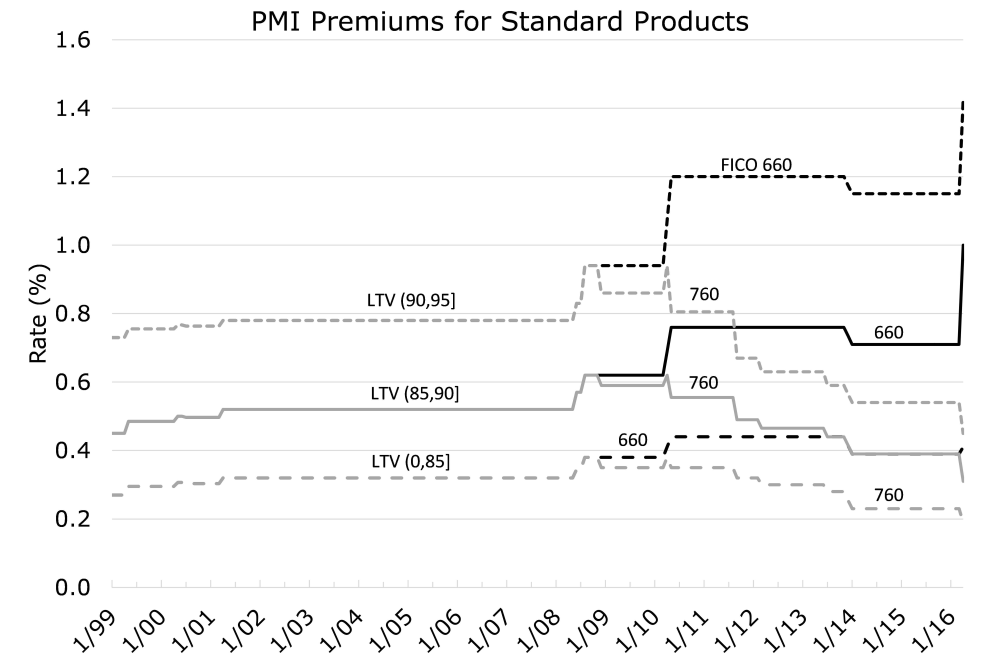
We also find the availability of insurance across risk characteristics changed substantially. During the boom, the range of mortgages that were insurable expanded enormously to include loans to borrowers with low FICO scores, high LTVs, or less than complete income documentation. PMI made them eligible for purchase by the GSEs, which further facilitated their growth. We collect PMI premium data from published rate sheets and publicly available archives of state insurance regulators, primarily Wisconsin (Wisconsin Office of the Commissioner of Insurance) and North Carolina (NC Department of Insurance). These states have the longest digital records for PMI prices. In addition, Wisconsin and North Carolina are the insurance regulators of domicile of two major private mortgage insurers which gives confidence in the accuracy and completeness of their records. Examination of rates from other sources indicates little and often no variation in premiums across states.

We limit our sample to the most common form of insurance: premiums that are fixed for the life of the insurance and paid monthly.[9] This also makes them comparable to mortgage interest rates. We further limit the sample to premiums on 30-year fixed rate mortgages, with "coverage rates" (the percentage of the initial principal that is insured) at standard levels set by the GSEs. Table I lists the standard coverage rates. Figure III demonstrates how the coverage rates in Table I translate into loss absorption provided by the private mortgage insurers, conditional on a loan's loan-to-value ratio (LTV). Since mortgage insurance is primarily a product for loans with LTV $>$ 80 percent, it is notable that the exposure for mortgage holder for insured loans in Table I are

[8] This pattern contrasts with Edelberg (2006), who finds that more granular pricing of default risk (including income, assets, and indebtedness information but not including credit scores) in a range of consumer loans began in the 1990s. She also finds evidence that as a consequence, credit was more widely available. We find, by contrast, that more granular pricing in mortgages only began after 2008, and it was accompanied by a reduction in the availability of credit, i.e. an increase in rationing.

[9] For borrowers current on their loans, PMI is automatically canceled when the ratio of the amortized loan balance to the assessed house price at origination is $\leq$ 78 percent. It can also be canceled if the house is re-appraised and this shows an updated LTV, reflecting both the current loan balance and new appraisal price, is below 78.

Figure I
Until 2008 PMI Rates were Generally Homogeneous Across Prime FICO Scores



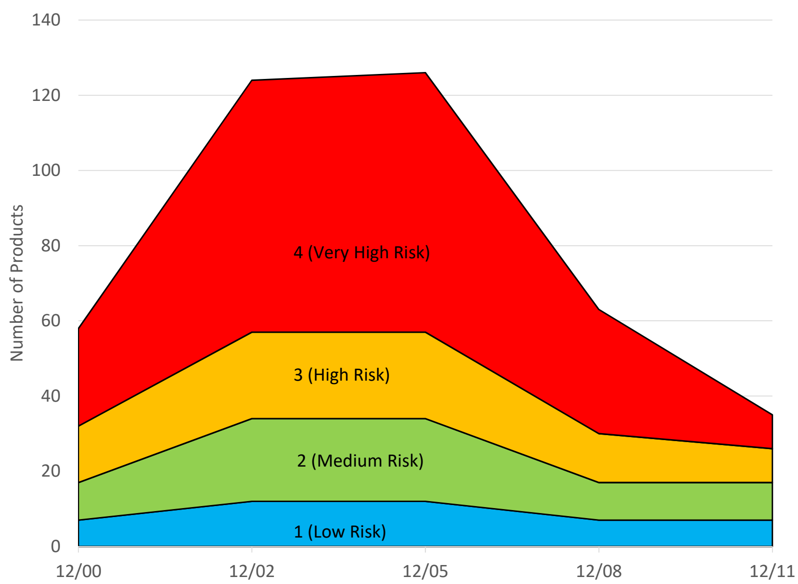
Source: WI and NC mortgage insurer regulatory filings

Table I
Standard coverage rates on 30 year mortgages

LTV	Coverage	Exposure
$\leq 85\%$	12%	$\leq 75\%$
$85.01 - 90$	25	≤ 67.5
$90.01 - 95$	30	≤ 66.5
$95.01 - 97$	35	≤ 63
≥ 97	40	≤ 60

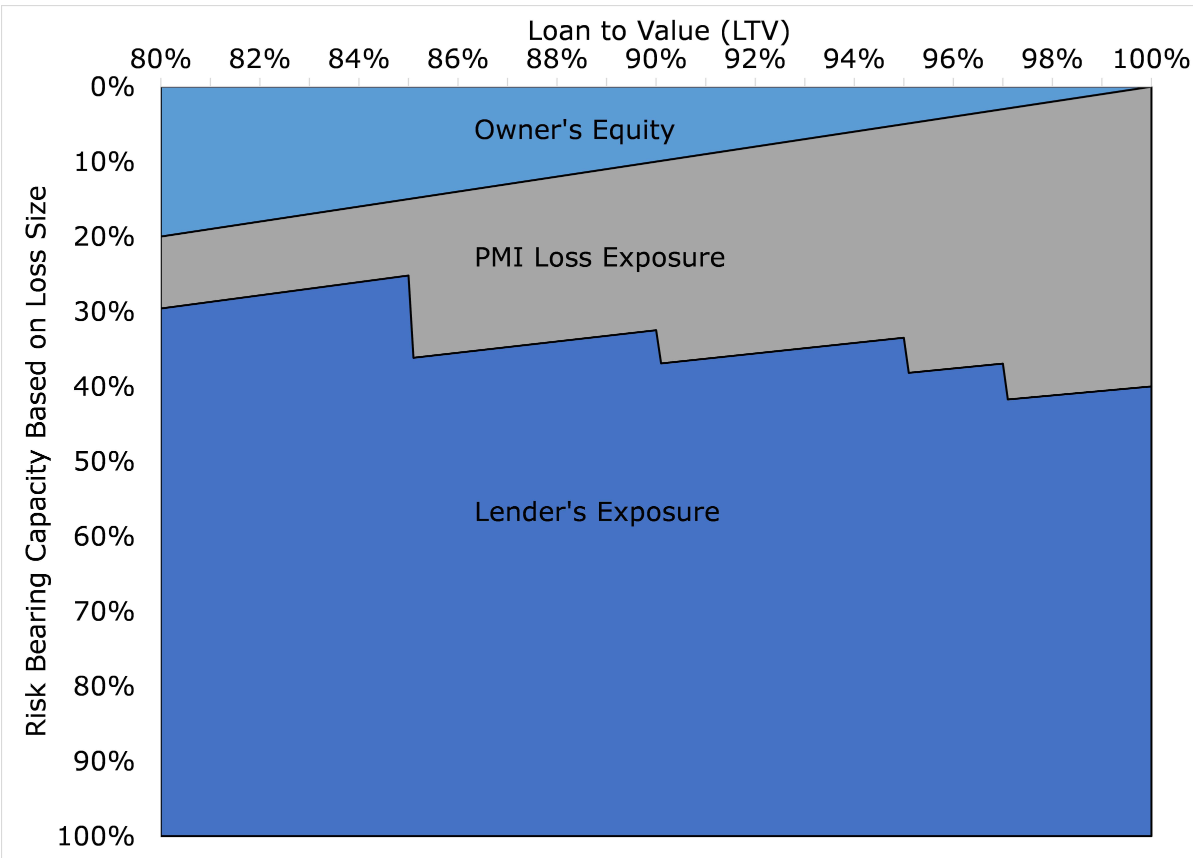
Source: MGIC (2017)

Figure II
Number of PMI Products Available by Risk Category



Section 2 details the methodology for defining PMI products and assigning them to risk categories. Sources: Lam et al. (2013) and author's analysis.

Figure III
Liability Structure of Loans with Privately Mortgage Insurance at Standard Coverage Rates



As losses increase, first the home owner's equity absorbs losses, then private mortgage insurer capital is at risk, and finally, for the largest losses, the bank or GSE's capital is at risk. Sources: MGIC (2017) and authors' calculations uniformly below 80 percent and generally declining in LTV. Presumably, this reflects the increasing likelihood of default and its associated costs. It suggests that the structure is intended to make lenders (or the ultimate holders of the mortgages) roughly indifferent to the borrower's choice of LTV over the 75 to 100 percent range. While the mortgage holder retains some default risk, this structure places the onus of underwriting differential risk by LTV almost entirely with the insurer.

In what follows, we will refer to the scope of insured "products," by which we generally mean combinations of FICO and LTV ranges, along with the level of documentation. We consider seven LTV bins: $[0, 80]$, $(80, 85]$, $(85, 90]$, $(90, 95]$, $(95, 97]$, $(97, 100]$, and > 100. The 11 FICO bins consist of ≥ 760 down to $600 - 619$ in increments of 20, plus $575 - 599$ and $550 - 574$. We also consider mortgages with full documentation ("Full Doc") and incomplete documentation ("Low Doc"). These bins correspond to how the premiums are generally published. For example, insurers would quote one rate for all standard coverage policies on all mortgages for all LTVs between 80 and 85, for all FICO scores between 600 and 619, and with full documentation.

The seven LTV bins, 11 FICO bins, and two levels of documentation result in 154 possible products. The data are essentially daily, since rate sheets typically give an effective date, but we aggregate to monthly using the rate in effect on the first day of the month, and in much of our subsequent analysis is at the annual frequency, since rate changes are relatively infrequent. We collected data from three companies (two active firms: Mortgage Guaranty Insurance Corporation and United Mortgage Guaranty, as well as Triad, which ceased operations in 2008 and subsequently went bankrupt). We averaged across quotes from multiple firms when available for the same product at the same date. However, there was very little and most commonly no difference across firms in pricing of comparable products.[10,11]

All rate changes must be reflected in regulatory filings. Therefore, we can be confident that between regulatory filings rates remain unchanged, and that they can be filled forward until the

[10] We also spot check PMI rate sheets from two additional companies (Radian and Republic) and find the same rates.

[11] In practice, there are many other dimensions that make small adjustments to the quoted PMI rate. These include, for example, second homes, investment properties, multiple units, very large loans, and condos. This paper studies only owner occupied, non-condo, single unit loans.

next rate sheet is filed. For example, if for a given product i we observe a rate sheet for date t and another in effect beginning date $t + \tau$, we can safely assume that rates were the same for dates $t, t + 1, ..., t + \tau - 1$.

Because the data vary both qualitatively (the scope of products available) and quantitatively (the level of premiums) over the nearly 18-year time-span of the sample, it is difficult to summarize concisely. We first consider just the riskiest Full Doc products (FICO scores ≤ 660 and $LTV \geq 90$), which is where some of the biggest changes in both rates and availability occurred. Table II shows sample rates for just these products from 1999, 2001, 2004, 2006, and 2011. We see that more risky products became available, first at a relatively low price. Then a steeper price gradient emerged, and the riskiest products disappeared post-2008. In the documentation dimension a similar but more extreme pattern occurred. Insurers began to offer rates on Low Doc mortgages as early as 2000 (in our data) for safer FICO-LTV combinations. Then products with lower FICO scores and higher LTVs appeared in 2003, albeit with high premiums (in some cases annually exceeding 5 percent if the loan value). In 2006 the riskiest Low Doc products disappeared. By 2009 all Low Doc products disappeared-in our rate sheets, though according to the loan level data, these loans are still occasionally issued (circa 2016-17), but in almost negligible quantities far below pre-crisis levels.

Whenever possible, in this analysis, we use observed PMI rates. When PMI rates cannot be observed we use a hedonic regression model to impute the premiums. We do this for two reasons. First, the CoreLogic dataset contains insured mortgages with combinations of date, LTV, FICO, and documentation for which we could find no quoted rates. Second, we will construct four risk-based PMI price indexes in Section 2, which requires prices for products one period after their disappearance and one period before their appearance. Our method is intended to emulate the methods that the Bureau of Labor Statistics (BLS) uses in constructing price indexes when products appear and disappear.

The goals of our imputation model are accurate in-sample fit for observed premiums and plausible out-of- sample fit for products with unobservable prices. The model (detailed in Appendix

Table II
Sample PMI Rates on Higher-Risk Products*

		Minimum FICO Score					
	LTV	660	640	620	600	575	550
1999	(95,97]	1.04	1.04	-	-	-	-
	(90,95]	0.75	0.75	-	-	-	-
	(85,90]	0.47	0.47	-	-	-	-
2001	(95,97]	0.99	0.99	1.33	1.48	1.70	1.70
	(90,95]	0.78	0.78	0.99	0.99	1.30	1.30
	(85,90]	0.51	0.51	0.62	0.62	0.73	0.74
2004	(95,97]	0.96	0.96	1.42	1.88	2.57	4.18
	(90,95]	0.78	0.78	1.00	1.32	1.80	2.92
	(85,90]	0.52	0.52	0.68	0.90	1.22	1.97
2006	(95,97]	0.96	0.96	1.54	2.05	2.97	4.18
	(90,95]	0.78	0.78	1.08	1.44	2.08	2.92
	(85,90]	0.52	0.52	0.74	0.98	1.41	1.97
2011	(95,97]	1.64	2.34	-	-	-	-
	(90,95]	1.20	1.36	-	-	-	-
	(85,90]	0.76	0.90	-	-	-	-
2013	(95,97]	1.53	1.53	-	-	-	-
	(90,95]	1.20	1.20	-	-	-	-
	(85,90]	0.76	0.76	-	-	-	-

* Units are in percentage points per year paid monthly.

Rates shown are those that prevailed for a majority of the year indicated.

Source: WI and NC mortgage insurer regulatory filings

A) is essentially a 3rd-order polynomial in FICO score and LTV, with time-varying coefficients and interaction terms with a dummy for low documentation. We allow most of the coefficients to vary by calendar year (we omit some year interactions early on when rates are stable). We write our regression specification such that the pure year effects to match the average premiums each year for the highest quality (FICO $\geq$ 760, LTV $\leq$ 80, Full Doc) products. We have 208 monthly observations (January 1999-April 2016) on up to 154 products (i.e. potentially 32,032 premium observations). However, since many of the products are not available throughout this period, we end up with 16,767 premium observations over 139 products.

The dependent variable in the regression is $\ln(\pi_{it}) - \ln(\pi_{0t})$, where π_{it} is the premium for product i at date t, and π_{0t} is the annual average of the premium for the $i = 0$ (FICO 760+, LTV 70, full documentation) product. The equation is specified so that the $i = 0$ product has all the dependent variables equal to zero. This model explains more than 95 percent of the in-sample variation in observed PMI rates. Figure IV depicts the actual premium surface in LTV-FICO space as of 2006, which is well approximated by the regression model. Note that while the very highest-risk mortgages face substantially higher premiums, the surface is flat over a broad range of FICO scores at all LTVs. We return to these results in the risk price indexation exercise in Section 2.

1.2 Government Insurance

In contrast with the private market, the providers of government insurance (the Federal Housing Administration (FHA), the Veterans Administration (VA), and the US Department of Agriculture (USDA)) set premiums based in part on policy objectives rather than an pure assessment of risk. Thus there is no presumption that these premiums are informative about risk, and indeed several grounds for suspecting otherwise. First, government insurance prices fewer aspects of borrower risk than does PMI, which suggests less concern with accurately pricing loan level risks and more with cross-subsidizing and in encouraging the use of these programs by a target constituency. Second, the prices of government insurance changed less frequently than private insurance, suggesting a weaker connection to financial fundamentals at the public insurers than in private insurers. Third, there are longstanding and well-documented criticisms about the way in which government insurers account for borrower risk (Deng et al. (1996), Pennington-Cross et al. (2000), Aragon et al. (2010), Elmendorf (2011), Chirico and Mehlman (2013), Ligon and Michel (2015)). Nevertheless, government insurance is an important part of the mortgage insurance market, especially during the period under study in this paper, so it is essential to incorporate it into our analysis.

Government insurance has had a pricing structure that differs qualitatively from most private insurance. Here we will focus on FHA loans, which are the largest category.[12] Since the 1960s these loans have allowed relatively high loan-to-value ratios, typically up to 95 percent plus closing costs. While the FHA has minimum FICO scores, they have never priced FICO scores as private insurers have (to varying degrees, as we have seen). The FHA also has relatively low minimum FICO score (generally 500, though at times 580 for loans with 95 or higher LTV).[13] Also, FHA insurance pricing has had at most two LTV tiers, 95 percent and above, or less than 95 percent. Finally, FHA insurance is structured to have an upfront fee plus a monthly premium.[14] This last feature complicates comparisons with PMI (which most commonly has only monthly premiums, though upfront options have sometimes been available). As with points, different borrowers may favor one or the other, depending on their anticipated horizon or time till prepayment. *Ceteris paribus*, an FHA loan will be relatively more attractive to a borrower with a longer horizon. There are also sometimes differences in eligibility requirements between FHA and private insurance that result in one or the other being unavailable to generally similar borrowers.

[12]The major other forms of government insured loans, VA and USDA loans, have special eligibility requirements like military service by the borrower or a home in a rural area, that make them less broadly available. This makes FHA the form of government insurance most comparable with PMI.

[13]For 10 weeks in 2008 (7/14/2008 - 9/30/2008) the FHA briefly used a matrix based premium pricing format similar to that of PMI. The Housing and Economic Recovery Act of 2008 provided for a one-year moratorium on the implementation of FHA's risk-based premiums beginning October 1, 2008 and they were never re-established (Federal Housing Administration, Office of the Assistant Secretary for Housing, and Federal Housing Commissioner (2008), Rumsey (2017))

[14]In addition to the upfront fee component, FHA insurance provides 100 percent coverage, versus 12 to 25 percent for PMI. In fact, though, between PMI and the buyer's down payment, lenders are protected for the first 30 to 40 percent of declines in home values (see Figure III). Because losses of such magnitude happen only for sharply negative housing return realizations that are rare, we calculate that the coverage differential can be neglected in comparing FHA and PMI premiums.

The bottom line is that government and private insurance do compete, but they are imperfect substitutes. In many instances, borrowers may not be able simply to choose the less expensive option. Moreover, the relatively coarse structure of FHA insurance pricing suggests little or no intent to compete directly with private insurance for borrowers with relatively low (say 90 percent and below) LTVs, and/or relatively high (say 700 and higher) FICO scores.

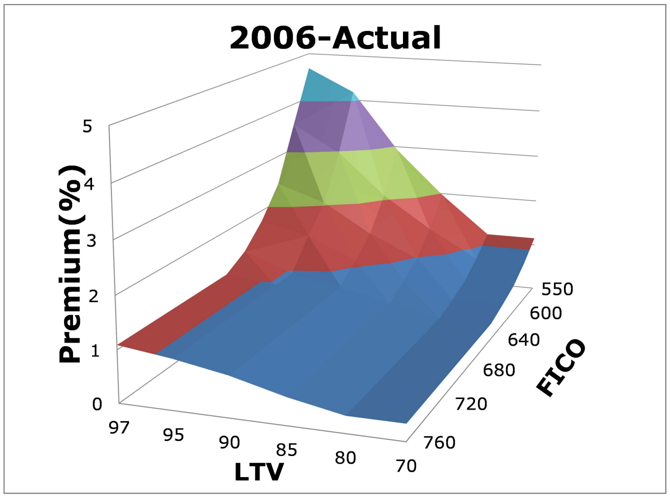
Table III provides sample data on actual FHA premiums for the same selected years in the manner that Table II did for PMI. Also included in this table is a calculation of a PMI equivalent rate. For this, the upfront payment is amortized with either a 7-year ($T = 7$) or 3-year ($T = 3$) horizon and added to the recurring premium. The 7-year horizon is intended for normal market conditions, but below we assume that at the peak of the boom the horizon was more like three years.[15]

A comparison of the two tables shows that using a 7-year horizon for comparison, FHA insurance in 1999-2001 is cheaper only for loans with LTVs exceeding 95 percent. But by 2004-2006, however, the 7-year horizon implies that it was actually cheaper than or competitive with PMI over a wider range of loans. In Section 3.2 we argue that the 7-year horizon might be too long for that particular time period. A 3-year horizon would have made FHA insurance competitive as usual only for the riskiest loans, those with LTV at least 95 percent or FICO scores below 640.

In addition to the private and government insured mortgage products, there were two additional ways of financing high LTV mortgages. First, some high LTV mortgages were completely uninsured. We lack data on the pricing of credit risk in such mortgages for the same reason as in other mortgages: credit spreads must be disentangled from other factors contributing to mortgage interest rates. Second, some homeowners financed their purchases with multiple mortgages. The first lien mortgage would be a conforming (GSE) mortgage with sub-80 LTV and without PMI.

[15] Seven years reflects the normal combined prepayment risk of refinancing, default, and sale. Median home tenure is about 15 years (Emrath (2009). In addition to prepayment due to sale, loans also end early due to refinancing and default. On average, from HMDA data, the refinancing rate for home owners is 8% per year which under a constant hazard rate would implies an average loan life expectancy of 11.5 years (median 8.5, Chen et al. (2013)). Foreclosure rates are about 1.5% per year (Neal (2015), Aron and Muellbauer (2016)). Combining the moving, default, and the refinancing hazards gives an annual hazard of 14% per year and a 7 year average loan life expectancy.

Figure IV
2006 Credit Surface



Source: WI and NC mortgage insurer regulatory filings and authors' calculations

Table III
FHA Premiums on Fixed-Rate Mortgages*

	LTV	Upfront	Recurring	Annual Equivalent T=7	T=3
1999	>95	2.25	0.50	0.90	1.33
	≦95	2.25	0.50	0.90	1.33
2001	>95	1.50	0.50	0.77	1.06
	≦95	1.50	0.50	0.77	1.06
2004	>95	1.50	0.50	0.76	1.05
	≦95	1.50	0.50	0.76	1.05
2006	>95	1.50	0.50	0.76	1.05
	≦95	1.50	0.50	0.76	1.05
2011	>95	1.00	1.15	1.32	1.51
	≦95	1.00	1.10	1.27	1.51
2013	>95	1.75	1.35	1.64	1.97
	≦95	1.75	1.30	1.59	1.97

* Units are in percentage points per year paid monthly.

Rates are those that prevailed for a majority of the year indicated.

Source:Mortgage Banker's Association, authors' calculations

The second lien would be a bank loan, also uninsured, that the bank would hold on its balance sheet or privately securitize. Our data, the CoreLogic LMMA 2.0 data, only contains first lien information, so we cannot see these loans directly. We use the CoreLogic data to identify loans that are uninsured or indicate that they involve a second lien (because their combined LTV (CLTV) is greater than the loan LTV). In 2006, at the height of the use of second lien mortgage financing, our pricing data represents at least one-half, and probably more than 60 percent of the > 80 CLTV single-family, purchase, owner occupied, and 30-year amortization loan market. In 2014, our pricing data represent 98+ percent of these loans.[16]

2 Risk Pricing over Time

This section summarizes the changes in the price of mortgage insurance over time. Section 2.1 constructs price indexes using only the PMI data. This simplifies comparing product pricing and allows for product entry and exit. Section 2.2 expands this analysis by adding government mortgages to the mix of products included in the indexes.

2.1 The Risk Structure of Private Insurance Premiums

In this section we construct price indexes for sub-aggregates of products by risk category. These indexes require price and quantity data. As described in the previous section, we use PMI prices

[16] In this population in CoreLogic LLMA 2.0, uninsured loans peak at eight percent (in 2006). Loans with identified second liens were never more than 18 percent of loans with CLTVs > 80 (also in 2006). However, particularly in the boom, there were first lien loans issued by issuers that did not know that part of the down payment came from another loan, a practice known as a "silent second" (Ashcraft et al. (2008)).

In our data, silent seconds would likely show up as uninsured sub-20 LTV loans. In our single family, owner occupied, purchase, 30-year amortization, > 80 LTV data, we find 651 thousand government or privately insured mortgages in 2006. CoreLogic covers about 60 percent of all US first lien mortgages, implying about 1.1 million insured loans in 2006. In the same period, Avery et al. (2007) estimate from HMDA data that there were 1.26 million second liens for all owner occupied purchase loans (our sample is 30 year term and fixed rate only, while in 2006 ARMs were about a third of the market and more than five percent had terms below 30 years (Goodman et al. (2018))). They estimate ten percent of these loans have CLTVs ≤ 80, selected not as an alternative to PMI but instead to keep the first lien at the conforming limit. This implies a maximum of 1.1 million second lien loans used to avoid government or private insurance. From 2008-2014 second liens were never more than 3.5 percent of the market and they were 1.8 percent in 2014 (Bhutta et al. (2015)). This implies that our own pricing data covers at least 50 percent of the high CLTV market.

that appear in rate filings and published rate sheets. With the CoreLogic data we compute the dollars of originated 30 year fixed rate home purchase loans for each product type in each year. We limit this exercise to Full Doc mortgages with PMI. The data cover the years 1999-2014.The quantity of each product is the total dollar value of the mortgages in the sample with corresponding characteristics.[17]

The two dimensions of risk across the 77 Full Doc products makes it difficult to rank the products by risk. Instead, we divide the products into four risk levels, based on findings in Lam et al. (2013) regarding foreclosure rates by FICO score and LTV in the financial crisis. These look at a coarser subset of these characteristics (four FICO scores and six LTV values), but by interpolation we get the partition in Figure V. We categorize products by their default rates in the stress of the financial crisis. Rates exceeding 7 percent are categorized as "Very High Risk"; 5-7 percent are "High Risk"; 3-5 percent are "Medium Risk"; and rates below 3 percent are "Low Risk." While this classifies products by their ex-post performance, we are only making categorical use of the data. Even if the ex ante assessments differed from the ex post performance, it is reasonable to assume that the ranking was similar.[18]

The result is an index of premiums for the four risk categories based on 77 products in all. Market shares in 1999 serve as starting weights for the premiums in constructing the index. The market shares used to calculate the four indexes are depicted in Figure VII. We use chain-weighting (the "Fisher ideal" index) to construct price indexes for each risk category going forward to 2014. This approach, the same one used by the Bureau of Labor Statistics since 1996 to measure US inflation, is robust both to substitution effects and product entry and exit. Fisher indexes (Fisher (1922)) are the geometric mean of two fixed-weighted indexes: a Laspeyres index (which uses the weights of the starting period) and a Paasche index (which uses the weights of the ending period). When products disappear, we use the regression imputation for the first year of the

[17]In robustness checks, not shown, the results were qualitatively similar when we included Low Doc and No Doc mortgages. They were also similar when we used only the covered portion of the mortgage rather than the balance at origination to measure mortgage size.

[18]Pinto (2014) does a similar exercise with different thresholds.

disappearance.[19]

The results appear in Figure VI. This exercise shows broad trends in PMI pricing with a methodology that controls for selection of borrowers into mortgage products and allows for product entry and exit. Most of the price variation is in the "Very High" category. There is a modest increase over 2005-2008, followed by a sharp jump in 2009-2011, then a decline in 2012 back to trend. The Low and Medium Risk products actually decline in price modestly after 2008. Later in Section 3 we will show how the pooling of disparate risks before 2008 resulted in high-risk products being underpriced and low-risk products being overpriced.[20] The spreading that we observe in this index, and that we also saw in Figure I is a reflection of this shift from pooling to separating by FICO scores.

The large jump in 2009-11 is consistent with a cyclical response to increased default risk during a downturn: Borrowers with a given FICO-LTV combination are more likely to default during a recession. While the "Very High" index had already increased slightly from 2005-2007, the most glaring change is that the price is some 65 basis points higher in 2013 than it was in 2005 at a similar point in an expansion (several years out from a cyclical trough). The spread relative to "Low Risk" products increased by more than that.

Table IV provides a snapshot of index values in 2005 and 2013.[21] In Figure VII we see large declines in market shares of the "Very High" risk products, with most of the slack taken up by the "Medium" risk products. This is broadly consistent with our claim that the price mechanism played an important allocative role in mortgage markets both pre- and post-2008. But we first need to consider the large role played by government insurance.

[19]We similarly use the regression imputation for the price in the year before the product appears. We also constructed a simple Laspeyres index with 2005 weights and obtained qualitatively similar results, though with stronger growth in the price of high-risk products.

[20]The mortgage insurers were not alone in getting the ordering of risk correct while making mistakes about the level of risk. Ashcraft et al. (2010) shows that rating agencies underestimated the risk of the riskiest products pre-crisis. The pattern of mispricing we find in mortgage insurance is similar. The rating agencies, like the mortgage insurers, similarly get the ordering of products with respect to risk is roughly correct, with worse ratings (here PMI pricing) associated with riskier mortgages.

[21]Our index values are available upon request.

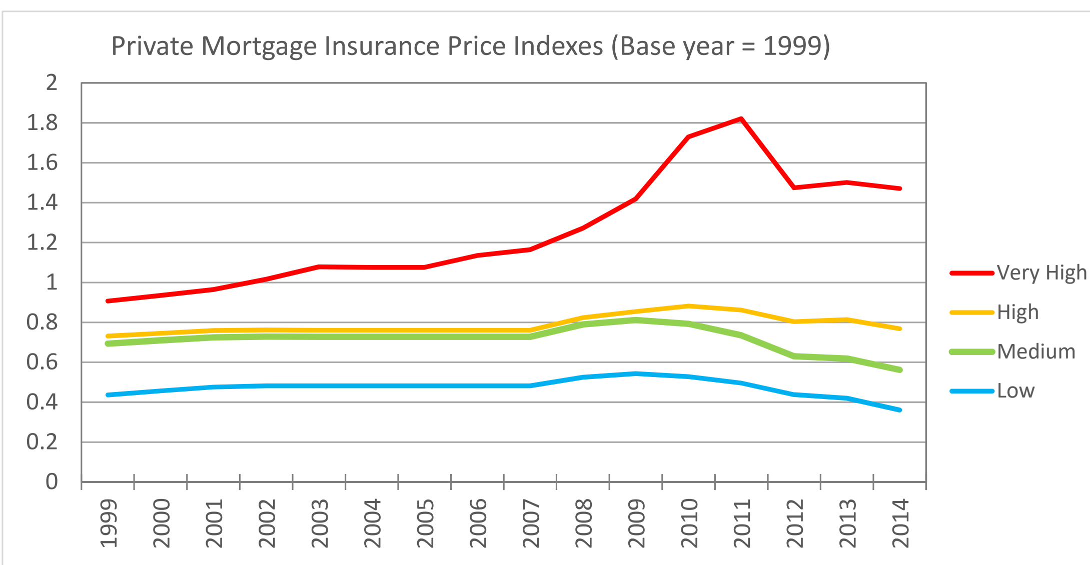
Figure V
Cumulative US Foreclosure Rates in the Financial Crisis by Loan LTV and Borrower FICO Score

Max LTV (%) \ Min FICO Score	760	740	720	700	680	660	640	620	600
103	*4.98*	6.43	*8.30*	10.16	*11.84*	13.51	*16.64*	19.77	23.49
100	4.48	5.80	7.52	9.24	10.77	12.30	15.23	18.16	21.66
97	*3.97*	5.17	*6.74*	8.31	*9.70*	11.08	*13.82*	16.55	19.83
95	*3.22*	4.20	*5.49*	6.77	*7.91*	9.05	*11.36*	13.66	16.43
90	*2.64*	3.44	*4.49*	5.53	*6.47*	7.41	*9.32*	11.22	13.51
85	*2.20*	2.85	*3.70*	4.54	*5.32*	6.10	*7.65*	9.20	11.06
80	*1.39*	1.79	*2.30*	2.81	*3.30*	3.79	*4.72*	5.65	6.76

Very High Risk
High
Medium
Low Risk

Source: Lam et al. (2013) The numbers (non-italicized) are the "cumulative foreclosure rates" for Full Doc products in Lam et al. (2013). The italicized numbers are interpolated.

Figure VI
Chain Weighted PMI Price Indexes



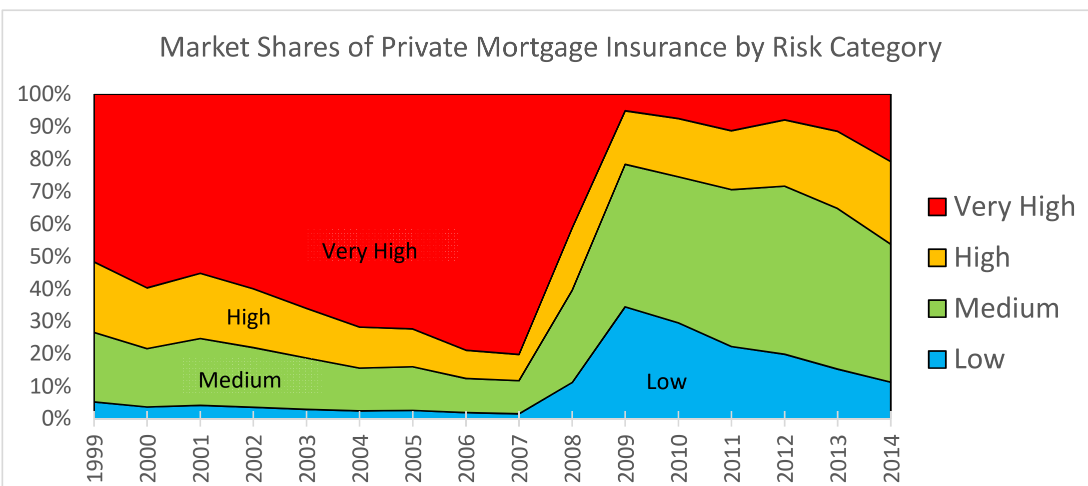
Source: Authors' calculations

Table IV
Average Premiums by Risk Level

Risk level	2005	2013
VH	1.76	2.40
H	0.84	0.88
M	0.83	0.73
L	0.36	0.32
Aggregate	1.45	1.49

*Units are in percentage points per year paid monthly. Source: Authors' estimates

Figure VII
Shares of Privately Insured Mortgages by Risk Level



Source: CoreLogic and Authors' Calculations

2.2 Incorporating Government Insurance

Between 2007 and 2010 there were large shifts in market share from private to government insurance. We have seen that for many higher-risk products, FHA insurance is considerably cheaper than PMI. Thus even though Section 2.1 shows that PMI became much more expensive for high-risk products, borrowers could substitute into FHA loans in response. Figure VIII depicts the market shares of insured (private and government) loans in the CoreLogic sample of 30-year fixed rate mortgages from 1999-2014, by risk category. Government-insured loans are represented by the shaded areas within each category. The FHA's share was always greater in the higher-risk categories, but jumped during the crisis and remained high thereafter. As noted, FHA insurance even became competitive for low-risk borrowers at that time, and we see some government-insured loans even in that category by 2008. If we now look at market shares for the combined private and government insured mortgages, we see a strikingly different picture from that of Figure VII.

The large move into government-insured loans occurred as (and to the extent) the insurance became cheap relative to PMI. Accounting for this substantially alters the price index for mortgage insurance. For this exercise, rather than treat government insurance as a set of products distinct from PMI, we treat it as similar to PMI (insurance for a loan with a particular FICO-LTV combination), albeit not necessarily a perfect substitute (as discussed in Section 1.2). We set the price of the composite product as a market share-weighted average of PMI and FHA premiums for the product's FICO score and LTV. We then construct a chain-weighted index depicted in Figure IX.

The resulting indexes exhibit a similar pattern to those in Figure VI until 2007, but then the large shift of the "Very High" risk mortgages from private into cheaper government insurance in the years 2008-2010 results in a drop in the cost back to 2001 levels. Then from 2010 to 2013 the cost of FHA insurance increases substantially. This leads to a higher cost of mortgage insurance for all but the low-risk category, as well as to a moderate recovery of PMI's market share in these categories (as seen in Figure VIII). This decline in the the cost of credit risk during the recession is surprising, but is driven entirely by FHA pricing. Premiums on PMI, as seen above, continued to rise, and FHA insurance became attractive by comparison. By 2014, once the turbulence of the crisis and recession years had receded, the cost of mortgage insurance rises substantially for all but the low-risk category. Spreads also widen relative to the pre-crisis years.

These indexes document the behavior of mortgage insurance premiums over this time period. This pattern of pricing is insufficient to establish that credit risk was underpriced pre-crisis. The indexes do not control for variation in the quantity of credit risk over time, either in aggregate or within each of the four categories. To establish mispricing we need to document to what extent did the cost of mortgage insurance vary—in the cross section or over time—because the amount of credit risk changed, versus a change in price of a given amount of credit risk. We interpret the latter as evidence of mispricing. For example, if two products with observably different default risks have the same premium, at least one of them must be mispriced.

3 Quantifying the Mispricing of Mortgage Insurance

In this section we undertake a quantitative analysis of the pricing (and apparent mispricing) of individual products as defined by a partition of FICO scores and LTV ranges. We are interested in comparing the pricing of these products before and after the financial crisis. For tractability, we focus only on "Full Documentation" products with LTVs between 80 and 100 percent, and with FICO scores of 575 or higher. "Low Doc" and "No Doc" products virtually disappeared after 2008. Focusing on a narrower set of products reduces our reliance on imputed prices. We are left with 50 products: 10 FICO bins and 5 LTV bins. Even some of these very nearly disappeared in the wake of the 2008 crisis: Specifically, market shares (in dollar value of loans) of nine of the products with FICO below 620 fell to less than 0.05%. Even so, we are able to price them accurately because they were almost entirely government-insured.

We focus on the years 2005 and 2013 as representative of the pre-crisis boom and post-crisis regimes. In doing so, we intentionally avoid the most volatile movements in insurance prices and product shares during the 2007-2009 period. Given the dramatic shift into lower-priced government insurance in those turbulent years, we would expect to find evidence of a a decline in the price of credit risk from that alone. But the years around 2008 were exceptional, arguably a transition between two pricing regimes, and consequently not ideal for the purposes of this paper. By contrast, both 2005 and 2013 were four years into an economic expansion, and in the middle of periods in which premiums and products were relatively stable. Each 12-month sample period contains well in excess of 100,000 insured fixed-rate mortgages.

The premiums on private insurance pre- and post-2008 provide a good illustration of the puzzling treatment of risk that we document in more detail below. PMI rates from 2005 and 2013 are depicted in Table V. In both periods they are increasing in LTV, but in 2005 they are completely flat with respect to FICO scores all the way down to at least 660, and in most cases to 640.[22]

In Section 3.3 we examine whether the 2005 prices were allocative, meaning whether they influenced quantities in the market in a manner consistent with a downward-sloping demand curve. The alternative is that higher-risk borrowers were rationed or otherwise screened from the pool. We find that the 2005 prices were allocative, that the mispricing of risk did distort the market. Under priced high-risk products attracted relatively more borrowers and overpriced lower-risk products relatively less than they would have in the absence of mispricing.

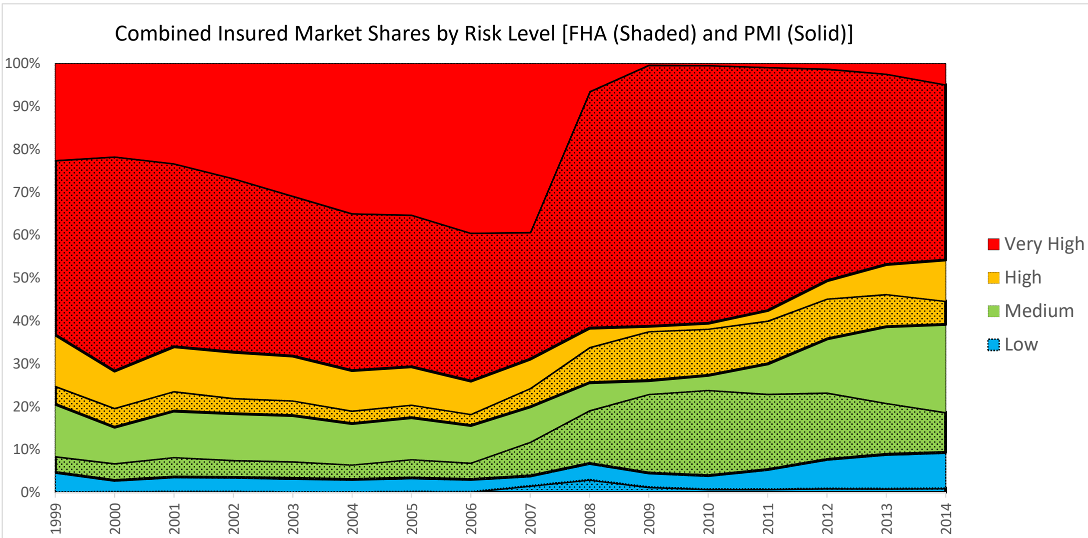
3.1 A Simple Lending Model

We begin with a two-period model of mortgage lending.[23] Suppose an agent purchases a house at $t = 0$, with a value normalized to one. At $t = 1$ the house is hit by a multiplicative value shock x, observed costlessly only by the owner, with mean $1+\mu$, and distribution function $G(x)$. The house depreciates at a deterministic rate δ. We assume x has bounded and compact support on $[\underline{x}, \bar{x}]$. Thus the expected value of the house at $t = 1$ is $1 + \mu - \delta$. We assume lenders are risk-neutral.

[22] This is the "spreading out" of PMI rates depicted in Figure I, but in more detail

[23] While a multi-period could provide a richer set of possibilities and greater realism, the two-period assumption is adequate for our purposes. Since the mortgage insurance premiums are annual, our approach essentially pulls out one representative year of a multi-year problem. The basic approach can be extended to incorporate a longer time horizon, but the results are similar.

Figure VIII
Shares of All Insured Mortgages by Risk Level and Insurer



Source: CoreLogic and Authors' Calculations

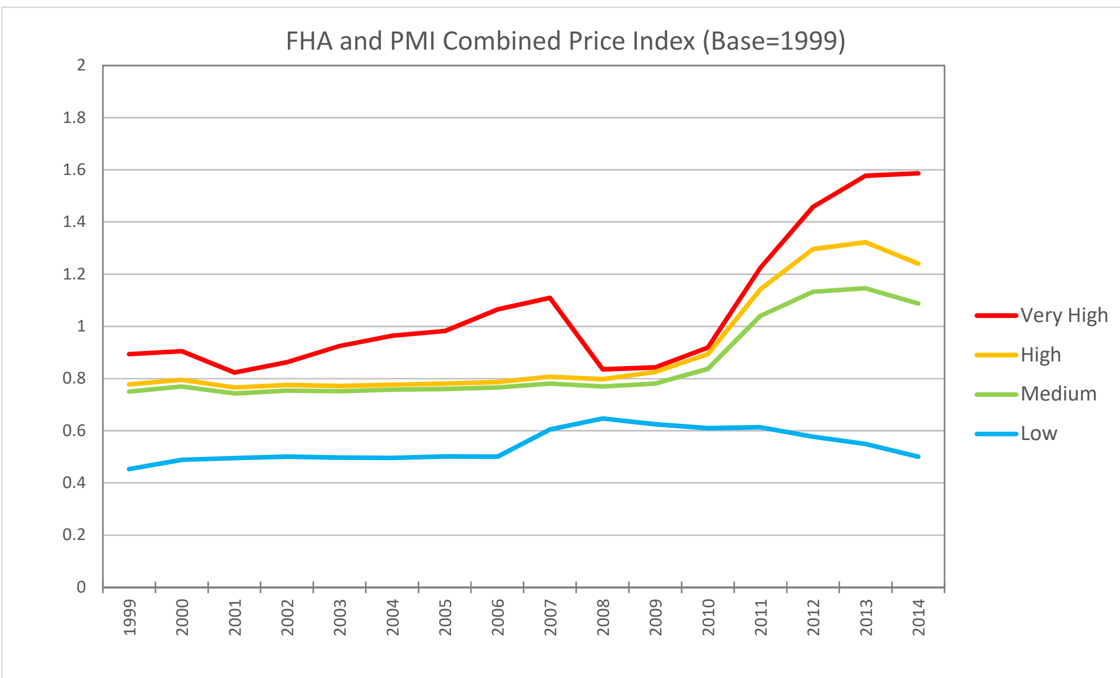
Table V
Private Mortgage Insurance Rates, 2005 versus 2013

		minimum FICO Scores									
	Max LTV	760	740	720	700	680	660	640	620	600	575
	85	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.41	0.53	0.72
	90	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.68	0.90	1.22
2005	95	0.79	0.79	0.79	0.79	0.79	0.79	0.79	1.00	1.32	1.80
	97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	1.42	1.88	2.57
	100	1.07	1.07	1.07	1.07	1.07	1.07	1.34	1.58	2.10	2.87
	85	0.29	0.32	0.32	0.38	0.38	0.44	0.44	*0.65*	*0.85*	*1.28*
	90	0.45	0.49	0.49	0.62	0.62	0.76	0.76	*1.06*	*1.35*	*2.03*
2013	95	0.61	0.67	0.67	0.94	0.94	1.20	1.20	*1.69*	*2.17*	*3.26*
	97	0.99	1.04	1.04	1.25	1.25	1.53	1.53	*2.32*	*3.01*	*4.59*
	100	*1.05*	*1.11*	*1.18*	*1.34*	*1.47*	*1.72*	*2.02*	*2.48*	*3.23*	*4.94*

*Units are in percentage points per year paid monthly. Rates in italics are imputed.

Source: WI and NC mortgage insurer regulatory filings, authors' calculations

Figure IX
Combined FHA and PMI Insurance Price Index



Source: CoreLogic and Authors' Calculations

To purchase the house, the agent borrows $z \in (0,1]$ at an interest rate of $\rho(\cdot)$, where ρ may depend on z and other observable characteristics. For concreteness we assume that interest $(z\rho)$ is always repaid. Repayment of the principle (z which due to the normalization of the house price is also the CLTV) is at the discretion of the borrower. The principle is secured by only by the value of the house (the loan is non-recourse).

We adopt the costly state verification framework of Townsend (1979) in which the lender must spend k to "verify" and recover the value of the collateral, which he only does in the event of default. Consequently, if a default occurs, the lender recovers $\min\{z, x-k\}$. The verification cost k includes the legal and other transactions costs involved in foreclosures, short sales, and other means lenders have of extracting value on default.

Normally, with one-period debt the optimal default decision is simple: default if and only if $x < z$, i.e. the house is "under water" on the loan. On both theoretical and empirical grounds such a default rule is inadequate and unrealistic. In multi-period models with default costs, it is generally not optimal to default whenever the house is under water, due to the option value of waiting. Moreover, empirically, default behavior does not generally conform even to richer optimal default models.[24] Defaults are triggered not just by the value of the collateral, but also on idiosyncratic individual characteristics (which credit scores attempt to measure), as well as other shocks such as declines in income or health. Mortgage insurance premiums reflect this reality, so our model must as well.

As a consequence, we model default decisions so that, with the appropriate choice of parameters, we can rationalize the observed premium data in 2013. Although this is admittedly ad hoc, for our purposes it is enough to get the conditional default probabilities (and losses conditional on default) approximately correct. We assume the relationship between the house price shock (x) and the default probability takes the form of a monotonic function of x relative to z. Let $H(x; z, \xi)$ be the probability of repayment of a loan with LTV z, where ξ is a characteristic of individuals to allow for heterogeneity, akin to a FICO score. This is a generalization of the

[24] Vandell (1995) surveys the extensive evidence of non-ruthless residential mortgage defaults.

simple default rule in which H is a step function that jumps from 0 to 1 at $x = z$. We assume $H : [\underline{x}, \bar{x}] \to [0, 1]$ is weakly monotonically increasing, and that it is monotonic in ξ, i.e. that $\xi < \xi' \implies H(x; z, \xi') < H(x; z, \xi)\ \forall x, z$.

We can motivate this by supposing that an individual's default decision depends on, in addition to equity $x - z$ and the characteristic ξ, the realization of an additional stochastic variable such as income, which we can denote by y. With this interpretation, $H(x; z, \xi)$ gives the probability that y exceeds the individual's threshold for repayment of the loan.

Let Π denote expected revenues for a representative risk-neutral lender. Let r denote the risk-free rate of return. Competition among risk neutral lenders determines ρ by equating the expected return on mortgage lending to r: $\Pi = (1 + r)z$. We allow for a servicing cost αz in addition to the risk free rate r. These assumptions imply that the mortgage interest rate ρ should satisfy:

$$\begin{aligned} z(1+r) = (\rho - \alpha)z + z\int_{\underline{x}}^{\bar{x}} H(x; z, \xi)dG(x) \\ + \int_{\underline{x}}^{z+k} (x-k)(1 - H(x; z, \xi))dG(x) + z\int_{z+k}^{\bar{x}} (1 - H(x; z, \xi))dG(x) \end{aligned} \tag{1}$$

In other words, the lender gets the interest payment (ρz), and pays the servicing cost (αz). If the borrower repays, the lender gets z, and if the borrower defaults the lender gets $\min\{z, x - k\}$.

To implement the model we choose functional forms for $G(x)$ and $H(x; z, \xi)$. For the sake of tractability (following Acikgoz and Kahn (2016)), we use a variant of the Beta distribution for x, the Kumaraswamy (see Jones (2009)), that has a closed form density and cumulative distribution function.[25] For the probability of repayment function (H) we use a logistic specification:

$$H(x; z, \xi, \psi) = \frac{e^{\psi(\frac{x}{z}-\xi)}}{1 + e^{\psi(\frac{x}{z}-\xi)}} \tag{2}$$

where $\xi \in (0, 1]$ and $\psi > 0$. A larger ψ implies a greater sensitivity of defaults to housing returns in the neighborhood of $x = \xi z$. In the limit, as $\psi \to \infty$, H becomes a step function. With this functional form a smaller value of ξ represents a better "credit score."[26]

Mortgage insurance does not generally cover all losses from defaults, so to model the premiums we need to be more specific. In the event of default, the lender recovers $\min\{z, x - k\}$. PMI is designed to cover some of the gap between what the lender recovers and the principal on the mortgage. We denote the coverage rate, as described above in Table I, by $\chi(z)$. This means that the insurer is actually only liable for

$$\min\{\chi z,\ \max\{0,\ z - x + k\}\}.$$

The box below provides examples of the PMI payout and lender losses in three scenarios.

[25] For the $[0, 1]$ domain the density κ and the distribution function K are

$$\kappa(x) = abx^{b-1}(1 - x^a)^{b-1}$$
$$K(x) = 1 - (1 - x^a)^b$$

where $a, b \geq 0$. If both a and b exceed one, the density will have the usual hump shape (when $a = b = 1$ the distribution is uniform on $[0, 1]$). It is straightforward to change the domain to $[\underline{x}, \bar{x}]$, and the choice of α and β implies the mean and standard deviation of x. We have obtained similar results, not shown, using lognormal distributions.

[26] Although $\lim_{z\to 0} H = 1$, to implement this numerically we have to add a small constant (through trial and error we find that 0.01 works well) to the denominator of $\frac{x}{z}$ in the H function so that the computer can handle values of z near zero.

Three Examples of PMI Risk Absorption After Borrower Default on 85 LTV ($z = 0.85$) **Loans with Standard Coverage** ($\chi = 0.12$)

1. Suppose $k = 0.10$, and a default occurs with $x = 0.7$ (a house price decline of 30 percent). The lender directly recovers 0.6. Additionally, PMI pays $\chi \cdot 0.85 = 0.10$. Therefore the lender recovers only 0.7 rather than the full 0.85 principal. PMI only covers 40 percent of the lender's loss.

2. If instead default occurred with $x = 0.8$, the lender would directly recover 0.7. PMI would again provide 0.10. PMI would then cover two-thirds of the loss.

3. If default occurred with $x = 0.9$, the lender would recover 0.8. PMI would cover the entire loss of 0.05.

We model the mortgage insurance premium $p(z; \xi, \psi, \alpha)$ as the expected value of the insurance payout ($\max\{0,\ z - x\} + k$) over the domain of losses ($[0,\ \chi z]$) plus a servicing fee (αz). With the conditional default probability modeled as $1 - H(x; z, \xi)$, we have

$$\begin{aligned} p(z; \xi, \psi, \alpha) \quad &= \quad \alpha z + \chi(z) z \int_{\underline{x}}^{z(1-\chi)+k} (1 - H(x; z, \xi, \psi)) dG(x) \\ &+ \quad \int_{z(1-\chi)+k}^{z+k} (z - x + k)(1 - H(x; z, \xi, \psi)) dG(x) \end{aligned} \tag{3}$$

The upper limit of $z + k$ on the second term reflects that if $x \geq z + k$, a default will still allow the lender to recover z, so there will be no liability for the PMI provider. Apart from the G and H functions, these premiums are model-free, in the sense that they can be conditioned on (z, ξ, ψ) without regard to how z is chosen. Therefore, we can calibrate ξ and ψ so that the implied default rates fit the premium data. We undertake this exercise in the next section.

3.2 Quantifying the Mispricing of Insurance

We next use the framework developed in the preceding section to measure the extent of mortgage insurance mispricing prior to 2008. We assume, as a benchmark, that post-crisis PMI rates satisfy the rational expectations hypothesis: Given information available at the time, they accurately reflect default risk. On this basis, we choose the parameters of the repayment function $H(x, z; \xi)$ to fit the 2013 premium data conditional on FICO scores and loan-to-value ratios. We apply this model to the pre-2008 market for PMI, assuming that the parameters of H are the same as in 2013, but allowing for differences in expectations about home price appreciation.[27]

Because our benchmark for quantifying 2005 mispricing is based in part on 2013 data, we need additional justification for our claim that the 2005 prices were wrong given information available at the time. To accomplish this we also examine loan performance data from 2005 and earlier. We show that FICO scores in the ranges that were not priced differentially (that is, were pooled together) in 2005 had indeed experienced notably different default rates over the previous five years.

We consider 35 LTV-FICO combinations: Five LTV categories denoted by z: $[0.801-0.85]$,$[0.851-0.90]$, $[0.901-0.95]$, $[0.951-0.97]$, $[0.971-1.00)$, and seven FICO categories: $[575-599]$, $[600-619]$, $[620-639]$, $[640-679]$, $[680-719]$, $[720-759]$, and $[760-900]$. These are standard ranges within which mortgage insurance premiums are constant. We only consider full documentation loans, so our data set for this exercise consists of premiums (either observed or imputed) for as many as 35 products in 2005 and 2013 $\{p_{ij}\}$ $(i = 1, .., 5;\ j = 1, ..., 7)$ corresponding to the different LTV and FICO score combinations.

We fit the parameters of the $H(x; z, \xi, \psi)$ function to target the average observed PMI rates from 2013. We set the parameters of the distribution of G in 2013 (denoted G_{13}) so that $\mu = 0.035$ based on the expectations surveys described in Case et al. (2012). We set $\delta = 0.025$ based on Harding et al. (2007), who measure the average gross depreciation of owner occupied housing at 2.5 to 2.9 percent per year using survey data in the 1983$-$2001 period. Recall that a, b, $\overline{x}$ and $\underline{x}$ are parameters of the G distribution. We set the standard deviation of x at 0.10 based on Flavin

[27] The assumption that H is time-invariant presumes that FICO scores are a consistent predictor of default given home equity. It does not require that default rates given FICO score be time-invariant, since both ex ante and ex post house price distributions will vary over time, only that for fixed expectations about house prices, the relationship between default and equity for a given FICO score is time-invariant.

and Yamashita (2002), and we also set $k = 0.10$ based on Cutts and Merrill (2008).[28]. The results are robust with respect to modest variations in these parameters.

Table VI
Table of Model Parameters, Calibrations, and Sources

Parameter	**Value**	**Source**	**Purpose**
k	0.100	Cutts and Merrill (2008)	Monitoring cost per unit of housing
μ_{2005}	0.035	Case et al. (2012)	Expected house price appreciation
μ_{2013}	0.070	Case et al. (2012)	Expected house price appreciation
δ	0.025	Harding et al. (2007)	Depreciation
σ	0.1	Flavin and Yamashita (2002)	Standard deviation of house price value shock
ξ_i	0.576 - 0.961	Calibrated to data	Controls probability of individual default, like FICO
α	0.227	Calibrated to data	Fixed component of PMI premium, servicing fee
ψ	6.987	Calibrated to data	Shape parameter for probability of repayment of a loan (H)
z	0.800-0.950	Calibrated to data	LTV at origination
a	3.986	Implied by other parameters	Kumaraswamy distribution parameter
b	33.084	Implied by other parameters	Kumaraswamy distribution parameter
$\underline{x}$	0.65	Implied by other parameters	Lower support of x
$\bar{x}$	Not sensitive	Implied by other parameters	Upper support of x

Because we do not have actual data for 2013 on premiums for the two highest LTV ranges (those above 95 percent) and lowest FICO scores (those below 640), we first fit the parameters of the H function only on the observed premiums. We choose $\xi_1, ..., \xi_4$, α, and ψ to minimize

$$(\sum_{i=1}^{4}\sum_{j=1}^{4}(p(z_{0i}, \xi_j, \psi, \alpha) - p_i)^2) \tag{4}$$

We find values of the six parameters that minimize (4) given the 16 data points. This results in estimates of ψ, α, and four FICO shift parameters $\xi_1 - \xi_4$ corresponding FICO scores $[760, 900]$, $[720, 759)$, $[680, 719)$, and $[640, 679)$. We then keep those values of ξ and α, and estimate the three remaining FICO parameters using the imputed premiums from our regression-based imputation

[28] Flavin and Yamashita (2002) actually estimate a cross-sectional standard deviation of house price changes somewhat larger than 0.10, but we find a slightly better fit of the premium data with $\sigma = 0.10$

as described in Section 1.1.[29]

The resulting parameter values are shown in Table VII. The model fits the sixteen observed premiums very well, with a root mean squared error (RMSE) of 6 basis points. When we extend the fit to the imputed data the overall RMSE is 26 basis points, which mainly reflects the fact that the imputed premiums are for high-risk products, and therefore an order of magnitude larger than the observed ones.

Using the 2013 pricing as a benchmark, there are three potential explanations for the 2005 PMI premiums:

1. Parameters (either of $H(x; z, \xi, \psi)$ or of $G(x)$) changed between 2005 and 2013, specifically so that FICO scores were uninformative about default risk in 2005 so long as they were at least 640;

2. There was rationing of credit to borrowers according to FICO scores, or selection among them based on criteria such as debt-to-income or other qualities somehow not reflected in (but correlated with) FICO scores.

3. Borrowers with observably different credit risks were pooled together, implying that credit risk was mispriced, and potentially resulting in adverse selection.

We address each of these in turn: the first now, the second two in Section 3.3. First, regarding parameter change: We could mechanically fit alternative parameters for $H(x; z, \xi, \psi)$ to the 2005 data. However, this would imply that default behavior conditional on borrowers' equity positions was for some reason very different in 2005 than 2013. It also would imply that this behavior was believed in 2005 to be identical for all FICO scores 640 and above, even for high-LTV loans. This seems implausible. The purpose of the FICO score is to capture default risk, so it would be surprising if default risks did not vary over such a wide range of scores, in 2005 as well as in 2013.

[29] We have also estimated the parameters in other ways. First, When we treat the imputed premiums as if they were observed, i.e. estimating nine parameters with 35 data points, we obtain very similar results (not shown). Second, we try a variety of alternative imputation methods, as detailed in Appendix A.I, we also obtain very similar results.

Fortunately, high-quality data are available to address the question of the relationship between FICO score and credit risk as of 2005. We examine default behavior in the public use Fannie Mae Single-Family Loan Performance Data. We calculate cumulative default rates (through the end of 2005) for loans that Fannie Mae acquired in the year 2000, by FICO and LTV groups, using the standard 20 or 25-point-wide grids in FICO and 5 percent-wide grids in LTV. Default is measured, in line with common practice, as a loan being 180 or more days delinquent (White (2008), Calem and Wachter (1999)). This provides a sense of what a mortgage insurer, operating in 2005, would have known about the default risks of their insured mortgages, conditional on LTV and FICO score.[30]

Table VIII presents the results of this analysis. As expected, cumulative default rates on mortgages monotonically decrease as loan FICO scores increase and as LTV decreases (aside from a couple of outliers in cells with a relatively small number of mortgages). The sample includes all 30-year, fixed-rate mortgages purchased by Fannie Mae in the year 2000, over 700,000 mortgages in total. Of these, we eliminate those with missing FICO scores (about 16,000) or scores below 575 (5,000). A little more than half of the remainder have LTVs of 80 or below.[31] This leaves some 315,000 mortgages that we can presume (because they have LTV higher than 80 and were purchased by Fannie Mae) have private mortgage insurance.

The performance depicted in Table VIII occurred under generally benign conditions with rising house prices (notwithstanding the brief 2001 recession), so it should be informative about risks in 2005—even if insurers unrealistically believed such conditions would continue unabated. The data show that loans with FICO scores of $[640, 659]$ had an overall default rates nearly ten times the rate of $[760, 900]$ loans (4.07% versus 0.43%), yet mortgage insurers charged them identical

[30]Other date cutoffs are sometimes used in the literature. Cowan and Cowan (2004) use 90 days, for what they call a "less stringent measure" of default. We get similar results with either measure. Elul et al. (2010) use a 60-day delinquency definition of default. However, White (2008) show that foreclosure starts in less than half of mortgages which are less than 180 days delinquent. Calem and Wachter (1999) show that the FHA generally initiates foreclosure only after 180 days of delinquency. Because a material number of loans flagged as defaulted under the less stringent measures may cure, we adopt the more stringent measure.

[31]The table includes default rates on mortgages in the 70 to 80 LTV range for comparison, even though such loans are not typically insured by PMI or the FHA.

Table VII
2013 Estimated Model Parameters

ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6	ξ_7	α	ψ
0.576	0.592	0.644	0.691	0.762	0.828	0.956	0.227	6.987

Source: Authors' estimates.

Table VIII
Observed Default* Rates for 2000 Vintage Mortgages Through 2005

FICO Group	(70,80]	(80,85]	(85,90]	(90,95]	(95,97]	Average by FICO†
[575,599]	4.06	5.65	5.77	7.79	12.71	7.82
	(2,560)	(230)	(1,022)	(1,824)	(543)	(3,619)
[600,619]	3.42	3.54	5.52	6.76	7.92	6.37
	(4,477)	(395)	(2,082)	(4,173)	(821)	(7,471)
[620,639]	2.56	5.00	3.97	5.74	7.77	5.45
	(8,719)	(961)	(4,337)	(10,320)	(1,712)	(17,330)
[640,659]	1.63	2.39	2.96	4.44	5.71	4.07
	(13,278)	(1,216)	(6,767)	(15,194)	(2,451)	(25,628)
[660,679]	1.25	1.03	1.98	2.97	4.17	2.75
	(18,599)	(1,556)	(8,617)	(18,685)	(3,720)	(32,578)
[680,699]	0.75	1.32	1.38	2.08	2.70	1.94
	(24,121)	(1,741)	(10,050)	(21,142)	(4,967)	(37,900)
[700,719]	0.44	0.59	0.88	1.46	2.27	1.38
	(29,644)	(2,028)	(10,889)	(22,543)	(5,778)	(41,238)
[720,739]	0.26	0.57	0.52	1.00	1.31	0.88
	(35,943)	(2,270)	(12,232)	(23,419)	(5817)	(43,738)
[740,759]	0.14	0.22	0.39	0.59	1.25	0.59
	(46,552)	(2,778)	(14,128)	(24,093)	(5,464)	(46,463)
[760,900]	0.10	0.11	0.30	0.46	1.04	0.43
	(89,837)	(4,494)	(20,774)	(29,168)	(5,195)	(59,631)
Average	0.54	1.01	1.25	2.07	2.81	1.86
by LTV	(273,730)	(17,669)	(90,898)	(170,561)	(36,468)	(315,596)

Note: For each FICO group, the top number is the delinquency rate in percent, the number in parentheses below it is the number of mortgages in the cell.

* A mortgage is classified as in default if it is 180 or more days delinquent.

†Volume weighted average only over mortgages with $LTV \geq 80$ percent.

Source: Fannie Mae and authors' calculations premiums in 2005.[32] By contrast, the average default rate for LTVs exceeding 95% was a more modest 2.81%, versus 0.54% for those with LTVs in the 70 to 80 percent range. Insurers' disparate treatment of risk, with premiums varying by LTV but not by FICO scores (except for sub-640 scores) is thus difficult to rationalize.

This strong relationship between FICO scores and default risk suggests that the flat pricing of insurance cannot be rationalized by beliefs that FICO scores did not help predict defaults. Thus we forgo the unappealing assumption of ad hoc changes in default behavior to "explain" the 2005 premiums. Instead we allow for different beliefs about house price appreciation (as expressed by the G function), common to borrowers and lenders, between 2005, in the midst of the boom, and 2013.[33] Expectations of higher appreciation rates during the boom could justify a flatter (though not entirely flat) structure of premiums with respect to FICO scores. This goes some way toward justifying the 2005 rates, and thereby makes the case for mispricing more of a challenge.

One data justification for assuming greater optimism in 2005 is that by most measures overall premiums were lower. This is apparent from Figure IX, which shows that that while low-risk products are priced similarly in 2005 and 2013, the premiums for the other three risk categories are substantially higher. But this ignores the important distinction between the change in the quantity and price of risk. So to discipline this exercise we rely on Case et al. (2012), who find using surveys that subjective expectations of home price appreciation were about 3.5 percentage points higher in 2005 than in 2013. In line with their results, we set μ, the mean of x, to 1.07 in 2005 and 1.035 in 2013.

To implement this change in beliefs, we first find parameters of the x distribution for the 2013 baseline that imply a mean of 1.035, a standard deviation of 0.10, and a lower bound consistent with a 2 percent default probability at 80 percent LTV. These three constraints determine three of the four parameters of the distribution, a, b, and $\underline{x}$. We choose the fourth parameter, $\bar{x}$, the upper bound of the support of x, somewhat arbitrarily. So long as it is sufficiently large, the

[32] More precisely, PMI premiums were identical. The composite prices in Table XI differ very slightly due to some borrowers opting for FHA insurance.

[33] Kaplan et al. (2017) argue that such a change in beliefs was crucial in explaining movements in house prices during this time period.

value of the density is indistinguishable from zero for x well below the upper bound. We set it at $1.6 - \delta + \mu - 1 = 1.61$, This results in $a = 3.986$, $b = 33.084$, and $\underline{x} = 0.650$. To obtain the distribution for 2005 we simply shift the support of G by 0.035, to $[0.685, 1.645]$, which increases the mean by 0.035 and leaves the standard deviation unchanged.

Figures X and XI respectively depict the x distributions and implied default probabilities given the parameter estimated for the H function. Table IX shows the result of combining the more optimistic beliefs about price appreciation in 2005 with the repayment probability function H based on 2013 insurance premiums. It indeed results in a flatter profile of premiums across FICO scores compared to the observed 2013 premiums. As a consequence it helps to match the actual 2005 rates: The RMSE is 33 basis points, whereas if we assumed the same beliefs in 2005 as in 2013, the RMSE would be 70 basis points. Nonetheless the model cannot rationalize the flat pricing with respect to FICO scores, particularly for higher LTV loans.

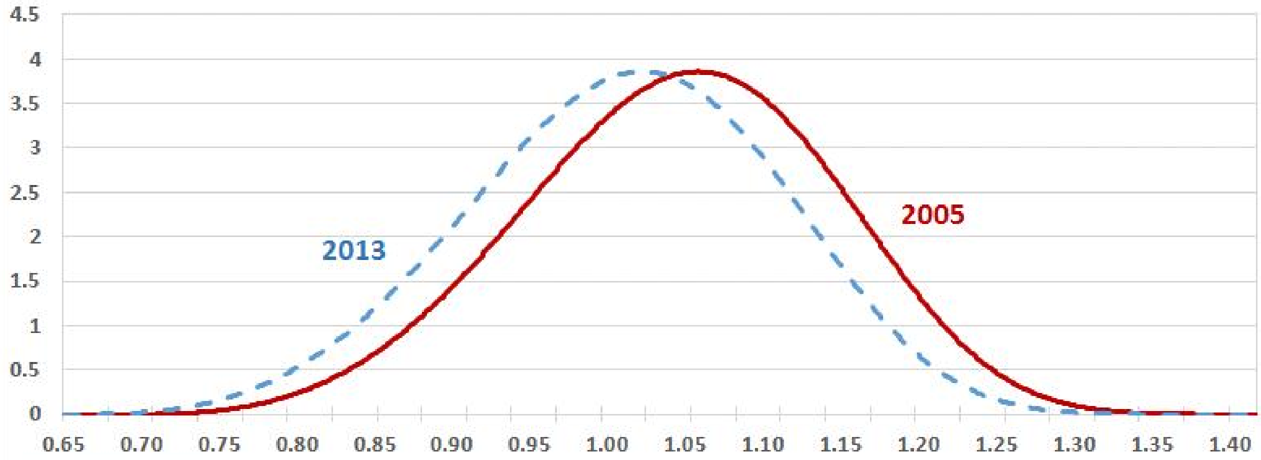
Recall (see Table V) that with one exception[34] actual PMI premiums in 2005 were constant across FICO scores 640 and higher. Those premiums lie in the middle of the range of model-implied premiums in Table IX. This suggests that the 2005 premiums were not systematically biased. Rather, these products were pooled together and charged a common rate conditional only on LTV. As a consequence, borrowers in 2005 borrowers with FICO scores roughly 680 and higher were subsidizing those with lower scores.

To summarize, we conclude that there were systematic pricing errors in PMI in 2005 resulting from insurers charging common premiums across a wide range of observable risk classes. This conclusion is robust to allowing insurers substantially more optimistic beliefs about house prices in 2005 than in 2013. We are unable to rationalize insurers' disregard of meaningful information from their recent experience about credit risk. We further quantify our findings based on the assumptions that, first, mortgage insurance was priced efficiently in 2013; and, second, that individuals exhibited the same default behavior, conditional on realized equity and FICO score, in 2005 and 2013.[35]

[34]That exception is LTV > 97 and FICO score $\in [640, 659]$.

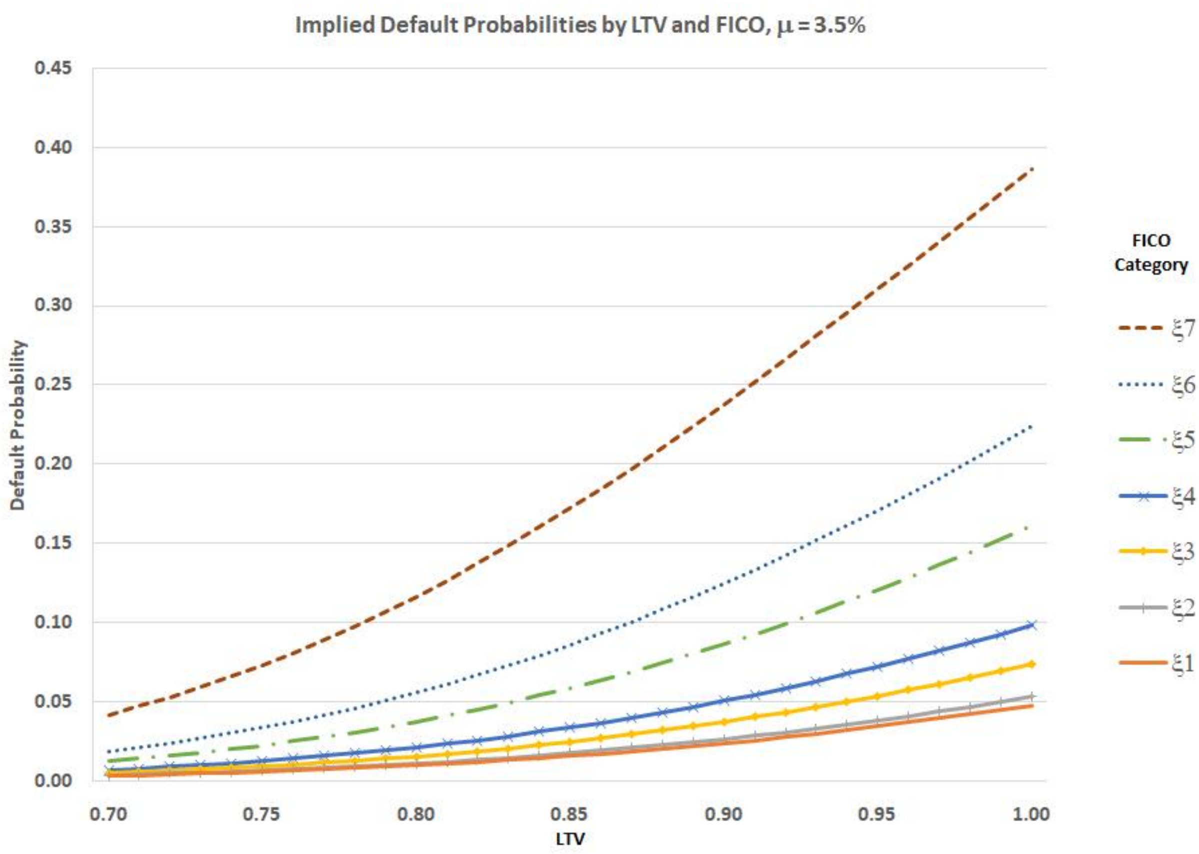
[35]As discussed above, borrowers in 2005 did have second lien mortgages as an alternative to mortgage insurance,

Figure X
Density Functions for the House Price Shock x



Source: Authors' calculations

Figure XI
Implied Default Probabilities by LTV and FICO Groups



Source: Authors' calculations. These default rates are based on the 2013 x distribution with $\mu = 0.035$. ξ_1 is the highest FICO category (most creditworthy) and ξ_7 the lowest (least creditworthy).

3.3 Credit Rationing or Adverse Selection?

A certain amount of pooling is inevitable in insurance products. But charging a common premium to widely disparate risks can result in adverse selection, unless insurers use some mechanism such as rationing to keep the customer pool from deteriorating.[36] Rationing might occur, for example, if insurers were constrained for some reason to charge a common premium to this pool of borrowers. They would then benefit from screening out higher-risk borrowers by refusing to insure their mortgages, or at least subjecting their applications to greater scrutiny, and perhaps to other requirements such as debt-to-income limits.[37]

Of course we can only judge the selection as "adverse" relative to some unobserved counter-factual. We suppose that PMI premiums in 2013, and the resulting distribution of FICO and LTV across privately insured mortgages, is the relevant counter-factual, adjusted for differences in expectations about house price appreciation. Thus if mortgage insurers (or lenders) were compensating for the lack of FICO score pricing in 2005 by rationing or screening, then lending patterns across FICO scores would not be systematically related to our measure of mispricing. Alternatively, if the price mechanism was operative (that is, borrowers in the pool could obtain mortgage insurance at the stated premium regardless of their FICO score), we would expect to find quantities responding to the mispricing that we observe under pooling; the higher FICO borrowers would be subsidizing the lower FICO borrowers, so the composition of the pool would shift toward the lower FICO borrowers.

but we are aware of no evidence that suggests there was self-selection of individuals with lower default risks toward one instrument or the other. Sherlund (2008) finds that borrowers with second liens had higher default rates, but these appear to have been associated with adjustable rate loans. In fact, Park (2016) finds the opposite, that controlling for FICO score and LTV, second liens had higher default rates than insured high-LTV loans.

[36]Conversations with insurers, regulators, and GSE staff have not yielded satisfactory explanations for why insurers charged common premiums over such a wide range of FICO scores. Perhaps it resulted from inertia, reflecting pricing structures from when subprime borrowers were relatively rare and the mix of borrowers was stable. Regulators may have had concerns about disparate impact, though this is belied by the allowance of higher premiums on mortgages with sub-640 FICO scores.

[37]Another possibility is that mispricing by insurers was offset by mortgage originators, either in their rates or lending decisions. But given that insurers have the primary exposure to credit risk, this seems unlikely. In fact, our understanding from discussions with insurance regulators and others from the industry is that, particularly pre-2008, mortgage insurers were passively providing insurance for whatever mortgages lenders were originating, which itself was driven by the standards that the GSEs would accept. See also Bhutta and Keys (2017).

3.3.1 Quantifying Mispricing

To investigate this, we take dollar quantities of mortgages by FICO and LTV from the CoreLogic data, and compute product shares in 2005 and 2013. To make the experiment as clean as possible, we again consider only 30-year fixed-rate mortgages, and include all mortgages with either private or government insurance. The product shares are shown in Table X. Not surprisingly, we see dramatic differences across products between 2005 and 2013. For example, the share of mortgages with FICO scores below 640 was nearly 25 percent in 2005, and barely exceeded 2 percent in 2013. That was largely balanced out by an increase in the share of mortgages with FICO scores of at least 740. The question is to what extent these differences can be explained by differences in insurance premiums.

As with our price index in Section 2, we use a market-share-weighted average of the PMI and FHA premiums (Table XI) to measure the price of insuring each FICO-LTV combination. Whereas in Section 2 we assumed a constant 7-year horizon to convert FHA premiums with their upfront fees to flat annual equivalents, in this exercise we assume a 3-year horizon in 2005 for comparing FHA and PMI premiums. We justify this as broadly consistent with higher expected home price appreciation in 2005, which enable borrowers to refinance sooner and be freed of mortgage insurance. The shorter horizon also helps to make sense of the substantial market share of private insurance even for high LTV products in 2005. As Table III shows, the 7-year equivalent FHA rate at that time was 0.76 percent, well below PMI rates for mortgages with LTVs exceeding 95 percent. A 3-year horizon makes the two types of insurance more competitive.

Comparing these blended rates to the PMI rates in Table V shows that the availability of government insurance as an alternative to PMI effectively caps the premiums on higher-risk mortgages. This is especially evident in 2005 when FHA rates were more broadly competitive with PMI. After 2013, PMI was largely not available or not price competitive for borrowers with FICO score below 640 or LTVs above 97 percent. Government insurance for these mortgages remained available but was more expensive than in 2005.

Although we assume that 2013 PMI premiums are actuarially fair, the blended 2013 premiums

Table IX
Model-implied Premiums with $\mu = .07$ for 2005

Max	Minimum FICO Scores									
LTV	760	740	720	700	680	660	640	620	600	575
85	0.27	0.28	0.28	0.30	0.30	0.32	0.32	0.38	0.44	0.60
90	0.36	0.38	0.38	0.44	0.44	0.50	0.50	0.68	0.85	1.27
95	0.53	0.56	0.56	0.69	0.69	0.84	0.84	1.21	1.56	2.39
97	0.63	0.68	0.68	0.85	0.85	1.04	1.04	1.53	1.98	3.20
100	0.83	0.89	0.89	1.14	1.14	1.43	1.43	2.12	2.75	4.18

* Units are in percentage points per year paid monthly.

Source: Authors' estimates

Table X
Insured Mortgage Product Shares, 2005 and 2013

(a) 2005

Max	minimum FICO Scores										
LTV	760	740	720	700	680	660	640	620	600	575	Total
85	0.77	0.32	0.29	0.34	0.36	0.36	0.34	0.31	0.18	0.14	3.41
90	2.95	1.33	1.30	1.50	1.51	1.50	1.48	1.20	0.64	0.40	13.80
95	3.72	1.93	2.00	2.32	2.64	2.81	2.85	2.26	1.20	0.69	22.42
97	1.05	0.71	0.73	0.85	1.06	1.13	1.13	0.98	0.64	0.47	8.76
100	4.70	3.23	3.74	4.38	5.64	6.84	7.50	7.14	4.59	3.86	51.61
Total	13.19	7.52	8.07	9.39	11.20	12.64	13.31	11.88	7.25	5.55	100.00

(b) 2013

Max	minimum FICO Scores										
LTV	760	740	720	700	680	660	640	620	600	575	Total
85	1.88	0.51	0.38	0.36	0.30	0.25	0.18	0.05	0.01	0.00	3.93
90	8.44	2.64	1.84	1.48	1.02	0.65	0.42	0.09	0.01	0.00	16.59
95	13.33	5.18	4.17	3.02	2.30	1.52	1.14	0.20	0.03	0.02	30.89
97	2.12	0.99	1.04	1.15	1.41	1.45	1.30	0.30	0.04	0.01	9.82
100	5.59	2.98	3.47	4.43	5.87	8.12	6.83	1.32	0.12	0.03	38.76
Total	31.36	12.32	10.90	10.43	10.91	11.98	9.87	1.96	0.21	0.06	100.00

Source: CoreLogic LLMA 2.0. Units are percentages of insured mortgages.

are (potentially) mispriced due to the presence of FHA insurance. We estimate the mispricing in 2005 as the difference between the actual 2005 premiums in Table XI and the fitted values in Table IX. The 2013 mispricing is estimated as the difference between the blended values in Table IX and the PMI only pricing in Table V.

The 2005 and 2013 mispricing estimates are reported in Table XII. A few observations: The riskiest loans (FICO below 640 and LTV above 95) are substantially underpriced in both years, because of the availability of government insurance. By contrast, premiums on loans to borrowers with FICO scores of 640 and above were not on average under- or overpriced. The lack of differentiation by FICO scores over this range in 2005 meant that insurance on the higher FICO (680 and above) loans tended to be overpriced, while insurance on loans with lower (640-679) FICO scores was underpriced.

We first consider the relationship between product share and mispricing by comparing the change in each between 2005 and 2013. Let p_{jt} refer to the insurance premium for product j at date t, in percentage points as in Table XI, and p^*_{jt} the "correct" premiums as in Table IX. Thus our measure of mispricing of product j at date t would be $p_{jt} - p^*_{jt}$. A positive number means that insurance for the product was overpriced. As formalized below, we look at changes (specifically, the differences between the values in 2013 and 2005) because the levels presumably depend on things like demographics and the distribution of income. These likely do not change dramatically over eight years, and since they are not of direct interest can be "differenced out." The relationship need not be linear in these differences. It turns out that a log-linear specification is a much better fit. Because some of the product shares in 2013 are virtually zero, we use the change in share relative to the midpoint, $\Delta\hat{s}_j \equiv \Delta s_{jt}/\bar{s_j}$, where s_{jt} denotes share of product j at date t relative to the total dollar value of insured mortgages in the sample, and $\bar{s_j}$ is the average of s_{j2005} and s_{j2013}. Note that by this definition $\Delta\hat{s}_j$ has a range of $[-2, 2]$.

Figure XII plots, product by product, the relative change in product share against the change in mispricing. Adverse selection, in the form of a negative relationship between market shares and mispricing, is evident in "prime" and "near-prime" products—those with FICO scores of at

Table XI
Mortgage Insurance Rates in 2005 and 2013 (PMI/FHA Composite)

		minimum FICO Scores									
	max LTV	760	740	720	700	680	660	640	620	600	575
2005	85	0.32	0.32	0.32	0.32	0.33	0.33	0.33	0.42	0.56	0.77
	90	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.69	0.92	1.13
	95	0.79	0.79	0.79	0.79	0.79	0.79	0.79	1.01	1.18	1.14
	97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	1.15	1.12	1.10
	100	1.07	1.07	1.07	1.07	1.07	1.07	1.12	1.12	1.09	1.07
2013	85	0.29	0.32	0.33	0.40	0.41	0.51	0.57	0.84	1.56	1.56
	90	0.45	0.49	0.49	0.63	0.64	0.82	0.92	1.20	1.53	1.57
	95	0.62	0.68	0.68	0.97	0.99	1.32	1.44	1.57	1.57	1.57
	97	1.11	1.17	1.18	1.44	1.49	1.61	1.61	1.61	1.61	1.62
	100	1.60	1.61	1.61	1.61	1.61	1.61	1.61	1.61	1.61	1.61

*Units are in percentage points per year paid monthly.

Source: WI and NC mortgage insurer regulatory filings, authors' calculations

Table XII
Mortgage Insurance Mispricing in 2005 and 2013

		minimum FICO Scores									
	Max LTV	760	740	720	700	680	660	640	620	600	575
2005	85	0.05	0.05	0.05	0.03	0.03	0.01	0.01	0.04	0.11	0.17
	90	0.16	0.15	0.15	0.09	0.09	0.02	0.02	0.01	0.07	-0.14
	95	0.26	0.22	0.22	0.09	0.09	-0.06	-0.05	-0.21	-0.39	-1.25
	97	0.35	0.30	0.30	0.13	0.13	-0.06	-0.06	-0.38	-0.85	-2.11
	100	0.24	0.18	0.18	-0.07	-0.07	-0.36	-0.31	-1.00	-1.66	-3.11
2013	85	0.01	0.01	0.01	0.02	0.03	0.07	0.14	0.19	0.72	0.28
	90	0.00	0.00	0.01	0.01	0.03	0.07	0.16	0.14	0.18	-0.45
	95	0.01	0.01	0.01	0.03	0.05	0.13	0.24	-0.11	-0.60	-1.69
	97	0.12	0.13	0.14	0.19	0.24	0.08	0.08	-0.71	-1.40	-2.97
	100	0.54	0.49	0.43	0.27	0.14	-0.11	-0.40	-0.87	-1.62	-3.33

Source: Authors' estimates

least 640, as indicated by darker points in the figure.[38] Products in the northwest quadrant are primarily those that were overpriced in 2005, and priced more fairly in 2013. They experienced increases in market shares. Those in the southeast quadrant tended to be underpriced in 2005 relative to 2013, so their product shares tended to decline.

For example, consider the 95% LTV, FICO 760 loan. This product was overpriced by 26 basis points in 2005, and by just one basis point in 2013. Its share of the insured market went from 3.72 percent in 2005 to 13.33 percent in 2013. This product would be indicated by the point (-0.25, 1.13) in Figure XII.

Table X shows, for example, that the 760 and higher FICO group went from 13.2 percent of the market in 2005 (when its premiums were overpriced) to 31.4 percent in 2013. Every group with FICO$\geq$ 700 increased its share while every group with FICO$<$ 700 decreased its share. For those products with FICO scores 640 and higher, our findings suggest this shift to safer products is consistent with a response to the change in the structure of insurance premiums, from one in which FICO scores were ignored to one in which they were priced analogously to LTV. This pooling before 2008 meant that insurance on mortgages with high FICO scores tended to be overpriced, in effect cross-subsidizing the underpriced insurance on mortgages with lower FICO scores.

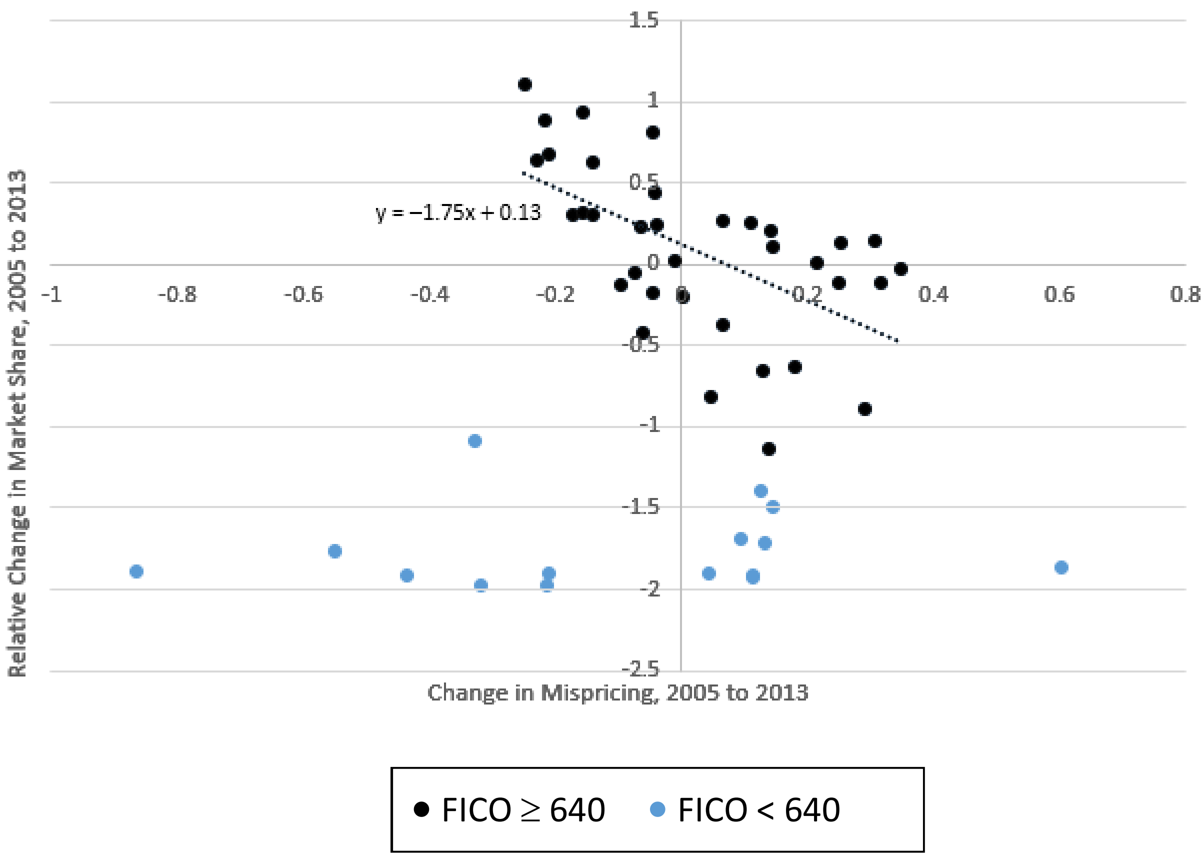
3.3.2 Estimating the Relative Contributions of Fundamental Prices and Mispricing

Figure XII shows products with FICO scores below 640 with lighter dots, mostly hovering near -2 on the vertical axis, meaning that they had largely disappeared from the market by 2013. These low FICO score products were more expensive in 2013 than in 2005, but were, by our estimates, as underpriced in 2013 as in 2005, if not more so for higher LTV products (see Table XII).

Thus while pricing and adverse selection played a significant role for the FICO $\geq$ 640 products, a different mechanism was in play for the lowest FICO score products. It could have been rationing, i.e. insurers (or lenders) simply rejecting borrowers with such low FICO scores. The "rationing" and price mechanism stories are not mutually exclusive, however. Moreover, standard models of

[38] Mortgages with FICO scores 640 and higher made up more than 80 percent of our sample of insured loans in 2005, and most were privately insured.

Figure XII
Relative Quantity Impact of Mispricing, 2005 vs 2013



Source: Authors' estimates

credit markets with asymmetric information and default costs (e.g. Bernanke and Gertler (1989)) imply that market shares also depend on the fundamental price. Under actuarially fair pricing, the presence of default costs implies that higher relative premiums (i.e. higher default risk) on a particular FICO-LTV combination will cause the relative quantity of borrowing to decline or even fall to zero. This motivates modeling s_{jt} as a function of the fundamental price in addition to mispricing.

Our empirical framework is therefore as follows: We assume that product share s_{jt} depends on p^*_{jt}, $p_{jt} - p^*_{jt}$, a product fixed effect d_j, and an error term ϵ_{jt}. The fixed effect represents the normal share of product j in the absence of unusual market conditions. The error term represents random changes in the cross-sectional demand for products. Our identification assumption is that the product supply curves are horizontal, i.e. premiums are not affected by variations in the cross-sectional demand for mortgage products. These markets are highly competitive, and each product is small relative to the mortgage market, and even smaller relative to financial markets as a whole.

This specification is in the spirit of the Dixit-Stiglitz model of a product demand system. In that framework, product shares depend on the product's own price relative to an aggregate price index. The symmetry of that model (equal cross-elasticities between all goods) makes it not precisely analogous, but it remains a useful approximation for the purposes of this exercise. We focus on shares rather than total quantities in order to emphasize the cross-section effects and abstract from aggregate factors that would affect all products in the same direction.

Our market share specification is therefore

$$\ln(s_{jt}) = \alpha + d_j + \gamma(p_{jt} - p^*_{jt}) + \theta p^*_{jt} + \epsilon_{jt}.$$

Differencing (and approximating $\Delta \ln(s_{jt})$ by the percentage change $\Delta \hat{s}_j$) yields[39]

$$\Delta \hat{s}_j = \gamma \Delta(p_j - p_j^*) + \theta \Delta p_j^* + \Delta \epsilon_j. \tag{5}$$

We would expect both coefficients to be negative, but not necessarily the same.[40] If they are the same, then Δp_j suffices to account for the price impact; whether it is due to mispricing or fundamentals makes no difference.

Because we have a limited dependent variable, we estimate this relationship with a logit specification (after transforming $\Delta \hat{s}_j$ to range between zero and one). The result of this regression on the full sample of 50 products is shown in Table XIII under Model I: The coefficient on $\Delta(p_j - p_j^*)$ is nearly identical to the slope for the FICO $\geq$ 640 mortgages in Figure XII, but this regression includes all 50 products and controls for the change in the fundamental price.

The Model I results suggest that the mortgages with FICO scores below 640 were priced attractively in 2005, and that their near disappearance by 2013 is at least in part, and perhaps primarily, attributable to the increase in their mortgage insurance premiums. As Table XI shows, the premiums were anywhere from 40 to 100 basis points higher in 2013 than in 2005. This first regression suggests that premium increases of those magnitudes could explain the near disappearance of these products.

A natural way to test the rationing hypothesis would be to include non-price criteria in the regression. To pursue this we extend the above specification a dummy variable for FICO $<$ 640, since these products' shares fell to near zero. The idea is that even after controlling for price-related factors, an economically and statistically significant coefficient on this non-price criterion would imply a role for rationing.

The result is shown in Table XIII under Model II. There are several things to note. The coefficient on the FICO $<$ 640 dummy is large and strongly significant. It suggests that the impact of having a FICO score below 640 is similar to that of least a 200 basis point increase in

[39]We suppress the t subscript since there are only two time periods.

[40]This can be seen from the fact that in the limiting case of no default costs and risk-neutrality, a fundamentals-driven change in the premium has no impact, whereas mispricing distorts decision-making.

the insurance premium. The R^2 for the regression increases from 0.557 to 0.837. Clearly something other than just premium increases contributed to the near disappearance of these products.

Finally, note that the coefficients on the two price variables, while still negative and significantly different from zero, are no longer significantly different from each other. This suggests a simpler specification: If we constrain the coefficients on these two variables to be the same, the price variable is simply Δp_j, the change in the premium between 2005 and 2013. The result for this specification is shown as Model III. It fits about as well as Model II, with an R^2 of 0.852 rather than 0.857 in the unconstrained version, and the standard F test fails to reject this specification.

Having the variables enter the equation as the sum has a helpful, simplifying interpretation. While our analysis has focused on the effects of mispricing, in fact it is the variation in premiums that matters, whether from mispricing or changes in fundamentals. If the premium on a product increased by a given number of basis points between 2005 and 2013, the effect on market share is the same whether the increase was the result of correction of mispricing, or a change in fundamentals. This does not mean that mispricing was unimportant: It was a major reason behind many of the premium changes, particularly for privately insured mortgages. Mortgages with FICO scores 640 and higher were mispriced in 2005 as a result of being charged a common premium, and the changes in premiums by 2013 reflect a correction of that mispricing.

Given that the dependent variable is transformed, the quantitative implications of Model III are not immediately apparent. The units of the explanatory variables in the regressions are changes in percentages, but the dependent variable is defined as a logistic transformation of percentage change in a market share. To explain the results quantitatively, Table XIV provides the implied 2013 market shares, given a 2005 market share of 10 percent and the indicated changes in insurance premiums, for products with FICO scores above and below 640.[41]

To understand these results, note that between 2005 and 2013 premiums increased by an average (unweighted) of 34 basis points. Product shares, of course, always sum to 100 percent. For the higher FICO score products, consequently, a 20 basis point increase in the insurance

[41] The implied 2013 shares are always proportional to the 2005 share. So if a product's 2005 share were 5 percent rather than 10 percent, its implied 2013 share would be half that shown in the table.

Table XIII
Product Share Regression Results[†]

variable	Model I	Model II	Model III
constant	0.591^{**}	0.354^{**}	0.347^{**}
	(0.163)	(0.105)	(0.093)
$\Delta(p_j - p_j^*)$	-1.404^{**}	-1.007^{**}	—
	(0.428)	(0.271)	
Δp_j^*	-3.017^{**}	-1.057^{**}	—
	(0.385)	(0.330)	
Δp_j	—	—	-1.023^{*}
			(0.249)
FICO < 640	—	-1.532^{**}	-1.548^{**}
		(0.177)	(0.143)
R^2	0.568	0.835	0.835

[†]The dependent variable is $\ln(y/(1-y))$, where $y \in [0,1] \equiv 1 + \Delta\hat{s_j}/2$ is the transformed change in product share. Standard errors are in parentheses.

Source: CoreLogic, WI and NC mortgage insurer regulatory filings, and authors' calculations.

**Significant with p-value < 0.01.

Table XIV
Implied 2013 Product Shares versus 10 Percent Share in 2005

Δp	FICO ≥ 640	FICO < 640
1.00	4.66	0.24
0.50	8.45	0.43
0.20	12.08	0.61
-0.20	19.43	0.98

Note: All values are in percent units.

Source: Authors' calculations.

premium nonetheless results in a bit more than a 20 percent increase in product share, i.e. from 10 percent to 12.08 percent. This is both because 20 basis points is below the average increase in premiums, and the low FICO score products were both priced and rationed nearly out of existence.

The 50 and 100 basis point increases are representative of what happened with the lowest FICO score products. The table shows that increases of those magnitudes have a substantial impact on product shares, but rationing would typically have a larger impact. For a 100 basis point increase, the price effect brings the market share down to below 5 percent, and rationing further lowers it to 0.2 percent. But a 50 basis point increase only lowers the share to 8.45 percent, while rationing brings it down to below 0.5 percent. So for these low FICO score products, premium increases contributed substantially to their near disappearance, but in most cases rationing was responsible for more than half their decline in market shares.

To gauge the impact of the three effects in the regression (mispricing, beliefs, and 'rationing'), we focus on the share of higher-risk insured mortgages, which we take to be those in the "high" and "very high" categories from Section 1. The dollar-weighted share of these mortgages was 81 percent in 2005, compared with 58 percent in 2013. To get some idea of the quantitative contribution of each of the effects to that change, we alternately zero out the variables in the regressions. Because the regression is non-linear, the effects interact, so there is a range of answers depending on ordering.

In addition, we interpret the "FICO$<$640" or "rationing" effect as an implementation of some combination of correcting mispricing and changes in beliefs, not an independent effect. If insurers realized that insurance on low FICO mortgages had been underpriced in 2005 even given the optimistic beliefs, and that given 2013 beliefs, with correct pricing, the market was not viable, then the rationing may be in part or entirely a response to mispricing. On the other hand, given that by 2013 the sub-640 FICO score mortgages are almost entirely FHA insured, the objective may have been to continue to subsidize this group of borrowers to the same extent as in 2005, but the 2013 beliefs made even that product non-viable. In that case the rationing effect would mainly reflect the changes in beliefs. There is no clear way to distinguish these effects in the data.

Consequently the quantitative impact on risky lending that we attribute to rationing is really some unknowable combination of the two.

With that in mind, we find that 2005 "mispricing" contributed in the vicinity of 2 to 5 percentage points (if the FICO<640 effect were purely optimism) to as much as 15 percentage points (if the FICO<640 effect were purely correcting mispricing) of the 24 percentage point difference in the share of high-risk mortgages in 2005 versus 2013.[42] The remainder would be mainly be attributable to changes in beliefs, along with a small residual unexplained component. See Appendix C for details.

4 Conclusions

Our examination of mortgage credit risk pricing yields two findings: The boom in high-risk mortgages was a response to both the mispricing of risk and to optimistic beliefs about house prices. As emphasized by Adelino et al. (2016), optimism would lead to increased borrowing across the board. The mispricing largely stemmed from pooling across widely disparate credit risks (as measured by FICO scores), and resulted in adverse selection within the pool. In addition, the riskiest mortgages were attractive because of the underpricing of risk by government insurers. The government continued to underprice risk after 2008. While less optimistic assumptions about housing markets resulted in higher premiums, prices were still too low. These premium increases reduced borrowing by borrowers with FICO scores below 640. However, the collapse in market share of these products appears primarily due to rationing rather than as a response to price changes.

These findings are thus consistent with Kaplan et al. (2017), who argue for the primary importance of common changes in beliefs on the part of both borrowers and lenders (as also argued by Adelino et al. (2016)), while at the same time acknowledging the presence of credit supply

[42] As mentioned above, in the pre-2008 period many borrowers were able to obtain piggyback loans as an alternative to PMI. The second lien loan, like PMI, prices much of the credit risk of the loan. However, for some high FICO borrowers, for whom PMI was likely overpriced, piggyback loans were less expensive. This suggests that 2005 loans to high FICO borrowers may not have been as overpriced as our estimates from PMI rates suggest. Less mispricing of the high FICO loans implies a steeper relationship between mispricing and market share in our analysis. Therefore, our estimates of the impact of mispricing in the misallocation of credit may be conservative.

shifts (as in Mian and Sufi (2009) and others). The supply shifts are evident in the mispricing of default risk by private mortgage insurers — specifically the pooling of observably different risks, with resulting adverse selection. In comparing the insurance premiums for these products in 2005 versus 2013, we see that they were not systematically higher in 2013, just differentiated by FICO score, as the errors implied by pooling were eliminated. For the highest-risk products, primarily insured by the government, insurance was underpriced both before and after 2008. But the insurance premiums themselves were substantially higher by 2013, consistent with more pessimistic beliefs about housing markets.

Thus, we can attribute the overall lower mortgage insurance premiums during the boom to optimism, which is not a failure to mitigate risk *ex ante*, only with hindsight. By contrast, the cross-sectional mispricing of those premiums is a failure to mitigate risk *ex ante*: It implies a prominent role for supply shifts, i.e. the underpricing of high-risk mortgages and overpricing of low-risk mortgages that was knowable at the time. Indeed and perhaps most important, we provide evidence that this mispricing affected choices meaningfully and exacerbated risk taking rather than mitigated it. Borrowers responded to this pattern of risk pricing, with increased borrowing in the underpriced higher-risk categories.

A deeper question raised by this analysis is, "What is the ultimate purpose of mortgage insurance?" As most mortgage credit risk is held by large institutions, notably the GSEs (e.g. Fannie Mae), idiosyncratic risk is largely diversified away and therefore should be of minor consequence. Presumably, the function of mortgage insurance is to shift some aggregate risk from the GSEs to mortgage insurers who were relatively better capitalized. The GSEs were highly levered: For example, Acharya et al. (2011) report that they had only a 0.45 percent capital requirement against mortgages they guaranteed. Even a modest decline in national house prices could threaten their solvency in the absence of mortgage insurance. Shifting of some of this aggregate risk to better capitalized (with a 4.00 percent capital requirement) mortgage insurers reduces the likelihood of GSE insolvency, increases the value of their guarantee to mortgage originators, and reduces the risk to the taxpayers providing an implicit guarantee to the GSEs. It also could reduce overall risk to the extent that the mortgage insurers are better able to evaluate and screen credit risks than are the holders of the mortgages.

Our results are complementary to other work on mortgage finance during the housing boom. That work primarily focuses on quantities (Foote et al. (2016), Mian and Sufi (2009), and Ambrose and Diop (2014)). One reason for this has been the difficulty in distinguishing the pricing of default risk from other components of mortgage interest rates. Another reason was a prevailing assumption that credit "rationing" or extensive margin decisions about who can get mortgages, and what types of mortgages to offer, were the primary allocation mechanisms. Our work demonstrates that the price mechanism is also operative in these markets. Moreover, the price data are informative about the underlying forces that gave rise to the subprime mortgage boom and bust.

A Imputing Missing PMI Rates

The regression estimation output appears in Table A.I.

Table A.I
PMI Imputation Model

Dependent Variable	PMI Rate (in percent)			
Periods included	208 (monthly, January 1999-April 2016)			
Cross-sections included*	139			
Total panel (non-missing) observations	16,767			
Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
$Constant$	3.203E-02	4.533E-03	7.065	0.000
$MinFICO$	1.061E-03	2.380E-04	4.465	0.000
$MaxLTV$	-8.754E-02	1.842E-03	-47.532	0.000
$MinFICO^2$	-4.770E-06	3.090E-06	-1.543	0.123
$MaxLTV^2$	1.130E-02	1.760E-04	64.112	0.000
$MaxLTV \times MinFICO$	-3.950E-05	1.950E-06	-20.191	0.000
$MinFICO^3$	-8.800E-08	1.050E-08	-8.409	0.000
$MaxLTV^3$	-2.370E-04	4.210E-06	-56.297	0.000
$LOWDOC$	3.509E-01	7.960E-03	44.077	0.000
$Y_{2007} \times MinFICO$	7.470E-05	2.410E-04	0.311	0.756
$Y_{2007} \times MaxLTV$	1.995E-03	1.952E-03	1.022	0.307
$Y_{2008} \times MinFICO$	3.760E-04	2.410E-04	1.557	0.120
$Y_{2008} \times MaxLTV$	2.315E-03	1.962E-03	1.180	0.238
$Y_{2009} \times MinFICO$	-8.660E-06	2.470E-04	-0.035	0.972
$Y_{2009} \times MaxLTV$	3.554E-03	1.973E-03	1.801	0.072
$Y_{2010} \times MinFICO$	-1.382E-03	2.550E-04	-5.423	0.000
$Y_{2010} \times MaxLTV$	6.591E-03	1.979E-03	3.331	0.001
$Y_{2011} \times MinFICO$	-1.910E-03	2.560E-04	-7.471	0.000
$Y_{2011} \times MaxLTV$	5.968E-03	1.984E-03	3.008	0.003
$Y_{2012} \times MinFICO$	-2.814E-03	2.570E-04	-10.933	0.000
$Y_{2012} \times MaxLTV$	-2.610E-04	2.001E-03	-0.131	0.896
$Y_{2013} \times MinFICO$	-2.866E-03	2.570E-04	-11.132	0.000
$Y_{2013} \times MaxLTV$	3.873E-03	2.001E-03	1.936	0.053
$Y_{2014} \times MinFICO$	-3.356E-03	2.570E-04	-13.036	0.000
$Y_{2014} \times MaxLTV$	1.146E-02	2.001E-03	5.730	0.000
$Y_{2015} \times MinFICO$	-3.356E-03	2.570E-04	-13.036	0.000
$Y_{2015} \times MaxLTV$	1.146E-02	2.001E-03	5.730	0.000
$Y_{2016} \times MinFICO$	-4.134E-03	2.990E-04	-13.829	0.000
$Y_{2016} \times MaxLTV$	1.219E-02	2.109E-03	5.781	0.000
$LOWDOC \times MinFICO$	1.760E-04	2.340E-04	0.753	0.451
$LOWDOC \times MaxLTV$	1.583E-02	1.750E-03	9.047	0.000
$LOWDOC \times MinFICO \times (YEAR > 2007)$	-1.535E-03	9.470E-05	-16.213	0.000
$LOWDOC \times MaxLTV \times (YEAR > 2007)$	7.588E-03	5.740E-04	13.219	0.000
$LOWDOC \times MinFICO^2$	2.160E-05	2.700E-06	7.996	0.000
$LOWDOC \times MaxLTV^2$	-1.971E-03	1.660E-04	-11.835	0.000
$LOWDOC \times MinFICO^3$	8.980E-08	8.700E-09	10.330	0.000
$LOWDOC \times MaxLTV^3$	3.860E-05	4.040E-06	9.551	0.000
$LOWDOC \times MaxLTV \times MinFICO$	2.140E-05	3.280E-06	6.549	0.000
$LOWDOC \times MinFICO \times Y_{2007}$	4.570E-04	7.370E-05	6.191	0.000
$LOWDOC \times MaxLTV \times Y_{2007}$	4.151E-03	4.520E-04	9.176	0.000
$Y_{2002} \times MinFICO$	1.715E-03	2.380E-04	7.198	0.000
$Y_{2002} \times MaxLTV$	3.814E-03	1.946E-03	1.960	0.050
$Y_{2003} \times MinFICO$	6.270E-04	2.380E-04	2.633	0.009
$Y_{2003} \times MaxLTV$	2.003E-03	1.946E-03	1.029	0.303
$Y_{2004} \times MinFICO$	6.230E-04	2.380E-04	2.614	0.009
$Y_{2004} \times MaxLTV$	2.244E-03	1.946E-03	1.153	0.249
$Y_{2005} \times MinFICO$	4.070E-04	2.380E-04	1.708	0.088
$Y_{2005} \times MaxLTV$	3.195E-03	1.945E-03	1.643	0.100
$Y_{2006} \times MinFICO$	2.590E-04	2.380E-04	1.086	0.278
$Y_{2006} \times MaxLTV$	3.530E-03	1.946E-03	1.814	0.070
$(YEAR > 2001) \times MinFICO^2$	4.090E-06	3.200E-06	1.278	0.201
$(YEAR > 2001) \times MinFICO^3$	-8.120E-08	1.100E-08	-7.394	0.000
$(YEAR > 2001) \times MaxLTV^2$	-7.150E-04	1.880E-04	-3.810	0.000
$(YEAR > 2001) \times MaxLTV^3$	2.210E-05	4.500E-06	4.915	0.000
R^2	0.967	Mean dependent var		0.924
$\overline{R}^2$	0.967	S.D. dependent var		0.613
S.E. of regression	0.112	F-statistic		9159.157
Sum squared resid	209.622	Log likelihood		12944.00

*Many cross-sections have missing observations due to product disappearance

$MinFICO$ = lower end of FICO range minus 760 (ranges from -185 to 0)

$MaxLTV$ = upper end of LTV range minus 65 (ranges from $+5$ to $+38$)

Source: WI and NC mortgage insurer regulatory filings and authors' calculations.

B PMI Imputation Sensitivity Analysis

Section 1.1 describes how we use imputation to provide a price for products that are no longer observable the market. These imputations are used in Sections 2.1 and 3.2 to price only these unobservable products. The results in the paper use the imputation specification from Appendix A. This section considers the robustness of the results in Section 3.2 to alternative imputation models.

We designed the model in Appendix A to capture the observed relationship between time and product level explanatory variables and observed PMI rates. Ultimately, it reflects our judgement about what features should be in such a model and was selected from amongst similar models. Out of a concern that our use of discretion is driving our results, we use cross-validated LASSO (Tibshirani (1996)) regression as an automated model selection technique free from our judgement. We choose a penalty parameter λ to pick a LASSO-optimal model with the lowest possible RMSE across cross-validated samples conditional on selecting a predetermined number of right-hand side variables. More informally, we are interested in finding the "best" model with exactly n explanatory variables, where the best models have minimal RMSE across sub-samples and the explanatory variables are all important.

We consider two models, Low Complexity (17 variables) and Medium Complexity (35 variables) which are both simpler than the model in the paper (54 variables). It is possible to create more complex models: the model with λ_{min}, which minimizes the RMSE across the sub-samples, has 186 variables but only improves on our base specification by a RMSE of 1.5 basis points. We also consider a Simple (8 variables) model, which we create to capture the coarse features of PMI pricing. These models appear is Table B.II.

Table B.III shows our results for the difference-in-difference product share regressions for each of these alternative imputation models. For all imputation specifications and product share models the results have coefficients of similar sign and magnitude. For our preferred specification, that includes the FICO$<$ 640 dummy variable (model II), the coefficients differ less than one standard deviation. Our conclusions are thus not sensitive to the imputation.

Table B.II
Comparison of PMI Imputation Models

Model Name	Variables	BP of RMSE	Model Form
Model in Paper	54	11.18	$XLPMI = FICO + LTV + FICO^2 + LTV^2 + LTV : FICO + FICO^3 + LTV^3 + Doc_{Low} + Year_{2007} : FICO + Year_{2007} : LTV + Year_{2008} : FICO + Year_{2008} : LTV + Year_{2009} : FICO + Year_{2009} : LTV + Year_{2010} : FICO + Year_{2010} : LTV + Year_{2011} : FICO + Year_{2011} : LTV + Year_{2012} : FICO + Year_{2012} : LTV + Year_{2013} : FICO + Year_{2013} : LTV + Year_{2014} : FICO + Year_{2014} : LTV + Year_{2015} : FICO + Year_{2015} : LTV + Year_{2016} : FICO + Year_{2016} : LTV + Doc_{Low} : FICO + Doc_{Low} : LTV + Doc_{Low} : FICO : Year_{>2007} + Doc_{Low} : LTV : Year_{>2007} + Doc_{Low} : FICO^2 + Doc_{Low} : LTV^2 + Doc_{Low} : FICO^3 + Doc_{Low} : LTV^3 + Doc_{Low} : LTV : FICO + Doc_{Low} : FICO : Year_{2007} + Doc_{Low} : LTV : Year_{2007} + Year_{2002} : FICO + Year_{2002} : LTV + Year_{2003} : FICO + Year_{2003} : LTV + Year_{2004} : FICO + Year_{2004} : LTV + Year_{2005} : FICO + Year_{2005} : LTV + Year_{2006} : FICO + Year_{2006} : LTV + Year_{>2001} : FICO^2 + Year_{>2001} : FICO^3 + Year_{>2001} : LTV^2 + Year_{>2001} : LTV^3$
Lasso Cross-validated Simple	8	27.54	$XLPMI = FICO + LTV + Doc_{Low} + LTV : FICO + FICO : Doc_{Low} + Year_{>2001} + Year_{>2007} + Year_{>2010}$
Lasso Cross-validated Low	17	18.34	$XLPMI = FICO + Doc_{Low} + LTV + Year_{>2001} + Year_{>2007} + Year_{>2010} + FICO^2 + LTV^2 + FICO : Doc_{Low} + LTV : FICO + LTV : Doc_{Low} + FICO : Year_{>2007} + Doc_{Low} : Year_{>2007} + LTV : Year_{>2007} + FICO : Doc_{Low} : LTV + FICO : Doc_{Low} : Year_{>2007}$
Lasso Cross-validated Medium	35	16.32	$XLPMI = FICO + Doc_{Low} + LTV + Year_{>2001} + Year_{>2007} + FICO^2 + LTV^2 + Doc_{Low} : LTV + FICO : Year_{>2001} + Doc_{Low} : Year_{>2001} + FICO : Year_{>2007} + FICO : Year_{>2010} + LTV : Year_{>2010} + LTV : FICO^2 + Year_{>2001} : FICO^2 + Year_{>2007} : FICO^2 + Year_{>2010} : FICO^2 + FICO : LTV^2 + Doc_{Low} : LTV^2 + Year_{>2007} : LTV^2 + Year_{>2010} : LTV^2 + FICO : Doc_{Low} : LTV + FICO : Doc_{Low} : Year_{>2001} + FICO : LTV : Year_{>2001} + FICO : Doc_{Low} : Year_{>2007} + Doc_{Low} : LTV : FICO^2 + Doc_{Low} : Year_{>2007} : FICO^2 + LTV : Year_{>2007} : FICO^2 + LTV : Year_{>2010} : FICO^2 + FICO : Doc_{Low} : LTV^2 + Doc_{Low} : Year_{>2001} : LTV^2 + FICO : Year_{>2007} : LTV^2 + FICO : Year_{>2010} : LTV^2 + Doc_{Low} : LTV : Year_{>2007} : FICO^2$

Table B.III
Alternative Product Share Regression Results†

variable	constant	$\Delta(p_j - p_j^*)$	Δp_j^*	Δp_j	FICO < 640	R^2
			Baseline Imputation			
Model I	0.909**	−1.874*	−5.149**	—-	—-	0.566
	(0.286)	(0.771)	(0.678)	—		
Model II	0.509**	−1.026*	−1.641**	—-	−2.759**	0.837
	(0.183)	(0.463)	(0.580)		(0.315)	
Model III	0.426*	—-	—-	−1.189*	−2.983**	0.832
	(0.167)			(0.445)	(0.256)	
			Medium Imputation			
Model I	0.602	−2.935**	−4.011**	—-	—-	0.392
	(0.331)	(0.868)	(0.741)	—		
Model II	0.431*	−1.173*	−1.201*	—-	−2.976**	0.832
	(0.176)	(0.488)	(0.469)		(0.271)	
			Simple Imputation			
Model I	0.192	−4.159**	−2.017*	—-	—-	0.432
	(0.319)	(0.720)	(0.976)	—		
Model II	0.352*	−1.406**	−0.728	—-	−2.866**	0.840
	(0.176)	(0.463)	(0.538)		(0.265)	

†The dependent variable is $\ln(y/(1-y))$, where $y \in [0,1] \equiv 1 + \Delta\hat{s_j}/2$ is the transformed change in product share. Standard errors are in parentheses.

Source: CoreLogic, WI and NC mortgage insurer regulatory filings, and authors' calculations.

*Significant with p-value < 0.05.

**Significant with p-value < 0.01.

Note: Model III is invariant to the imputation. F-tests fail to reject the restriction on Model II implied by Model III for all three imputations.

C Estimating the Marginal Effects on Risky Mortgage Share

To estimate the contributions of changes in optimism and mispricing to the 2005 boom, we use model II from Table XIII. We assume the regression $\beta's$ are distributed multivariate normally with $\mu's$ and Σ from our regression estimates. We preform 10,000 simulations, which allows us to account for the joint uncertainty in the parameters of interest. This model decomposes the change in market shares between 2005 and 2013 into three parts, the effect from mispricing, the effect from optimism, and the effect from FICO scores < 640, which could be caused by either optimism or mispricing. Due to this indeterminacy, we estimate the effect both ways, considering FICO scores < 640 as part of mispricing (and not a part of optimism) and FICO scores < 640 as part of optimism (and not part of mispricing). Because this is a non-linear (logistic) model, the marginal effects depend on the value of the other variables. To handle this issue, we consider all possible combinations of Mispricing $\in \{\text{Actual}, 0\}$, Optimism $\in \{\text{Actual}, 0\}$, and <640 $\in \{\text{On}, \text{Off}\}$. We index these combinations by i, and call the resulting data x_i

This gives $\ln(y(\hat{x_i})/(1 - y(\hat{x_i}))$. Market share in 2005 is then (2013 market share) $\cdot\, e^{\hat{y}(x_i)}$. The share of risky mortgages (risky $\widehat{\text{market}}$ $\text{share}_t(x_i)$) is then the total share of all mortgages we classify as "High" or "Very High" risk mortgages (in Section 1) normalized by the total market share. Therefore, the predicted marginal are the changes in risky market share_t (risky $\widehat{\text{market}}$ $\text{share}_t(x_i)$ - risky $\widehat{\text{market}}$ $\text{share}_t(x_j)$) as we bring variables in and out of the predictions. Figure C.I shows the results. For comparison, in 2005, risky mortgages were 81 percent of the total insured market. In 2013, risky mortgages were 58 percent of the insured market for a decline of 24 percentage points.

References

Acikgoz, O. and Kahn, J. (2016). A Quantitative Model of "Too Big to Fail," House Prices, and the Financial Crisis. Technical report, Yeshiva University.

Adelino, M., Schoar, A., and Severino, F. (2016). Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class. *The Review of Financial Studies*, 29(7):1635–1670.

Ambrose, B. W. and Diop, M. (2014). Spillover effects of subprime mortgage originations: The effects of single-family mortgage credit expansion on the multifamily rental market. *Journal of Urban Economics*, 81:114–135.

Aragon, D., Caplin, A., Chopra, S., Leahy, J. V., LeCun, Y., Scoffier, M., and Tracy, J. (2010). Reassessing fha risk. Technical report, National Bureau of Economic Research.

Aron, J. and Muellbauer, J. (2016). Modelling and forecasting mortgage delinquency and foreclosure in the uk. *Journal of Urban Economics*, 94:32–53.

Ashcraft, A. B., Goldsmith-Pinkham, P., and Vickery, J. I. (2010). Mbs ratings and the mortgage credit boom. Technical report, Federal Reserve Bank of New York.

Ashcraft, A. B., Schuermann, T., et al. (2008). Understanding the securitization of subprime mortgage credit. *Foundations and Trends® in Finance*, 2(3):191–309.

Avery, R. B., Brevoort, K. P., and Canner, G. B. (2007). The 2006 hmda data. *Fed. Res. Bull. A73*, 93.

Bernanke, B. and Gertler, M. (1989). Agency costs, net worth, and business fluctuations. *The American Economic Review*, pages 14–31.

Bhutta, N. and Keys, B. J. (2017). Eyes wide shut? Mortgage insurance during the housing boom. Technical report, National Bureau of Economic Research.

Bhutta, N., Popper, J., and Ringo, D. (2015). The 2014 home mortgage disclosure act data. *Fed. Res. Bull. 4*, 101.

Brueckner, J. K., Calem, P. S., and Nakamura, L. I. (2012). Subprime mortgages and the housing bubble. *Journal of Urban Economics*, 71(2):230–243.

Calem, P. S. and Wachter, S. M. (1999). Community reinvestment and credit risk: Evidence from an affordable-home-loan program. *Real Estate Economics*, 27(1):105–134.

Case, K. E., Shiller, R. J., and Thompson, A. (2012). What have they been thinking? Home buyer behavior in hot and cold markets. Technical report, National Bureau of Economic Research.

Chen, H., Michaux, M., and Roussanov, N. (2013). Houses as atms? mortgage refinancing and macroeconomic uncertainty. Technical report, National Bureau of Economic Research.

Chirico, C. and Mehlman, S. (2013). Fha's single-family mortgage guarantee program: Budgetary cost or savings?

Cowan, A. M. and Cowan, C. D. (2004). Default correlation: An empirical investigation of a subprime lender. *Journal of Banking & Finance*, 28(4):753–771.

Cutts, A. C. and Merrill, W. (2008). Interventions in mortgage default: Policies and practices to prevent home loss and lower costs. *Borrowing to live: Consumer and mortgage credit revisited*, pages 203–254.

Deng, Y., Quigley, J. M., Van Order, R., and Mac, F. (1996). Mortgage default and low downpayment loans: the costs of public subsidy. *Regional science and urban economics*, 26(3-4):263–285.

Doherty, N. and Smetters, K. (2005). Moral hazard in reinsurance markets. *Journal of Risk and Insurance*, 72(3):375–391.

Edelberg, W. (2006). Risk-based pricing of interest rates for consumer loans. *Journal of monetary Economics*, 53(8):2283–2298.

Elmendorf, D. W. (2011). Accounting for fha's single-family mortgage insurance program on a fair-value basis.

Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D., and Hunt, R. (2010). What" triggers" mortgage default? *American Economic Review*, 100(2):490–94.

Emrath, P. (2009). How long buyers remain in their homes. *HousingEconomics.Com*.

Epperson, J. F., Kau, J. B., Keenan, D. C., and Muller, W. J. (1985). Pricing default risk in mortgages. *Real Estate Economics*, 13(3):261–272.

Federal Housing Administration, Office of the Assistant Secretary for Housing, and Federal Housing Commissioner (2008). *Federal Housing Administration (FHA) Single Family Mortgage Insurance: Implementation of Risk-Based Premiums; Notice*, volume 73. National Archives and Records Administration.

Fisher, I. (1922). *The making of index numbers: a study of their varieties, tests, and reliability*. Number 1 in 1. Houghton Mifflin.

Flavin, M. and Yamashita, T. (2002). Owner-Occupied Housing and the Composition of the Household Portfolio. *The American Economic Review*, 92(1):345–362.

Foote, C. L., Loewenstein, L., and Willen, P. S. (2016). Cross-sectional patterns of mortgage debt during the housing boom: evidence and implications. Technical report, National Bureau of Economic Research.

Goodman, L., McCargo, A., Golding, E., Parrott, J., Pardo, S., Hill-Jones, T. M., Zhu, J., Bai, B., Kaul, K., Ganesh, B., Strochak, S., and Reyes, A. (2018). Housing finance at a glance: A monthly chartbook.

Harding, J. P., Rosenthal, S. S., and Sirmans, C. F. (2007). Depreciation of housing capital, maintenance, and house price inflation: Estimates from a repeat sales model. *Journal of urban Economics*, 61(2):193–217.

Jones, M. C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, 6(1):70–81.

Justiniano, A., Primiceri, G. E., and Tambalotti, A. (2016). The mortgage rate conundrum. Technical report, Technical report.

Kaplan, G., Mitman, K., and Violante, G. L. (2017). The housing boom and bust: Model meets evidence. Technical report, National Bureau of Economic Research.

Lam, K., Dunsky, R. M., and Kelly, A. (2013). Impacts of down payment underwriting standards on loan performance-evidence from the GSEs and FHA portfolios. *FHFA Working Paper*, pages 13–3.

Ligon, J. L. and Michel, N. J. (2015). *The Federal Housing Administration: What Record of Success?* The Heritage Foundation.

MGIC (2017). Mortgage Insurance Coverage Requirements.

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

Neal, M. (2015). Lien-ing in: What is behind the continued recovery of mortgage default rates?

Park, K. A. (2016). Fha loan performance and adverse selection in mortgage insurance. *Journal of Housing Economics*, 34:82–97.

Pennington-Cross, A., Yezer, A., and Nichols, J. (2000). Credit risk and mortgage lending: Who uses subprime and why? *Research Institute for Housing America*.

Pinto, E. (2014). Periodic Table of Housing Risk Home Purchase Loans v. 2.19.14.

Rumsey, T. (2017). Fha historical mi premium chart.

Sherlund, S. M. (2008). The Past, Present, and Future of Subprime Mortgages. *Finance and Economics Discussion Series*, (2008-63).

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic theory*, 21(2):265–293.

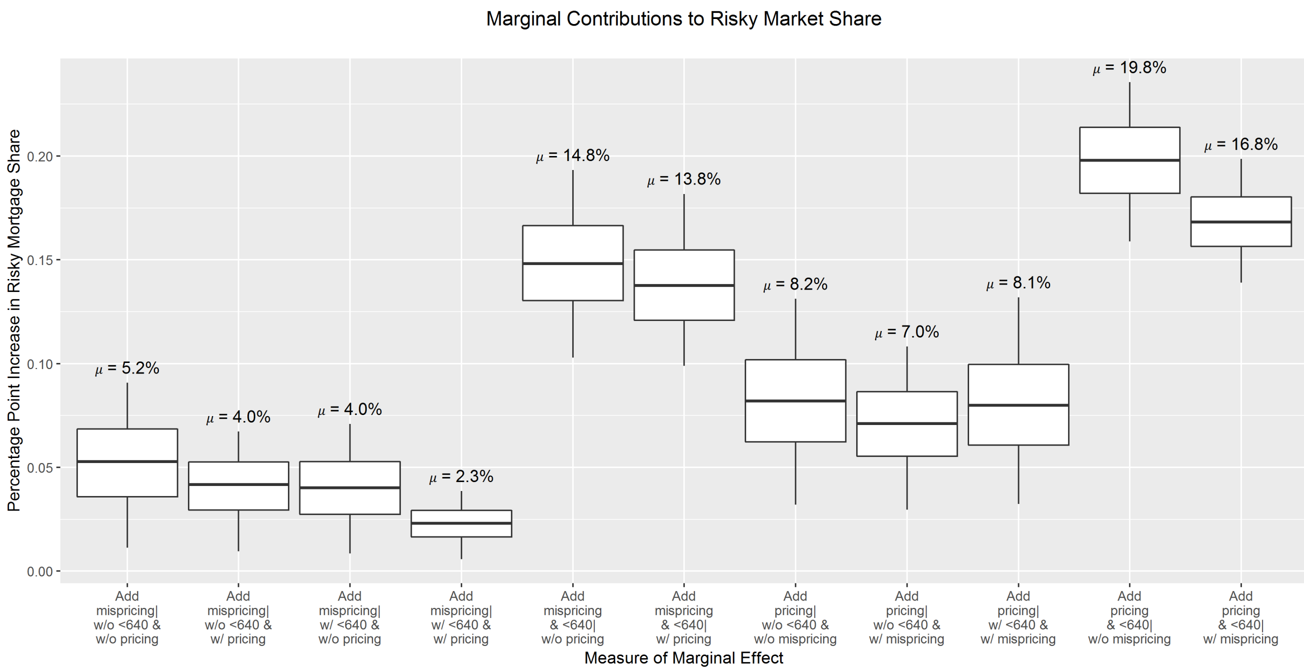
United States Financial Crisis Inquiry Commission (2011). *The financial crisis inquiry report: Final report of the national commission on the causes of the financial and economic crisis in the United States*. PublicAffairs.

Urban Institute (2017). *Mortgage Insurance Data at a Glance*. Urban Institute.

Vandell, K. D. (1995). How ruthless is mortgage default? A review and synthesis of the evidence. *Journal of Housing Research*, 6(2):245.

White, A. M. (2008). Deleveraging the American homeowner: The failure of 2008 voluntary mortgage contract modifications. *Conn. L. Rev.*, 41:1107.

Figure C.I
Simulation Estimates of Contributions to Risky Mortgage Market Share Scores



Source: Author's Estimates

Previous volumes in this series

874 August 2020	Demographic origins of the decline in labor's share	Andrew Glover and Jacob Short
873 July 2020	Effects of Fed policy rate forecasts on real yields and inflation expectations at the zero lower bound	Gabriele Galati and Richhild Moessner
872 July 2020	Effects of credit restrictions in the Netherlands and lessons for macroprudential policy	Gabriele Galati, Jan Kakes and Richhild Moessner
871 July 2020	The Matthew effect and modern finance: on the nexus between wealth inequality, financial development and financial technology	Jon Frost, Leonardo Gambacorta and Romina Gambacorta
870 July 2020	International spillovers of forward guidance shocks	Callum Jones, Mariano Kulish and Daniel M Rees
869 June 2020	How well-anchored are long-term inflation expectations	Richhild Moessner and Előd Takáts
868 June 2020	Debt De-risking	Jannic Cutura, Gianpaolo Parise and Andreas Schrimpf
867 June 2020	The effectiveness of macroprudential policies and capital controls against volatile capital inflows	Jon Frost, Hiro Ito and René van Stralen
866 May 2020	Model risk at central counterparties: Is skin-in-the-game a game changer?	Wenqian Huang and Előd Takáts
865 May 2020	The drivers of cyber risk	Iñaki Aldasoro, Leonardo Gambacorta, Paolo Giudici and Thomas Leach
864 May 2020	Global and domestic financial cycles: variations on a theme	Iñaki Aldasoro, Stefan Avdjiev, Claudio Borio and Piti Disyatat
863 May 2020	Pension contributions and tax-based incentives: evidence from the TCJA	Ahmed Ahmed and Anna Zabai
862 May 2020	On the instability of banking and other financial intermediation	Chao Gu, Cyril Monnet, Ed Nosal and Randall Wright
861 May 2020	Dealers' insurance, market structure, and liquidity	Francesca Carapella and Cyril Monnet
860 April 2020	Dollar invoicing, global value chains, and the business cycle dynamics of international trade	David Cook and Nikhil Patel

All volumes are available on our website www.bis.org.