FHFA STAFF WORKING PAPER SERIES

A Quarter Century of Mortgage Risk

Morris A. Davis
William D. Larson
Stephen D. Oliner
Benjamin R. Smith

February, 2022 (revised)
January, 2019 (original)

Working Paper 19-02

FEDERAL HOUSING FINANCE AGENCY
Division of Research & Statistics
Office of Economic Research
400 7th Street SW
Washington, DC 20219, USA

Please address correspondence to William Larson (william.larson@fhfa.gov).

A Quarter Century of Mortgage Risk

Morris A. Davis, William D. Larson, Stephen D. Oliner, Benjamin R. Smith
FHFA Staff Working Paper 19-02

February, 2022 (revised)
January, 2019 (original)

Abstract

This paper provides a comprehensive account of the evolution of default risk for newly originated home mortgages over the past quarter century. We bring together several data sources to produce this history, including loan-level data for the entire Enterprise (Fannie Mae and Freddie Mac) book. We use these data to track a large number of loan characteristics and a summary measure of risk, the stressed default rate. Among the many results in the paper, we show that mortgage risk had already risen in the 1990s, planting seeds of the financial crisis well before the actual event. Our results also cast doubt on explanations of the crisis that focus on borrowers with low credit scores. The aggregate series we present in this paper are available for download at `https:/www.fhfa.gov/papers/wp1902.aspx`.[1]

Keywords: mortgage risk · housing boom · default · foreclosure · house price · leverage

JEL Classification: E32, G21, G28, H22, R31

Morris A. Davis
Department of Finance and Economics
Rutgers Business School
Rutgers University
1 Washington Park Room 627
Newark, NJ 07102
mdavis@business.rutgers.edu

William D. Larson
Federal Housing Finance Agency
Division of Research & Statistics
Office of Economic Research
400 7th Street SW
Washington, DC 20219, USA
william.larson@fhfa.gov

Stephen D. Oliner
American Enterprise Institute for Public Policy Research
1789 Massachusetts Avenue, NW
Washington, DC 20036
stephen.oliner@aei.org

Benjamin R. Smith
The Wharton School
University of Pennsylvania
3733 Spruce Street
Philadelphia, PA 19104
benjamrs@wharton.upenn.edu

[1] These indices are works in progress and all data, tables, figures, and other results in this working paper are subject to change. Earlier versions of this paper were posted in January, March, and October 2019, and May, 2021. The January and March 2019 versions were posted under the title "Mortgage Risk Since 1990."

1. Introduction

Since the global financial crisis, there has been an outpouring of research to understand the developments in the U.S. home mortgage market that precipitated the crisis.[1] Nonetheless, as valuable as this research has been, there is still no comprehensive account of the changes in mortgage risk that produced the worst foreclosure wave since the Great Depression.

This paper is an effort to fill that gap, covering essentially the entire market for home mortgage loan originations in the United States from 1990 to 2019. We bring together several data sources to construct this comprehensive historical picture, including the full set of home mortgages guaranteed by the government sponsored enterprises Fannie Mae and Freddie Mac (the Enterprises). To our knowledge, this is the first time the entire Enterprise book has been used in publicly-available research on mortgage risk.[2] We supplement the Enterprise dataset with data covering more than 90 percent of the loans in private mortgage-backed securities and data from mortgage servicers for a large fraction of loans in the rest of the market (loans guaranteed by the Federal Housing Administration (FHA) and the Department of Veterans Affairs (VA) as well as unguaranteed loans held in the portfolio of banks and other lenders).

We track many borrower and loan characteristics that influence loan performance. These characteristics include the borrower's credit score, the debt-payment to income ratio (DTI), the combined loan-to-value ratio (CLTV) that accounts for any subordinate liens at origination, loan type (fixed or adjustable mortgage rate), loan term, loan purpose, whether the borrower's income is fully documented, and whether the mortgage has a feature that modifies the paydown of loan principal (such as a period of interest-only payments).

Given the multitude of loan characteristics, it is important to develop a summary measure of default risk. This measure should convey the risk of default in a simple, straightforward

[1] For readers seeking to explore this vast literature, the appendix table in Ferreira and Gyourko (2015) lists more than 30 papers on various aspects of the mortgage market, Foote and Willen (2018) discuss the post-crisis surge of research on the determinants of mortgage default, and Mian and Sufi (2018) cover research on the expansion of mortgage credit supply during the housing boom and its implications for house prices.

[2] The Enterprise data source is the Mortgage Loan Information System (MLIS) at the Federal Housing Finance Agency (FHFA). No personally identifiable information is contained in the tables, figures, or the database associated with this paper. Results presented pertaining to these data are aggregates and are rounded where appropriate.

manner, and to be of most use to policymakers with prudential oversight responsibilities, it should focus on default under severely stressed conditions. To this end, we calculate what we call the "stressed default rate" for all loans in the dataset based on the observed default experience of similar loans originated nationwide in 2006 and 2007, just before the financial crisis. The stressed default rate for a given loan thus represents its expected counterfactual default rate if it were hit shortly after origination with a replay of the financial crisis, including the observed national average decline in house prices and ensuing policy responses. Because the drop in house prices varied enormously across localities, the stressed default rate calculated in this way will only be valid for a national portfolio of loans. As an extension, we also calculate a stressed default measure for states and metropolitan areas. This extension embeds the estimated magnitude of a severe local house price shock in each year using the framework in Smith and Weiher (2012) and Smith et al. (2016).

The stressed default rates presented in this paper build on those currently published by the AEI Housing Center and the Urban Institute. AEI unveiled its National Mortgage Risk Index (NMRI) in 2013, while the Urban Institute introduced its Housing Credit Availability Index (HCAI) a year later, drawing in large part on the NMRI methodology. The NMRI uses highly accurate and complete data for government-guaranteed loans but only goes back to late 2012 and does not cover private portfolio loans or loans in private mortgage-backed securities. The HCAI covers the entire market for owner-occupied home purchase loans back to 1998 but it omits other types of mortgages, excludes the 1990-1997 period that we cover, uses incomplete data for pre-2013 Enterprise loans, and pays less attention to imputing missing loan information rates.[3] In short, we substantially improve the scope and quality of stressed default measures that already have a wide audience.

Our paper makes numerous contributions to the literature. The first, and most basic, is that we document the characteristics of the entire home mortgage market back to 1990 and

[3] The NMRI has since been renamed the National Mortgage Default Rate (NMDR). For more on the NMDR and the HCAI, see https://www.aei.org/housing/mortgage-risk-index/ and https://www.urban.org/policy-centers/housing-finance-policy-center/projects/housing-credit-availability-index. The HCAI is based on analysis in Li and Goodman (2014).

provide a summary measure of risk going back almost as far. As a result, researchers and policymakers now have at their disposal more complete and more accurate historical information than before.[4] All the data presented in the paper and many other series are available for free download from the FHFA website at https:/www.fhfa.gov/papers/wp1902.aspx.

Among the insights provided by the longer historical analysis is that seeds of the financial crisis were planted in the 1990s. We show that the average CLTV and average DTI on both home purchase and refinance loans increased over the decade, and there was some decline in the average credit score. By 2000, the stressed default rate had already risen considerably for home purchase loans and for refinance loans as well after controlling for the effects of changes in refinance volume on loan risk (more on this important control below). This finding calls into question the common view that mortgage lending conditions were normal in the early 2000s (CoreLogic, 2017; Goldman Sachs, 2014; Urban Institute, 2018).

Although the stressed default rate cannot identify the influence of credit supply versus credit demand on mortgage risk, we provide separate evidence that credit supply expanded in the early and mid-2000s. We do so by documenting a compression of mortgage rate spreads between the riskiest and least risky loans. To cleanly identify changes in credit supply, we limit the analysis to loans for which lenders and investors fully bear the credit risk – private sector loans with no government guarantee and no private mortgage insurance. To our knowledge, this is the most detailed exploration to date of spreads across risk tiers.[5]

Our results contribute as well to the ongoing debate about the role of "subprime" borrowers in the housing boom and bust.[6] Mian and Sufi (2009, 2014, 2017a) have focused on an

---

[4] While we focus on mortgage risk, we wish to make clear that there may be substantial benefits to increased access to credit. We hope the metrics in this paper provide the basis for future benefit-cost analysis of policy decisions.

[5] For other discussions of mortgage spreads, see Demyanyk and van Hemert (2011), Justiniano, Primiceri, and Tambalotti (2021), Levitin and Wachter (2020), and Mian and Sufi (2017a). Anenberg et al. (2019) complement this evidence by showing that the credit supply frontier – defined as the largest loan available for a given set of loan and borrower characteristics – shifted up during the early and mid-2000s.

[6] Throughout the paper, we define low-credit-score borrowers as those with a credit score of less than 660 at origination. Such borrowers are often referred to as "subprime" in the academic literature. In other contexts, there is no agreed-upon definition of the term. Accordingly, we refrain from using the subprime label throughout the paper except in reference to others' work.

increase in credit availability for subprime borrowers, which they define as borrowers with a credit score less than 660, while several recent papers have argued that the credit expansion was more widespread (Adelino, Schoar, and Severino, 2016, 2017; Albanesi, Di Giorgi, and Nosal, 2017; Conklin et al., 2018; Ferreira and Gyourko, 2015). Our results favor the latter view. We find that the below-660 share of mortgage loans was flat on net from 2000 through 2006, the primary years of the housing boom, though there was some increase in the 1990s. In addition, the stressed default rate on loans to below-660 borrowers rose largely in sync with that on loans to higher-score borrowers.

Another question of interest is the role played by loans with risky product features in the buildup of risk during the boom. We define such loans to be those having one or more of the following features: low or no documentation of borrower income, non-standard amortization of loan principal, and a term greater than 30 years.[7] We find that less than half of the overall rise in the stressed default rate owed to these risky product features; the majority of the rise stemmed from more "plain vanilla" forms of borrower risk, such as an increase in DTIs. Thus, a narrative that focuses primarily on risky product features overstates their role during the boom and underplays the risk-increasing effect of more prosaic forms of leverage.

The next part of the paper bring house price risk explicitly into the analysis, using a method based on Smith and Weiher (2012) and Smith et al. (2016) to construct severe potential house price shocks.[8] The shocks vary over time and geography based on FHFA house price indices for states and metropolitan areas. We adjust each loan's reported CLTV at origination to reflect the lower house value from the assumed negative shock to house prices. With the large house price increases starting in the late 1990s, these "shock CLTVs" move further and further

[7] This definition adopts some of the key product features that make a loan ineligible for Qualified Mortgage status under the Dodd-Frank Act and regulations adopted by the Consumer Financial Protection Bureau. Non-standard amortization refers to loans with an interest-only period, a negative amortization period, or a balloon payment.
[8] A modified version of this method is now in use by FHFA to determine capital requirements for Fannie Mae and Freddie Mac. See https://www.fhfa.gov/Media/PublicAffairs/Pages/Final-Rule-on-Enterprise-Capital.aspx for a summary of the revised Enterprise Regulatory Capital Framework, effective in 2021; for additional details, see https://www.federalregister.gov/documents/2020/12/17/2020-25814/enterprise-regulatory-capital-framework.

above the reported CLTVs. The average shock CLTV peaks in 2006 at more than 120 percent, illustrating the enormous house price risk borne by lenders and mortgage insurers at that time.

Although our full dataset ends in 2019, the final part of the paper uses a subset of the data to provide an update extending into the COVID-19 episode. This update includes Enterprise, FHA, and VA loans and runs through the third quarter of 2021. The update shows that the riskiness of mortgage originations fell during this period as the mix of both home purchase and refinance loans shifted toward lower-risk borrowers.

2. Data

We rely on several sources of loan-level data over the period 1990-2019. For Enterprise loans, we use the internal data maintained by FHFA, the Mortgage Loan Information System (MLIS). The MLIS dataset covers all mortgage loans acquired by the Enterprises, providing detailed information about loan characteristics and performance.[9] As noted in the introduction, we believe our paper is the first to use the full book of Enterprise loans over such a long history.[10]

The internal FHFA data are much more comprehensive than the public-use loan-level Enterprise datasets that first became available in 2013, which were limited to fixed-rate mortgages with full documentation and standard amortization of loan principal and that excluded many seasoned loans, Alt-A loans, and loans made under affordable housing programs.[11] As a result, those public-use datasets understated the risk in the Enterprise guarantee book, especially during the housing boom. In April 2021, the Enterprises released expanded public-use datasets at the behest of FHFA to promote greater transparency. The data now available cover about 95

---

[9] MLIS includes all Enterprise acquisitions of whole loans regardless of whether the loans were securitized or retained in their portfolios. It excludes the private-label mortgage-backed securities purchased by the Enterprises. A separate dataset, described below, captures the private-label securities purchased by the Enterprises and other investors.

[10] Other research has studied loan performance using internal Fannie Mae or Freddie Mac data over shorter periods. For example, Fout et al. (2018) use internal Fannie Mae loan data over 2002-2013, while Firestone, Van Order, and Zorn (2007) use internal Freddie Mac data over 1993-1997.

[11] The publicly available data are Fannie Mae's Single-Family Loan Performance Dataset and Freddie Mac's Single-Family Loan-Level Dataset, which are posted at http://www.fanniemae.com/portal/funding-the-market/data/loan-performance-data.html and http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.html, respectively.

percent of the home purchase and refinance originations since 2000 that were acquired by the Enterprises. However, some key loan characteristics, such as loan documentation status, are only reported by one of the Enterprises. In addition, the internal FHFA data that we use continue to be the only comprehensive source of Enterprise loan-level data before 2000.

For loans securitized in the private market, we use CoreLogic's dataset on non-agency residential mortgage-backed securities. The CoreLogic dataset is the most comprehensive source for loans included in private-label securities (PLS). It covers all segments of the PLS market, with the loans in the dataset accounting for more than 90 percent of the entire market. Many previous studies have used this dataset, including Adelino, Frame, and Gerardi (2017), Demyanyk and Van Hemert (2011), Houghwout, Peach, and Tracy (2008), Keys et al. (2010), Mayer, Pence and Sherlund (2009), and Palmer (2015). Our analysis builds on these studies by constructing much longer historical time series.

For loans held in portfolio by private lenders, we use the Loan-Level Market Analytics (LLMA) dataset from CoreLogic and the McDash dataset from Black Knight, Inc. Both datasets compile information provided by large loan servicers, though the set of servicers differs across the two datasets. Studies that have used one of the servicer datasets include Bubb and Kaufman (2014), Courchane, Kiefer, and Zorn (2015), DeFusco, Johnson, and Mondragon (2020), Foote et al. (2010), and Keys, Seru, and Vig (2012), among others. In contrast to these studies, we merge the two datasets to provide broader coverage than with either one alone.

Distinguishing portfolio loans from other conventional loans in the servicer datasets is not always straightforward. LLMA and McDash include investor codes to identify whether a loan was acquired by the Enterprises, but the information in this field is sometimes incomplete or missing. In addition, neither LLMA nor McDash identify whether a loan was securitized in the PLS market. We use the procedure described in Appendix A to remove conventional loans that we assess to be Enterprise or PLS loans. The remaining conventional loans constitute our portfolio loan dataset.

For FHA and VA loans, we use the two servicer datasets through 2012 and then switch for 2013-2019 to Ginnie Mae data processed by the AEI Housing Center. The Ginnie Mae dataset is a near-census of FHA and VA loans and provides detailed information about risk characteristics at origination. It is more comprehensive than the servicer datasets but is only available starting in late 2012.[12]

For all of these market segments, we use data on first-lien, 1-4 unit home mortgage loans that pass a variety of data quality filters. These filters remove, for example, loans that lack information on such variables as loan term or property location.[13] In all, the filters remove less than 2 percent of Enterprise loans in the source data, about 5 percent of FHA/VA loans, and about 10 percent of both PLS loans and the private-sector loans that are candidates for inclusion in our portfolio loan dataset.[14] After applying the filters and removing duplicate loans, the dataset includes more than 200 million home mortgage loans originated from 1990 through 2019.

Despite using the best available mortgage data, information on key risk factors is missing for many loans. This is especially true for the early years of our analysis, when the reporting of credit scores and DTIs is very spotty. Although we could simply drop observations with missing data, we would lose the non-missing data fields that contain useful information on the likelihood of default. Furthermore, this information is not missing solely by chance, as the likelihood of a missing risk factor is correlated with observable borrower and loan characteristics. Hence, dropping observations with missing information would bias the sample and resulting statistics.

To be able to use loans with incomplete information in our analysis, we impute missing values for the following risk factors: DTI, credit score, CLTV, loan documentation status, amortization status, occupancy status, and type of refinance loan (rate-and-term versus cash-out). Table 1 shows the share of each of these risk factors that must be imputed in each market

---

[12] The Ginnie Mae data cover not only FHA and VA loans, but also loans guaranteed by the Rural Housing Service (RHS), which are not included in this analysis. The source data for FHA, VA, and RHS loans are available at https://www.ginniemae.gov/Pages/profile.aspx?src=%2fdata_and_reports%2fdisclosure_data%2fPages%2fdatadownload_bulk.aspx.

[13] See Appendix A for full detail on these data quality filters and all other aspects of our data preparation.

[14] These private-sector loans are only candidates for the portfolio dataset because we still have to remove the PLS loans in the servicer datasets to avoid double-counting these loans. In addition, because we merge data from the servicer datasets for portfolio loans and FHA/VA loans, there are many duplicates that must be removed. The percentages cited in the text represent the shares of loans in both groups that are removed by the data quality filters over and above the removal of duplicates and PLS loans.

segment. We impute half of the DTIs for PLS loans and more than half for portfolio and FHA/VA loans. In addition, about 35 percent of portfolio loans require imputations for credit scores and 40 percent require imputations for documentation status. For the other cells in Table 1, imputations are less frequent or are not required at all.[15]

Figure 1 displays the share of imputations by year for the entire dataset of purchase and refinance loans. As shown, nearly all credit scores and DTIs are imputed for 1990-1992. The imputation share for DTIs falls dramatically in 1993 and the share for credit scores drops notably in 1993, 1994, and 1996. Improved reporting for Enterprise loans largely accounts for these changes. For the other risk factors shown in the figure, the imputed share is relatively low in all years.

The near-complete reliance on imputed DTIs and credit scores in the early 1990s raises concerns about the accuracy of those results. To ensure that our findings are anchored to real data, we do not present results for DTIs and credit scores until the first year in which these risk factors are well reported by the Enterprises, the dominant institutions in the mortgage market. This rule implies that our DTI and credit score results begin in 1993 and 1994, respectively. By 1993, more than 90 percent of Enterprise loans had a reported DTI, up sharply from less than 10 percent in prior years. In 1994, one of the two Enterprises began to systematically report credit scores, with the other following suit in 1996. Because the credit profiles for the two Enterprises are similar, the data available as of 1994 provide a very good estimate for the full Enterprise book. Given the central role of credit scores as a risk factor, we begin reporting the stressed default rate in 1994. Anchoring to well-reported Enterprise data supports the accuracy of our

---

[15] Table 1 indicates that we do not impute occupancy status for FHA/VA loans. Although some FHA/VA loans in our source datasets have missing data on occupancy status, we assume that these loans are for primary owner-occupied properties given that second-home loans and investor loans make up a very small share of the loans with reported occupancy status. For Enterprise loans, documentation status is well reported in the MLIS dataset (we impute this variable for only 2 percent of loans), but we use the reported information differently across the two Enterprises. For one Enterprise, we classify a loan as having less than full documentation if it is either reported as such by the documentation status variable or is classified as an Alt-A loan; for the other Enterprise, we rely solely on the documentation status variable. For both Enterprises, we determined that the resulting share of loans with less-than-full documentation is in general agreement with information from the Securities and Exchange Commission and Financial Crisis Inquiry Commission.

results for the mortgage market as a whole, even though many non-Enterprise loans require imputations in 1994 and beyond.

As detailed in Appendix A, the imputations for all variables are done under the general framework of multiple imputation (Rubin, 1976).[16] Within this framework, we employ predictive mean matching (PMM) as developed by Little (1988). The PMM procedure uses regression-based predictions to find the nearest "donors" of a given variable for loans with missing values. We randomly select one of the eight closest donors to provide the missing value. PMM is well suited for imputing variables that are not normally distributed, such as credit scores and DTIs, and it avoids the convergence problems that can affect other methods, such as logit models, for predicting categorical variables.

The PMM models are estimated separately for each market segment. To allow the estimated parameters to vary over time, we use the shortest estimation window that we judge to have sufficient data for reliable estimation. For Enterprise loans, we estimate separate PMM regressions every month starting in 1995, with some pooling of months for earlier years; for other market segments, we use a mix of windows ranging from monthly to grouped years. In addition, all the models use a rich set of loan characteristics as explanatory variables.

Because the coverage of portfolio and FHA/VA loans in the servicer datasets is incomplete, we weight the included loans to be representative of the full national market using data reported by lenders under the Home Mortgage Disclosure Act (HMDA). We adjust the loan counts in HMDA to include loans not subject to HMDA reporting and then distribute the counts across a four-way matrix of origination year, loan type (purchase or refinance), state, and loan amount ranges. We then construct cell-level weights for portfolio loans and FHA/VA loans that equal the adjusted HMDA count in each cell divided by the count from our servicer datasets.

Despite this weighting, the servicer data still may not be fully representative of the national market if the loans in a given cell have risk characteristics that differ from the universe

---

[16] Multiple imputation preserves both the first and second moments of the underlying data, which is critical to our exercise; the relation between risk factors and default is often highly non-linear, making preservation of the second moment in the imputed data of paramount concern.

of loans in that cell. This issue can only be addressed by benchmarking to a representative external database. The National Mortgage Database (NMDB), an ongoing project jointly managed by FHFA and the Consumer Financial Protection Bureau, has the potential to provide such an external benchmark. The NMDB, still under development, is drawn from a 5 percent random sample of mortgage files at a national credit repository dating back to 1999 (Avery et al., 2017). As described in Appendix A, we make some use of the NMDB and other benchmark information.

3. Stressed Default Rates

3.1 Definition and Interpretation

We calculate a stressed default rate for every loan in the dataset based on the observed default experience of similar loans originated nationwide in 2006 and 2007. We define a loan to have defaulted if it was ever 180 days delinquent or was terminated with less than full repayment of the outstanding loan balance. Stressed default rates are calculated by measuring lifetime loan performance for fine disaggregations of the 2006-2007 cohort and then applying the cell-level default rates to all loans originated between 1994 and 2019. The cells are highly granular and span a number of loan-level risk factors.

It is important to be clear about the interpretation of our default measure. The stressed default rate for a given loan represents its expected performance had it been hit shortly after origination with a replay of the financial crisis (including the observed policy response) ***and*** experienced the national average decline in house prices. The second part of the interpretation holds because the observed performance of 2006-2007 originations nationwide is used to estimate the stressed default rates. Given that the drop in house prices during and after the crisis varied enormously across localities, the stressed default rate calculated in this way will only be valid for a national portfolio of loans. In section 6, we introduce an extension of the stressed default rate that incorporates location-specific house price shocks and thus is useful for states and metropolitan areas.[17]

The stressed default rate is the analogue in the mortgage market to crash tests for motor vehicles or wind ratings for doors and windows in hurricane zones. In all of these cases, the goal is to assess performance under severe stress. This measure of mortgage risk should be of primary interest to policymakers because the stability of the mortgage market depends on its ability to withstand extreme events.

The stressed default approach rests on an identifying assumption – that unobserved characteristics of mortgage borrowers or unobserved aspects of the origination process do not vary substantially over time. This assumption comes into play because we use default tables based on 2006-2007 originations to calculate the stressed default rate for other origination cohorts. That is, we match loans originated in years other than 2006-2007 to the 2006-2007 loans in its cell of a default table. By doing so, we assume there are no "vintage" effects at the cell level. Although such vintage effects may exist, there is no consensus in the small literature on this topic about the size of any such effects (see Foote and Willen, 2018).

One piece of evidence provides some reassurance about the matching exercise. We check for drift over time in the average values within the credit score, CLTV, and DTI buckets that we use to construct stressed default rates. Finding widespread instances of drift would call into question the validity of the matching. As it turns out, the time variation for all but a handful of buckets is minimal.[18]

[17] The stressed default rate is similar in spirit to a counterfactual exercise in Palmer (2015) that estimates what the default rate for subprime PLS loans originated from 2003 to 2007 would have been if each vintage had been exposed to the house price declines experienced by the 2006 vintage. For the 2003 vintage, exposure to that stress would have doubled the observed default rate.

[18] The greatest variation over time is in the lowest credit score bucket (scores of 300-579), However, this is a sparsely populated bucket, and the variation reflects year-to-year noise rather than drift. There is also some variation in the highest credit score, DTI, and CLTV buckets, but the differences are too small to materially affect default risk (in the case of the credit score and DTI buckets) or affect only a few years (in the case of the CLTV bucket).

3.2 Calculation Details

To capture the wide variation in default rates across the 2006-2007 originations, we calculate eight separate default tables for home purchase loans and do the same for rate-and-term refinance loans as well as for cash-out refinances. In each case, as shown in Table 2, four tables apply to different types of fixed-rate mortgages, with a parallel set of tables for adjustable-rate mortgages (ARMs). The four default tables cover each combination of documentation status (full doc or low/no doc) and amortization of principal (standard or non-standard amortization). Each table contains 320 cells to account for all combinations of eight credit score buckets, eight CLTV buckets, and five DTI buckets.[19] Given the eight tables and 320 cells per table, we allocate the 2006-07 home purchase originations across 2,560 cells in total, and do the same for rate-and-term refinances and for cash-out refinances.[20]

We produce separate default tables for Enterprise, PLS, FHA, and VA loans. For portfolio loans, we elected to use the default tables for Enterprise loans due to the more comprehensive information regarding loan performance in the FHFA data than in the servicer datasets. The maintained assumption is that portfolio loans originated in 2006-07 with the same risk characteristics would have the same default experience as Enterprise loans.[21]

All of the default tables pertain to 30-year mortgages secured by owner-occupied properties. We then apply adjustment factors to these tables for loans with 15- or 20-year terms, loans with 40-year terms, loans originated to investors, and loans used to purchase second

[19] The credit score buckets are 579 or less, 580-619, 620-639, 640-659, 660-689, 690-719, 720-769, and 770+; the CLTV buckets (in percent) are 60 or less, 61-70, 71-75, 76-80, 81-85, 86-90, 91-95, and 96+; and the DTI buckets (in percent) are 33 or less, 34-38, 39-43, 44-50, and 51+. All non-integer values for CLTVs and DTIs are rounded up to the nearest integer.

[20] For the cells in the tables with 100 or more loans, we use the actual default rate for the loans in that cell. For cells with less than 100 loans, we use the average estimated default rate from a logit model built using the loans that generate the default tables; details of the logit model are available on request. Cells with no loans use the average estimated default rate from the logit model for a loan with the median value for each risk factor that defines the bucket.

[21] For FHA and VA loans, we face a similar issue concerning incomplete performance histories in the servicer datasets but were not as comfortable using the Enterprise default tables. Instead, for FHA and VA loans with performance records that are censored after *m* months, we estimate the probability of default from loans with full performance histories conditional on having survived to month *m*. The resulting default tables for VA loans show that these loans default much less often than FHA loans with similar characteristics, consistent with results in Goodman, Seidman, and Zhu (2014).

homes. To calculate these adjustment factors, we estimate separate logit default models for Enterprise, PLS, FHA, and VA loans originated in 2006-2007, each divided into purchase loans, rate-and-term refinances, and cash-out refinances, for a total of 12 models. Each logit includes dummy variables for credit scores, DTIs, and CLTVs using all the buckets in the default tables, along with dummies for loan type (fixed rate vs. ARM), documentation status (full doc vs. low or no doc), amortization status (standard vs. non-standard), loan term (15-20 year, 30-year, 40-year), and occupancy status (primary owner-occupied, second home, investor). After estimation, we run loans with one of the characteristics of interest (say, investor loans) through the model as reported and a second time with the characteristic set to the baseline value (primary owner-occupied in this case). The average default rate with the characteristic as reported divided by the average default rate with the characteristic "turned off" represents the adjustment factor.[22] The estimated adjustment factors reduce the stressed default rates for 15- and 20-year loans, increase or occasionally leave unchanged the rates for investor loans and second-home loans, and increase the rates for 40-year loans with one exception. The exception is for Enterprise loans, where the adjustment factor reduces the stressed default rate for 40-year loans relative to observably identical 30-year loans. Although this result is counter-intuitive, it affects only 0.2 percent of the Enterprise loans in our dataset.[23]

3.3 Example

As an illustration of the default tables, Table 3 displays two slices from the default table for Enterprise 30-year fixed-rate home purchase loans with full documentation and standard amortization. The top panel shows the slice of the table for credit scores of 720-769 and the bottom panel shows the slice for credit scores of 660-689. To help visualize the pattern of default

[22] The actual adjustment factors used in the analysis include a refinement beyond this basic description. After calculating the logit-based estimate of each loan's default rate, we divide the loans with the characteristic in question into two groups – those with estimated default rates above the median and those below the median. We then calculate the adjustment factor for each group separately. This refinement allows the adjustment factors to have a different effect on defaults for relatively low risk and relatively high risk loans.

[23] We apply the Enterprise adjustment factors to portfolio loans, consistent with our use of the Enterprise-based stressed default tables for these loans. Note that the Enterprise adjustment factor for 40-year loans affects only about 3 percent of portfolio loans in our dataset.

rates in the table, the cells with default rates of 7 percent or less have green shading, those with default rates of 7.01 to 14 percent have orange shading, and those with higher default rates have red shading.[24]

Both panels show that higher DTIs and CLTVs increase stressed default rates, though the CLTV effect is considerably stronger. Moving from the lowest CLTV bucket to the highest in either panel increases the stressed default rate by a factor between 6 to 12, while moving from the lowest DTI bucket to the highest raises stressed defaults by a factor of only 1.5 to 2.5. The effect of credit scores can be seen by comparing the same cells across the two panels. This comparison indicates that lower scores are associated with substantially higher default rates. Although Table 3 does not show the lowest and highest credit score buckets, the difference in default rates across those buckets is comparable to that for CLTVs. Thus, the default tables embed standard results from empirical studies of mortgage default – that credit scores and CLTVs at origination are highly predictive of loan performance, with DTIs contributing some additional information (see, for example, Mahoney and Zorn, 1997; Haughwout, Peach, and Tracy, 2008; Dunsky, Kelly, and Lam, 2013; and Fout et al., 2018).[25]

The similarity of the default patterns in Table 3 to those in the literature suggests that default tables estimated using origination years other than 2006-2007 would generate a similar pattern for stressed default rates over time even if the absolute level of the stressed default series were different from that based on the 2006-2007 cohort. We tested this proposition using

[24] The default tables do not control for the influence of house prices because they pool loans originated across the U.S., as mentioned earlier. In Section 6, however, we amend the baseline default tables to incorporate loan-level effects of house price movements. Those tables display the same general effects of CLTVs, credit scores, and DTIs on stressed defaults as in Table 3.

[25] A few studies have asserted that DTIs not only have less predictive power than credit scores and CLTVs, but add little to default models (Avery et al., 1996; Foote et al., 2010). This conclusion is not supported by our results. The early research cited by Avery et al. (1996) was conducted in an era when underwriting practices generally kept DTIs within strict limits (Quercia and Stegman, 1992; Herzog and Earley, 1970), which may explain why these studies found that DTIs had so little explanatory power. More recently, Foote et al. (2010) studied the influence of DTIs on defaults for mortgages originated during 2005-2008, finding much smaller effects than those shown in Table 3. Their analysis, however, only tracked loan performance through the end of 2008, thus omitting about half of the total jump in the unemployment rate during the financial crisis and the associated income shock. In contrast, our analysis – and that in DeFusco, Johnson, and Mondragon (2020) – tracks loan performance over a much longer period, which allows the full effect of payment burdens to emerge.

Enterprise loans originated in 1998-1999, a cohort with much lower overall defaults than the 2006-2007 cohort. We found that the two stressed default series line up very closely after allowing for the difference in level due to the much greater severity of the shock to the 2006-07 cohort than to the 1998-99 cohort. This tight correlation implies that the time variation in the stressed default series presented in Section 5 is not an artifact of using the 2006-2007 cohort to construct the default tables, but instead reflects long-standing connections between risk factors and default outcomes.

3.4 Comparison to Other Default Models

The default-table approach corresponds to estimating a large set of standard logit or probit default models, each with the same structure as the default tables. To see this, consider a single default table – say, the table for Enterprise purchase loans with fixed rates, full documentation, and standard amortization. That table (and every other default table) contains 320 cells covering every possible combination of the credit score, CLTV, and DTI buckets, with the value in each cell equal to the default rate for the loans in that cell. Instead of directly assigning the default rates, we could have used the loans in that table to estimate a logit or probit default model in which the explanatory variables mimic the structure of the default table. That is, the explanatory variables would consist of 320 dummies that correspond to every combination of the credit score, CLTV, and DTI buckets. This model would return the mean probability of default for the loans in each cell, which is the same as the directly-assigned default rate. Since we have a total of 96 default tables (eight tables for each of the three loan purposes in the Enterprise, PLS, FHA, and VA market segments, or 8 x 3 x 4), that implies the estimation of 96 separate default models to replicate our default tables.

The default-table approach also can be compared to machine-learning models. In contrast to the default tables, which impose a pre-defined cell structure, machine-learning models let the data determine an optimized partition of loans into cells with different default risk. Regression trees, first proposed in Breiman et al. (1984) are a popular machine-learning approach. In our context, a regression-tree model begins by comparing splits of the loan data on every included risk factor to find the most informative split for predicting defaults. It then repeats this step multiple times to fill out the decision tree. Cross-validation checks determine the amount of tree complexity associated with the maximum predictive power. Random forests, proposed in Breiman (2001), are an extension of regression trees. They average over multiple regression trees, where each tree is constructed from a subset of risk factors and loans. An advantage of random forests is that they can reduce the risk of overfitting the data with a single regression tree.

As described in Appendix B, we estimate both machine-learning models as a robustness check on our default-table methodology.[26] The results show that the stressed default rates produced by the regression trees and the random forests are very similar to those obtained with the default tables, which provides important support for our results. However, the regression trees require many fewer cells in order to effectively separate loans by default propensity. In particular, while the default tables include 2,560 cells for any loan purpose in a given market segment (eight tables, each with 320 cells), the regression trees generate similar results with 250 to 700 cells. This shows that the *a priori* structure of the default tables is more granular than would be necessary had loans been grouped through machine learning.[27]

4. Origination Volumes and Market Shares

Figure 2 displays origination counts for home purchase loans and refinance loans. As shown in the upper panel, refinance volume exceeded purchase volume in a majority of years, with refinance loans representing about 55 percent of the total loan count over 1990-2019. Refinance activity is much more volatile year to year than is purchase lending, as borrowers take advantage of drops in mortgage rates to refinance existing loans and then largely withdraw from the market when the refinance option is out of the money. The bottom panel scales the number of refinance loans by the prior-year stock of mortgages to indicate the share of the stock that was

[26] We create the regression trees with the package rpart in R and the random forests with the package Rborist.
[27] That said, the default tables provide a much simpler visualization of risk than the complex flow chart in a regression tree. The first split in all of our regression trees is on either the credit score or CLTV, which is intuitive, but the subsequent levels of the tree quickly become harder to keep in mind because of the expanding number of conditional splits of the data.

refinanced.[28] This share ranges from about 2 percent in 1990 to nearly 35 percent in 2003, which dwarfs the share in all other years. Apart from the 2003 refinance tsunami, the largest refi waves were in 1993, 1998, 2001, and 2002. As we will demonstrate below, this annual variation in refinance activity has a significant effect on the risk characteristics of refinance loans.

Figure 3 shows the annual shares of home purchase loans (top panel) and refinance loans (bottom panel) by market segment over 1990-2019. The shares in both panels are based on loan counts. For purchase loans, the Enterprises generally accounted for about 40 to 50 percent of the market over this period, with a spike above 50 percent in 2007. FHA/VA loans represented roughly another 20 percent of the market in the 1990s, then shrank during the boom years in the 2000s, as they gave up market share to PLS loans, before rebounding in importance after the financial crisis. Altogether, government guaranteed loans have accounted for 70 to 80 percent of the purchase loan market since 2009, a historically high share. The increased share has coincided with a significant reduction in loans held in portfolio and the near-disappearance of the PLS market.

For refinance loans, the Enterprises have played an even larger role than for purchase loans, accounting for slightly more than half of the market on average over 1990-2019. Before the financial crisis, the Enterprise share rose sharply during years with heavy refinancing activity and then dropped back when the refi wave abated.[29] Since the crisis, the Enterprise share has been higher on average and more stable, with a peak in 2011-2013 when the Home Affordable Refinance Program (HARP) was most active.[30] In contrast to the Enterprises, FHA and VA had been relatively small players in the refinance market before the financial crisis. However, their

---

[28] The data for the stock of mortgages come from the American Community Survey, Table S2506, starting in 2005. For earlier years, we use data from the American Housing Survey, Table 3-15. The AHS data are only published every other year, and we linearly interpolate between consecutive surveys to fill in the missing years. For both the ACS and the AHS, the stock is measured as the number of owner-occupied units with a mortgage.

[29] This pattern likely reflects, at least in part, capacity constraints on lenders when refi volume spikes. In this situation, lenders prioritize relatively high-quality loans that require limited effort to underwrite (Sharpe and Sherlund, 2016). Such loans are prime candidates for sale to the Enterprises.

[30] HARP allowed borrowers to refinance Enterprise loans that would otherwise be ineligible for refinancing because the loans had too little (or negative) equity. See https://www.fhfa.gov/AboutUs/Reports/ReportDocuments/Refi_4Q2016.pdf for details.

role grew after the crisis, and since then Enterprise and FHA/VA loans together have accounted for 70 to 80 percent of the refinance market, about the same share as for purchase loans.

5. Main Results

This section reports the main results from our analysis of the data described in Section 2, starting with the stressed default rate and individual risk factors, before turning to three topics that pertain to the housing boom: the role of risky product features, the path for mortgage rate spreads, and evidence concerning the role of low-credit-score borrowers.

5.1 Stressed Default Rate

Figure 4 shows the results for the stressed default rate – our summary measure of risk – from 1994 to 2019. The top left panel plots the series for home purchase loans, while the other panels cover refinance loans and the combination of purchase loans and refis. All panels use the same scale to facilitate comparisons across the loan types.

The top left panel shows that mortgage risk for purchase loans held in a narrow band from 1994 to 1998. It then moved up to a higher range for 1999-2003, followed by a steep ascent over 2004-2006, peaking at 35 percent in 2006. The jump from an average of 17 percent for loans originated over 1994-1998 to 35 percent for loans originated in 2006 implies that the earlier origination cohorts would have been much more resilient than the 2006 cohort had they been exposed to the same shock. Nonetheless, even a 17 percent default rate is extremely high, which speaks to the rigor of the stress test that we perform.

From 2007 through 2013, the stressed default rate fell sharply as credit standards tightened in the wake of the housing bust. Since then, the risk measure has edged higher, but as of 2019 it remained near the bottom of the range observed since 1994. The low level in recent years owes importantly to the high credit scores for most borrowers and to the very limited use of loans with risky product features such as a period with interest-only payments.

The top right panel shows the stressed default rate for refinance loans (the solid blue line), along with refinance volume as a share of the prior-year stock of loans (the dashed grey line, repeated from the bottom panel of Figure 2). As with purchase loans, the stressed default rate for refis peaked in 2006 at a level far above those in the 1990s, then fell sharply during the financial crisis and remained relatively low through 2019.

However, compared to purchase loans, the stressed default rate for refis is more volatile year-to-year, especially prior to the financial crisis, when the rate dropped in 1998 before jumping in 2000 and falling again in 2001-2003. This volatility is associated with refi booms and busts, as can be seen from the negative correlation between the stressed default rate and refi activity. During refi booms, higher-quality borrowers with no immediate need for cash enter the market to lower their mortgage rate. Their entry reduces the average risk of the borrower pool, which pushes down the stressed default rate. In contrast, when the option to refinance is generally out of the money, these borrowers move to the sidelines, leaving the pool tilted toward higher risk, credit constrained borrowers who cash-out some of their accumulated equity or lower their monthly payment by extending their mortgage term even if the new rate is no lower than their old rate. This adverse shift in the borrower pool increases the stressed default rate.

We use a simple regression method to adjust for this refi volume effect. Let R_t denote the number of refinance loans scaled by the prior-year stock of mortgages, the measure of refi volume shown in the top right panel of Figure 4. We then estimate the following regression for each variable of interest, Y_t, which includes the stressed default rate and a number of loan characteristics:

$$Y_t = \alpha + \beta_1 D_1 (R_t - \bar{R}) + \beta_2 D_2 (R_t - \bar{R}) + \sum_i \delta_i t^i + \epsilon_t$$

where D_1 is a dummy variable that equals one for years through 2007, D_2 is a dummy for the years 2008-2019, $\bar{R}$ is the mean of R_t over the entire period through 2019, and $\delta_i t^i$ represents the i^{th} term in a polynomial function of time. The dummies allow the effect of refi volume to differ between the pre-crisis period and later years, and the polynomial in time allows Y_t to vary flexibly over the sample period after controlling for the refi volume effect.[31] Our adjustment to Y_t to account for changes in refi volume is $\beta_1(\bar{R} - R_t)$ for years through 2007 and $\beta_2(\bar{R} - R_t)$ for

[31] We estimated the regression for each variable Y_t with polynomials ranging from first to sixth order and used the Bayesian information criterion (BIC) to select the order of the polynomial. The BIC provides a consistent estimate of the appropriate order of the polynomial (Stock and Watson, 2015).

later years, in effect substituting the mean level of refi activity for the actual level in each year. Note that we only apply the adjustment when the relevant β coefficient is significant at the five percent level. In particular, if β_1 is significant but β_2 is not (the usual pattern), we apply the adjustment through 2007 but not for later years.

The lower left panel of Figure 4 shows the adjusted and unadjusted stressed default rate for refinance loans. The two series coincide over 2008-2019 because the estimate of β_2 is insignificant. However, for earlier years, the adjusted series (the dashed line) smooths out the wide swings in the unadjusted series before the financial crisis, revealing a sustained uptrend in the stressed default rate for refis from the mid-1990s to the mid-2000s.

The results for the aggregate of purchase and refinance loans are displayed in the lower right panel. Note that the adjusted series for the all-loan aggregate is a weighted average of the purchase loan series and the adjusted refi series for years when β is significant and the unadjusted refi series for years when it is not. The annual weights in the adjusted series reflect the actual count of purchase loans and a "normalized" count for refi loans, where the normalized count is the count implied by using the average ratio over 1990-2019 of refis to the prior-year stock of loans. The normalized weights control for an additional way that changes in refi volume can affect the aggregate results. Even if changes in refi volume do not have a significant effect on a particular refi series, the change in volume can affect the all-loan aggregate if the level of the purchase loan series differs from the level of the refi series.

As shown in the lower right panel, the adjusted stressed default rate for the all-loan aggregate trends up starting in the mid-1990s, while the unadjusted series inherits a damped version of the pre-crisis swings in the unadjusted refi series. During and after the crisis, the adjusted and unadjusted series move virtually in lockstep.[32]

We can use these results to evaluate the common view that the early 2000s represent a normal period in the mortgage market, a benchmark by which to assess whether lending standards in other periods are loose or tight (CoreLogic, 2017; Goldman Sachs, 2014; Urban

[32] The barely visible differences between the series over 2008-2019 are due to the differences in weighting discussed in the previous paragraph.

Institute, 2018). The shading in Figure 4 covers the period 2000-2003 to facilitate the comparison to earlier years. For home purchase loans, the average stressed default rate over 2000-2003 was considerably above the average over 1994-1999. The increase is even more apparent for refinance loans and for the aggregate of purchase loans and refis after adjusting for changes in refinance volume.

Determining what constitutes "normal" lending conditions is inherently subjective. The lower right panel of Figure 4 shows why some analysts could have regarded the early 2000s as a normal period, as risk for the unadjusted all-loan aggregate had not yet moved up decisively. However, once the risk of refinance loans is adjusted to account for changes in refi volume, an alternative story becomes clear: the 2000-2003 window is near the midpoint of a large and steady increase in risk between the mid-1990s and 2006. Thus, the evidence presented here calls into question the view that the early 2000s represented a period of normal lending conditions.

The stressed default rates in Figure 4 pertain to the entire home mortgage market. Similar figures for the major segments of the market (Enterprise, portfolio, PLS, and FHA/VA) can be found in Appendix C. The takeaway from those figures is that the rise in risk both before 2000 and from 2000 to the market peak was widespread. Only FHA/VA loans had little increase in their stressed default rate; for these loans, the rate started at a high level and remained so through the market peak.[33]

5.2 Risk Factors

The time series for stressed default rates reflect changes in a variety of underlying loan and borrower characteristics. We briefly describe the evolution of these risk factors.

[33] As noted in Section 3, the definition of default underlying the stressed default rate is that a loan was ever 180 days delinquent (D180) or was terminated with less than full repayment. We include D180 in the definition of default in large part to capture loans that very likely defaulted but had incomplete performance information in our data, making it impossible to observe the final disposition of the loan. As a check, we used the Enterprise data to see how the stressed default rate would change if we had used the narrower default definition that requires termination with less than full repayment. We found that substituting the narrower definition slightly reduced the average level of the stressed default rate and, importantly, had essentially no effect on the pattern over time.

5.2.a. *DTIs, CLTVs, and Credit Scores*

Figure 5 shows annual data for the average DTI, average CLTV, and average credit score for the aggregate of all home purchase and refinance loans. These series have been adjusted for the effect of changes in refinance volume using the same methodology described above for adjusting the stressed default rate.

As shown in the top panel, the average DTI rose sharply on net from 1993 to 2007, with about half of the rise in place by 2000.[34] The average DTI then retraced a good part of this increase during and shortly after the financial crisis, before increasing again. As of 2019, the level had returned to that seen in the early 2000s, a move back toward the high debt payment burdens that were common during the boom.

The average CLTV (middle panel) rose substantially on net in the early 1990s, with further increases through 2006.[35] The average CLTV, like the average DTI, then fell sharply during the financial crisis. Since then, the annual pattern has been choppy, due in part to the impact of the HARP program. The loan-to-value limits in the HARP program were removed in late 2011, enabling many Enterprise borrowers to refinance loans with CLTVs well above 100 percent. As the HARP program wound down in subsequent years, the boost to the average CLTV diminished (see Appendix D for more on HARP loans).[36]

The bottom panel of the figure indicates that the average credit score declined over 1994-2000, was little changed over 2000-2006, and then jumped during the financial crisis. That

[34] Although Foote, Loewenstein, and Willen (2019) do not directly measure DTIs, they present evidence consistent with this increase in the average DTI. Using loan-level data, they regress the purchase loan amount at origination on borrower income and a set of control variables. They find that the relationship between the loan amount and income became weaker over the 1990s and into the 2000s, indicating that income constraints on loan size had loosened.

[35] The increase in the first half of the 1990s was concentrated among Enterprise loans and owed to a sharp rise in the share of loans with CLTVs of 95 percent or more. The longest similar measures of CLTVs in the literature start in 1996 (Adelino, McCartney, and Schoar, 2020) or 1997 (Ferreira and Gyourko, 2015), and thus omit the large run-up in the first half of the 1990s, highlighting the value of our longer history.

[36] A possible issue with the CLTVs in our dataset is that they may not fully capture the second liens taken out in connection with home purchases (so-called "piggyback" loans), which became increasingly common as the housing boom went on. We examine this question in Appendix A and conclude that the under-reporting of piggybacks in our dataset does not materially bias the results shown in Figure 5.

increase was still largely in place in 2019.[37] Importantly, the results for 2000-2006 show that the rise in risk for home mortgage loans during the height of the housing boom was not driven by a shift toward lower-score borrowers. We provide more detail below on the low-credit-score part of the mortgage market.

5.2.b. *Other risk factors*

The top panel of Figure 6 shows the share of mortgage loans with less than full documentation of the borrower's income or assets. In 1990, about 20 percent of home mortgages had less than full documentation. However, low/no doc loans then largely disappeared for the next several years as lenders experienced worse-than-expected performance on earlier originations (Sichelman, 1990; Pacelle, 1991). Starting in the late 1990s and continuing through the first half of the 2000s, low/no doc loans grew again in popularity. By 2006, these loans accounted for nearly 40 percent of the market. Many borrowers used low/no doc loans to overstate their true income and hence their borrowing capacity in the face of rising house prices (see, for example, Mian and Sufi, 2017b). Since the housing bust, low/no doc loans have become much less common because the ability-to-repay rules in Dodd-Frank require the documentation of borrower income with limited exceptions. Most of the low/no doc loans since the financial crisis have been streamline refinances of existing FHA and VA loans.

The bottom panel of the figure shows the share of loans with non-standard amortization, which includes interest-only, negative amortization ("neg-am"), and balloon loans. From 1990 through 2002, fewer than 10 percent of mortgage loans had non-standard amortization. The share then moved sharply higher, reaching 30 percent in 2005 and 2006. Like low/no doc loans, these loans were used to maintain borrowing capacity at a time when household income was rising much more slowly than house prices (Amromin et al., 2018; Garmaise, 2018). After the housing

[37] Because we compare credit scores over many years, an important question is whether a given score indicates the same level of credit risk under successive FICO scoring models. At the time of a model changeover, FICO ensures that borrowers assessed to have the same risk under both the old and the new models will have no change in their FICO score. Some borrowers, though, will be assessed as more risky under the new model than the old, and their scores will fall. Others will be assessed as less risky, and their scores will rise. These "swaps" could have a small impact on the distribution of scores. Overall, however, changes in scoring models do not appear to seriously compromise the comparability of credit scores over time. We thank Joanne Gaskin, Vice President, Scores and Analytics, at FICO for this information, obtained via email communication dated June 27-July 1, 2019.

bust, these loans became extremely uncommon, largely because interest-only and neg-am loans and the vast majority of balloon loans do not qualify for Qualified Mortgage (QM) status under the rules issued by the Consumer Financial Protection Bureau.[38]

Other salient risk factors, shown in Appendix C, include the ARM share of loans, the investor share, and the share with terms of 15 or 20 years. Changes in all three factors contributed to the rise in risk. The ARM share jumped from 2001 to 2005, the investor share rose over the same period, and share with terms of 15 or 20 years trended down from the early 1990s through the mid-2000s.

5.3 Role of Risky Product Features During the Boom

How much of the build-up of risk during the housing boom can be traced to standard forms of borrower leverage (such as rising DTIs) versus the growing use of risky products (such as interest-only loans)? We address this question by comparing the stressed default rate shown above, which incorporates the full set of risk factors, to an alternative stressed default rate that omits a specified set of risky product features.

We define risky product features as those characteristics that make a loan ineligible for QM status – non-standard amortization, less than full documentation, and a term greater than 30 years.[39] To calculate our alternative stressed default rate that excludes risky product features, we run low/no doc loans and loans with non-standard amortization through the default tables as loans without those features and we remove the adjustment factor for loans with a 40-year term.

Table 4 displays the results of this exercise for the aggregate of purchase loans and refis. The first line shows the baseline stressed default rate that embeds all risk factors, while the next line shows the alternative stressed default rate that excludes risky product features. We focus on the change over 1994-2000, 2000-2006, and the full period from 1994 to 2006. The data in the table have not been adjusted for changes in refi volume; this adjustment has no material effect on the results.

[38] For a summary of the QM rule, see Consumer Financial Protection Bureau (2014).
[39] Technically, balloon loans originated by small lenders in rural or underserved areas remain eligible for QM status, but we do not distinguish these loans from other balloons for simplicity.

As shown, the baseline stressed default rate rose 8.4 percentage points from 1994 to 2000, while the alternative series that excludes risky product features increased 6.7 percentage. Thus, 80 percent of the rise in mortgage risk during this early period owed to "plain vanilla" risk factors; risky product features played only a small role. Over 2000-2006, the contribution from plain-vanilla factors dropped to roughly 40 percent (5.4/12.3). It should be noted, however, that this percentage likely understates the contribution from plain-vanilla factors, as one of these factors, the reported DTI, was held down by the overstatement of income for low/no doc loans during the boom. Over the full period 1994-2006, plain-vanilla factors accounted for more than half of the rise of mortgage risk, even though risky product features have garnered a great deal of attention.[40] These results imply that a more balanced view would be appropriate.

5.4 Mortgage Rate Spreads

We have documented a sharp rise in the riskiness of home mortgage loans over a long period leading up to the housing bust. This finding, on its own, does not identify whether the increase in risk owed to an expansion of credit supply, greater credit demand, or both. However, the literature provides ample evidence that an easing of credit supply was at least part of the story.[41]

One of the key pieces of evidence supporting an expansion of credit supply is that mortgage rate spreads came down during the boom period (Demyanyk and van Hemert, 2011; Justiniano, Primiceri, and Tambalotti, 2021; Levitin and Wachter, 2020 (Fig. 7.9); Mian and

[40] Li and Goodman (2014) perform a similar decomposition and find a larger role for risky product features. This difference likely relates to their estimate of a very small rise in stressed defaults for Enterprise loans, with portfolio and PLS loans accounting for nearly the entire increase in the market-wide stressed default rate. Since risky product features were more common in portfolio and PLS loans than in Enterprise loans, these features end up dominating plain-vanilla leverage in the Li-Goodman decomposition. However, our results imply that they substantially understated the rise in risk for Enterprise loans through their use of the publicly-available data at the time.

[41] See Mian and Sufi (2017a) for a discussion of the evidence. Anenberg et al. (2019) provide a nice summary measure of the expansion of mortgage credit supply during the boom by estimating a supply frontier, which represents the maximum loan size that could be obtained for a given set of observable borrower and loan characteristics; they show that the frontier moved substantially higher from 2001 to 2005-2006. Despite this evidence, Albanesi, De Giorgi, and Nosal (2017) question whether there was any increase in credit supply based on their finding that mortgage debt growth was strongly correlated with income growth over this period. This correlation, however, is consistent with a broad-based increase in credit supply that left intact the demand-side linkage between income and debt growth.

Sufi, 2017a). This decline implies that the supply of credit increased by more than enough to accommodate any outward shift in mortgage demand. Our data add to this body of evidence. In particular, we show that rate spreads for the mortgages with the highest stressed default rates compressed sharply during the boom years, which indicates that lenders and mortgage investors became significantly more willing to extend credit to high-risk borrowers. To our knowledge, this is the most detailed exploration to date of spreads across risk tiers.

To compute the spreads, we use only 30-year fixed-rate mortgages (FRMs) to avoid mixing loans with different terms and different product types. We also screen out loans with implausibly high or low rates. On the high side, we remove any loan with a reported mortgage rate above 20 percent, and on the low side, we remove any FRM with a reported rate more than 100 basis points below the 30-year FRM rate in Freddie Mac's Primary Mortgage Market Survey (http://www.freddiemac.com/pmms/).

Figure 7 presents the results. We limit our attention to loans for which lenders and investors fully bear the credit risk – PLS and portfolio loans without any private mortgage insurance – to get a clean read on changes in supply. Each panel in the figure shows the average annual spreads from 1998 to 2007 for loans with stressed default rates in progressively higher risk buckets measured relative to the mortgage rate for loans in the lowest stressed default rate bucket (0-10 percent). We report spreads beginning in 1998 because our data show limited pricing for risk in earlier years; risk pricing likely became commonplace only with the widespread use of credit scores.[42]

For PLS purchase and refinance loans, the level of spreads corresponds one-for-one with the ordering of the risk buckets, confirming that the market was pricing for risk. Spreads for the highest risk buckets fell dramatically between 2000-2001 and 2005, before turning up in 2006-

[42] The spreads we calculate are based on the loan's note rate, which excludes the effect of fees and points paid to the lender at closing. These fees and points are not reported in the servicer or PLS datasets. Separate data on conventional home purchase loans from FHFA's Mortgage Interest Rate Survey indicate that the effective mortgage rate that includes these charges averaged only about 10 basis points more than the note rate over 1998-2007, with a modest decline in the gap over the period. This was true both for the subset of loans with LTVs of 95 percent or more (which we use as a proxy for higher-risk loans) and the subset with LTVs of 80 percent or less (a proxy for lower-risk loans). Thus, using an effective rate rather than the note rate would have no material effect on our results.

2007, when the performance of subprime PLS loans began to deteriorate (Mayer, Pence, and Sherlund, 2009). For portfolio loans, spreads also compressed relative to the lowest risk bucket.[43] Overall, these results provide additional evidence of an expansion of credit supply for riskier borrowers during the housing boom.[44]

5.5 Borrowers with Low Credit Scores

An ongoing debate in the literature concerns the role of "subprime" borrowers in the housing boom and bust.[45] Mian and Sufi (2009, 2014, 2017a) have emphasized the importance of mortgage lending to marginal borrowers, who traditionally would have been denied credit. However, a number of recent papers have questioned the focus on this sector, arguing that (a) the growth of mortgage debt was spread across the credit score distribution, (b) loan defaults during the crisis were more skewed toward middle- and higher-income borrowers than in previous periods, and (c) house prices increased the most in areas with relatively weak lending to borrowers with low credit scores (Adelino, Schoar, and Severino, 2016 and 2017; Albanesi, De Giorgi, and Nosal, 2017; Conklin et al., 2018; Ferreira and Gyourko, 2015).

We make two contributions to this debate. First, using our data for the entire mortgage market, we track the low-credit-score share of loan originations. And, second, we examine whether low-credit-score loans became markedly riskier than other loans during the boom. Our

[43] It is not clear why the spread for the highest risk bucket for portfolio purchase loans lies below the spread for the next-highest bucket in some years. This anomaly aside, the portfolio loan spreads align properly with risk.

[44] Fuster, Ho, and Willen (2017) discuss an alternative measure of credit supply – the price of mortgage intermediation, calculated as the difference in price between a mortgage in the primary and secondary markets. They estimate this price of intermediation over 2008-2014 and find that its movements did not always line up closely with the spread that they consider (the primary market rate less an MBS yield), leading them to question the accuracy of spreads as a measure of credit supply. While acknowledging their valuable contribution to the literature, we stand by the usefulness of our spread results for two reasons. First, the spread decline that we document for the riskiest loans is several times larger than the maximal disconnect they find between the price of intermediation and the spread they consider. Second, the spread we calculate – the difference in the average rate on risky loans and the least risky loans – effectively differences out any market-wide factors that would cause divergent movements in the price of intermediation as they define it and the spread.

[45] As stated previously, we define low-credit-score borrowers as those with a credit score of less than 660 at origination. Such borrowers are often referred to as "subprime" in the academic literature, but in other contexts, there is no agreed-upon definition of the term. Accordingly, we refrain from using the subprime label except in reference to others' work.

definition of low-credit-score loans as those with scores below 660 follows Adelino, Schoar, and Severino (2016, 2017), Conklin et al. (2018), and Mian and Sufi (2009).

The top panel of Figure 8 displays the below-660 share over 1994-2019. Focusing on the period before the financial crisis, the share increased over 1994-2000 and then changed little on net over 2000-2006, mirroring the pattern for the average credit score in Figure 5. Thus, while there was some shift toward low-score borrowers during the 1990s, perhaps spurred by the increasing use of automated underwriting models, there is no evidence that this shift continued from 2000 through 2006, when the housing boom was in full swing. On balance, these results provide little support for the view that low-score borrowers drove the growth of mortgage lending during the boom.

The bottom panel of the figure compares the stressed default rate for loans below the 660 credit score threshold versus loans with higher scores. Notably, the stressed default rate for low-score loans moved in close alignment with that for higher-score loans before the crisis, indicating there was no systematic increase in the riskiness of low-score loans relative to other loans.

Taken together, these findings are not favorable to a "subprime-centric" view of the financial crisis. Although mortgage lending shifted toward low-score borrowers in the second half of the 1990s, there was no such shift from 2000 to 2006. In addition, the stressed default rate for low- and high-credit-score borrowers rose in tandem over the full 1994-2006 period.[46]

6. Incorporating House Price Shocks

6.1 Measuring Shock CLTVs

The national stressed default rate presented above is based on default tables that amalgamate the widely varying outcomes for house prices across the U.S. during and after the Great Recession. Given that house prices plunged in some places and were fairly stable in others, it would not be appropriate to use the national default tables to construct a risk indicator for any

[46] An alternative definition of subprime loans uses 620 as the credit score threshold. We think this alternative definition is of less interest than the 660 cutoff because the below-620 share of the mortgage market is small – only about 15 percent on average over 1994-2006. That said, if we redo Figure 8 with 620 as the threshold credit score, the results are no more favorable to the subprime-centric view of the housing boom.

particular locality. Such an indicator would require a measure of location-specific house price risk. In this section, we implement a straightforward local risk measure under stress by incorporating hypothetical local house price shocks in the risk indicator.[47]

We define a "shock combined loan-to-value" ratio at origination (SCLTV) as the ratio of the unpaid principal balance at origination (UPB) divided by the "shock house value," which is set equal to the house value at origination (V) multiplied by one minus a projected 3-year house price shock that we associate with a severe stress scenario.

$$SCLTV = \frac{UPB}{V \times (1 - \Delta V^s)} \tag{1}$$

The SCLTV represents the expected combined loan-to-value ratio of a loan three years after the onset of severely stressed conditions; the three-year period allows the price shock to play out over a realistic time horizon. ΔV^S represents the assumed shock to house prices, expressed in absolute value. Consequently, equation (1) shows that the SCLTV is always higher than the stated CLTV at origination.

Our price shock variable is constructed following Smith and Weiher (2012) and Smith et al. (2016), who describe a simple method for constructing a house price path associated with severe economic stress. The concept behind this method is that house prices tend to fall below a long-term trend during a period of severe stress. The trend serves as a reduced-form proxy for economic fundamentals, such as the house price-to-income ratio or the house price-to-rent ratio. Different locations face different degrees of variation around this trend, so both the house price level in relation to the trend and the history of price declines are taken into account when setting the shock value. As noted in the introduction, a version of this approach was implemented in FHFA's revised Enterprise Regulatory Capital Framework, effective in 2021.

[47] We view the house price shock as a stand-in for a broader set of economic shocks during the financial crisis, including shocks to income or unemployment. This interpretation is supported by the strong correlation between changes in house prices and the unemployment rate at the county level during this period. This close association does not always hold, as highlighted by Cherry et al. (2021) for the COVID-19 pandemic, when house prices surged while many households suffered substantial hits to income.

We begin with a real house price series constructed using a nominal series divided by the consumer price index for all urban consumers (CPI). The construction of the SCLTV then proceeds in 4 steps:

1. Fit a trend line with a constant growth rate to the series for the log of real house prices for a given locality using the FHFA annual indices over 1975-2020 described in Bogin, Doerner, and Larson (2019).[48] If the trend slope is negative, set it to 0. Convert the real trend back to nominal terms using the CPI, as all subsequent calculations are in nominal terms.
2. Considering the entire period 1975-2020, find the year in which the house price series was furthest below trend in percentage terms. Let T* denote that year and let L(T*) denote the maximum percentage deviation below trend. L(T*) represents the amount by which the house price index is assumed to drop below trend under severe stress. We set L*(T) to be 5 percent if the unconstrained calculation yields a smaller drop below trend.
3. Calculate the stress loss in house value for each year, $\Delta V^S(T)$, as the percentage loss in house value given two pieces of information: (a) the house price index in year T and (b) the trend house value three years in the future adjusted by the maximum historical amount by which the house price index dropped below trend.[49] For each year, we enforce a minimum stress loss of 5 percent (i.e., $\Delta V^S(T) \geq 0.05$ for all T).
4. For each loan in our dataset originated in year T (T = 1990 , … , 2019), use $\Delta V^S(T)$ to calculate the SCLTV in equation (1) above.

Figure 9 illustrates this procedure for the Phoenix, AZ metropolitan statistical area (MSA). The trend increase in real house prices is estimated to be 0.6 percent per year, which we convert to nominal terms in the figure. The largest deviation below the nominal trend, L(T*), occurs in 2011, when the house price index fell to 31 percent below trend. Using L(T*) = 0.31,

[48] We use all available data to estimate the trend; 2020 was the final year with annual data when we finished the paper.
[49] Written out explicitly, $\Delta V^s(T) = 1 - [(1 - L(T^*)]Trend(T+3)/HPI(T)$. This formula expresses the stress loss as a positive number, as noted in the discussion of equation (1). For T = 2019, calculating the trend three years ahead requires an assumed value for the CPI for 2022. We use the latest available 2022 CPI forecast from the Survey of Professional Forecasters (https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/spf-q4-2021).

the maximum stress loss is nearly 53 percent in 2006 (i.e., $\Delta V^S(T=2006) \approx 0.53$), which is calculated using the house price index for 2006 and the stressed level for 2009 of 0.69 times the 2009 trend. The estimated stressed loss then moved down through 2012, reflecting the sharp drop in house prices. Since then, house prices have increased much faster than trend, pushing the stressed loss back to an elevated level.

This procedure is repeated for each MSA and each state. For each loan in the database, we calculate the SCLTV using the house price index for its MSA if available. When the MSA index is unavailable – as in small cities where city-specific indices are not available or in rural areas that are not part of a MSA – the maximum stress loss is constructed using the state index.

After calculating loan-level SCLTVs, we implement the procedure outlined in Section 3 to construct the stressed default rate for each loan based on SCLTVs substituted for CLTVs at origination. The buckets for SCLTVs in the default tables are (in percent) 60 or less, 61-80, 81-100, 101-115, 116-130, 131-145, 146-160, and 160+. We do not change any other risk factor in the new default tables, which continue to reflect the performance of loans originated in 2006-2007.[50] We use these new tables to compute the stressed default rate for each loan in each year and then average the loan-level stressed default rates year by year to construct national, state, and MSA stressed default series.[51] We refer to the resulting series as "shock stressed default rates."

Before presenting results, it may be useful to compare the shock method of computing stressed defaults with our baseline method. The baseline method assigns the same stressed default rate to all loans that have the same reported characteristics at origination. By doing so, it effectively embeds the average house price shock experienced by the 2006-2007 cohort, with no adjustment for the specifics of the house prices prevailing at origination in a given loan's

---

[50] In principle, one might also consider adjusting DTIs to account for a local income shock that is correlated with the local house price shock. However, this would complicate the analysis while having little effect on the results. The effect would be small because income at the MSA or state level fell much less than house prices, so the resulting adjustment to DTIs would be too small to affect default risk by much. Even though income shocks clearly matter at the household level, aggregated income data are a poor proxy for these shocks, a point stressed by Foote and Willen (2018) in their review of mortgage default research.

[51] We also estimate revised adjustment factors for 15-20 year, 40-year, second-home, and investor loans using the logit methodology described in Section 3 but with the SCLTV buckets replacing the reported CLTV buckets in the logit model.

locality. The shock method, in contrast, adjusts the baseline method to account for the level of house prices in the loan's local geography relative to trend at the time of origination. For example, a mortgage with a reported CLTV of 80 percent at origination would have a higher shock stressed default rate when house prices are well above trend than when they are well below.

6.2 Results

The average SCLTV for the nation as a whole is shown in the left panel of Figure 10.[52] By construction, the SCLTV is higher than the unadjusted CLTV for every loan because the value of the home is lower than the value used to construct the CLTV. The extent to which the average SCLTV exceeds the CLTV depends on the location-specific house price shocks averaged over the country as a whole. The figure shows that the average SCLTV varies much more over time than the unadjusted average CLTV. In 2006, the national average SCLTV exceeded 120 percent, indicating the enormous collateral risk due to above-trend house prices.

The right panel shows that the stressed default rate calculated using the SCLTV is more pro-cyclical than when using the unadjusted CLTV. This can be seen from the larger rise in the stressed default rate over 1994-2006 with the SCLTV than the CLTV, the greater decline over 2006-2011, and the somewhat steeper increase over 2011-2019. The series based on the SCLTV is more pro-cyclical because of the changes in house price risk over the cycle. The stressed default rates converge in 2006-2007 by construction, as both sets of default tables reflect the observed performance of loans originated in those years.

The narrowing gap between the two series since 2011 highlights one of the insights of this extension of the standard measure. Namely, rising house prices can boost default risk under stress even when there are only modest overall changes in borrower and loan characteristics.

The second main use of SCLTV-based default risk estimates for individual loans is to construct measures of local risk. We calculate stressed default rates for all 50 states, the District of Columbia, and all 383 MSAs by aggregating the loans within these geographies. These

[52] This figure and the other figures in this section present data without the adjustment for changes in refi volume. The salient features of the figures are not affected by the adjustment.

indicators show wide variation in stressed default risk both across areas and over time, reflecting region-specific lending and house price patterns.

To illustrate this variation, Table 5 shows the shock stressed default rate and average shock CLTV for 2006-2007 originations in the states with the highest and lowest levels of risk. The “Sand States” (Arizona, California, Florida, and Nevada) had by far the highest risk, with Maryland the closest state behind them. The very high default risk in the Sand States at the onset of the Great Recession owed to the combination of risky loan characteristics and the potential for severe house price declines – most notably in Nevada, which had an average shock CLTV of 200 percent. The states with the lowest risk had shock stressed default rates that were only a fraction of those in the Sand States, in part reflecting average shock CLTVs that were much lower.

7. COVID-19 Update

The COVID-19 pandemic began in early 2020 and had not relented by the end of 2021 when we completed this research. This section provides a partial update of our results through the third quarter of 2021. To obtain the most timely results possible, we limit the analysis to Enterprise, FHA, and VA loans, as data for these market segments are available with the shortest lag. Because Enterprise, FHA, and VA loans have accounted for roughly three-quarters of all first-lien originations in recent years, there is only a modest loss of market coverage by restricting attention to government-guaranteed loans.

Table 6 summarizes the results of this update. The clear conclusion is that the riskiness of mortgage originations fell during the pandemic. As shown, the baseline stressed default rate dropped from 2019 to 2020 and edged down a bit more during the first three quarters of 2021. The decline in risk reflected a rise in the average credit score and a drop in both the average DTI and average CLTV.

Early in the pandemic, lenders tightened standards in the face of financial stress and uncertainty about future economic activity. However, the influence of this factor on origination risk waned over time. According to the Federal Reserve Board’s Senior Loan Officer Opinion Survey on Bank Lending Practices (https://www.federalreserve.gov/data/sloos.htm), the tightening of standards ended in the third quarter of 2020 and was at least partially reversed over subsequent quarters.[53] This reversal makes sense: With immediate support and various actions from fiscal and monetary policy to help stabilize markets, the COVID-19 recession lasted only two months (see https://www.nber.org/research/business-cycle-dating), and the economy rebounded quickly from its trough.

Two other factors have had a longer-lasting effect on the decline in origination risk. First, the steep drop in mortgage rates spurred a wave of refinancing. As discussed above, the stressed default rate declines when refinancing volume is high as lower-risk borrowers flood the market to take advantage of attractive rates. The second factor relates to the hot housing market that emerged in mid-2020 in response to low mortgage rates, the rapid economic recovery, and the desire by households for more space.[54] In this market, lower-income homebuyers have been at a disadvantage when competing against buyers with stronger financial profiles; they also have become less able to afford homes given the rapid price increases that began in the second half of 2020. Accordingly, the composition of home purchase loans has shifted toward lower-risk borrowers, as shown in the table by the rising Enterprise share of purchase loans at the expense of the FHA and VA.

A final point is that the steep increase in house prices has raised the risk of a price reversal in a stress event. This heightened risk operates through the shock CLTV in our framework. In 2020, the average shock CLTV moved about in sync with the average reported CLTV, as house prices did not accelerate much when measured on the annual-average basis used in our framework. However, the data in hand for 2021 through Q3 imply an extremely large rise in house prices for the year as a whole. Consequently, when the full-year data become available, they likely will show that the average shock CLTV for new originations moved up substantially in 2021.

---

[53] Taken literally, the survey results imply that the reversal was not complete and that lending standards remained tighter in 2021:Q3 than they were before the pandemic. However, the survey historically has shown a bias toward reporting tighter mortgage lending standards, so the net change in standards since the start of the pandemic is unclear.

[54] See Harvard Joint Center for Housing Studies (2021) for details.

8. Conclusion

Understanding the evolution of risk in the mortgage market before the financial crisis and after requires a long historical account built from comprehensive and accurate data. Until now, that historical record did not exist, and this paper takes a major step toward filling the gap. We bring together several sources of data, including the entire Enterprise book, to cover essentially the entire market for home mortgage loans from 1990 to 2019. We track important loan characteristics and compute a summary measure of risk under stressed conditions.

We use the data to reach a number of conclusions. First, we show that loan risk had already risen by 2000, calling into question the common view that the early 2000s represented a period of normal lending conditions. Second, we provide new evidence that credit supply expanded in the early and mid-2000s by documenting a compression of mortgage rate spreads between the riskiest and least risky loans in the PLS market and in lenders' portfolios. Third, the rise in risk before the financial crisis was similar for borrowers with low credit scores and those with higher scores. This fact, combined with little change in the low-score share of loans from 2000 through the peak of the housing boom, undercuts explanations of the crisis that focus on low-score borrowers. Fourth, plain-vanilla factors such as higher DTIs and CLTVs played a key role in driving up risk before the financial crisis, suggesting the typical focus on risky product features misses an important part of the story. Finally, tighter underwriting standards during and after the crisis sharply reduced mortgage risk. Although the level of risk based on loan and borrower characteristics remains relatively low, the substantial house price appreciation over the past decade has created the potential for a sizable price drop in a stress event.

References

Adelino, Manuel, William B. McCartney, and Antoinette Schoar. 2020. "The Role of Government and Private Institutions in Credit Cycles in the U.S. Mortgage Market." NBER Working Paper 27499.

Adelino, Manuel, W. Scott Frame, and Kristopher Gerardi. 2017. "The Effect of Large Investors on Asset Quality: Evidence from Subprime Mortgage Securities." *Journal of Monetary Economics* 87: 34-51.

Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2016. "Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class." *Review of Financial Studies* 29(7): 1635-70.

Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2017. "Dynamics of Housing Debt in the Recent Boom and Bust." *NBER Macroeconomics Annual* 32: 261–311.

Albanesi, Stefania, Giacomo De Giorgi, and Jaromir Nosal. 2017. "Credit Growth and the Financial Crisis: A New Narrative." NBER Working Paper 23740.

Amromin, Gene, Jennifer Huang, Clemens Sialm, and Edward Zhong. 2018. "Complex Mortgages." Forthcoming in *Review of Finance*.

Anenberg, Elliot, Aurel Hizmo, Edward Kung, and Raven Molloy. 2019. "Measuring Mortgage Credit Availability: A Frontier Estimation Approach." *Journal of Applied Econometrics* 34(6): 865-82

Avery, Robert B., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner. 1996. "Credit Risk, Credit Scoring, and the Performance of Home Mortgages." *Federal Reserve Bulletin* 82 (July): 621-48.

Avery, Robert B., Kenneth P. Brevoort, Tim Critchfield, Ian H. Keith, Ismail E. Mohamed, Forrest W. Pafenberg, Saty Patrabansh, Jay D. Schultz, and Claudia E. Wood. 2017. "National Mortgage Database Technical Report 1.2." NMDB Technical Report Series, October 30. https://www.fhfa.gov/PolicyProgramsResearch/Programs/Documents/NMDB-Technical-Report_1.2_10302017.pdf

Bogin, Alexander N., William M. Doerner, and William D. Larson. 2019. "Local House Price Dynamics: New Indices and Stylized Facts." *Real Estate Economics* 47(2): 365-98.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45(1): 5-32.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees*. Chapman and Hall/CRC: Boca Raton, FL.

Bubb, Ryan, and Alex Kaufman. 2014. "Securitization and Moral Hazard: Evidence from Credit Score Cutoff Rules," *Journal of Monetary Economics* 63: 1-18.

Cherry, Susan, Erica Jiang, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2021. "Government and Private Household Debt Relief during COVID-19", *Brookings Papers on Economic Activity*, Conference Draft, September 9. https://www.brookings.edu/wp-content/uploads/2021/09/Government-and-Private-Household-Debt-Relief-_Conf-Draft2.pdf

Conklin, James, W. Scott Frame, Kristopher Gerardi, and Haoyang Liu. 2018. "Villains or Scapegoats? The Role of Subprime Borrowers in Driving the Housing Boom." Federal Reserve Bank of Atlanta Working Paper 2018-10, August. Forthcoming in *Journal of Financial Intermediation*.

Consumer Financial Projection Bureau. 2014. "Ability-to-Repay and Qualified Mortgage Rule: Small Entity Compliance Guide." November 3. https://files.consumerfinance.gov/f/201411_cfpb_atr-qm_small-entity-compliance-guide.pdf

CoreLogic. 2017. "Housing Credit Index: First Quarter 2017". CoreLogic Insights Blog, June 20. https://www.businesswire.com/news/home/20170620005466/en/CoreLogic-Reports-Mortgage-Credit-Risk-Edges-Slightly

Courchane, Marsha J., Leonard C. Kiefer, and Peter M. Zorn. 2015. "A Tale of Two Tensions: Balancing Access to Credit and Credit Risk in Mortgage Underwriting." *Real Estate Economics* 43(4): 993-1034.

DeFusco, Anthony A., Stephanie Johnson, and John Mondragon. 2020. "Regulating Household Leverage." *Review of Economic Studies* 87(2): 914-58.

Demyanyk, Yuliya, and Otto Van Hemert. 2011. "Understanding the Subprime Mortgage Crisis." *Review of Financial Studies* 24(6): 1848–80.

Dunsky, Robert M., Austin Kelly, and Ken Lam. 2013. **"**Impacts of Down Payment Underwriting Standards on Loan Performance – Evidence from the GSEs and FHA Portfolios." FHFA Working Paper 13-3, December. https://www.fhfa.gov/policyprogramsresearch/research/paperdocuments/2013-12_workingpaper_13-3-508.pdf.

Ferreira, Fernando, and Joseph Gyourko. 2015. "A New Look at the Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012." NBER Working Paper 21261.

Firestone, Simon, Robert Van Order, and Peter Zorn. 2007. "The Performance of Low-Income and Minority Mortgages." *Real Estate Economics* 35(4): 479-504.

Foote, Christopher, Kristopher S. Gerardi, Lorenz Goette, and Paul S. Willen. 2010. "Reducing Foreclosures: No Easy Answers." In *NBER Macroeconomics Annual 2009*, edited by Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, University of Chicago Press, pp. 89-138.

Foote, Christopher L., Lara Loewenstein, and Paul S. Willen. 2019. "Technological Innovation in Mortgage Underwriting and the Growth in Credit, 1985-2015." Federal Reserve Bank of Boston Research Department Working Paper No. 19-11. https://www.bostonfed.org/publications/research-department-working-paper/2019/technological-innovation-in-mortgage-underwriting-and-the-growth-in-credit-1985-2015.aspx

Foote, Christopher L., and Paul S. Willen. 2018. "Mortgage-Default Research and the Recent Foreclosure Crisis." *Annual Review of Financial Economics* 10: 59-100.

Fout, Hamilton, Grace Li, Mark Palim, and Ying Pan. 2018. "Credit Risk of Low Income Mortgages." Forthcoming in *Regional Science and Urban Economics*.

Fuster, Andreas, Laurie Goodman, David Lucca, Laurel Madar, Linsey Molloy, and Paul Willen. 2013. "The Rising Gap Between Primary and Secondary Mortgage Rates." Federal Reserve Bank of New York, *Economic Policy Review*, December: 17-39. https://www.newyorkfed.org/medialibrary/media/research/epr/2013/1113fust.pdf

Fuster, Andreas, Stephanie H. Lo, and Paul S. Willen. 2017. "The Time-Varying Price of Financial Intermediation in the Mortgage Market." Federal Reserve Bank of New York, Staff Report No. 805, August. https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr805.pdf?la=en

Garmaise, Mark J. 2018. "Alternative Mortgage Contracts and Affordability – Overview." Forthcoming in *Regional Science and Urban Economics*.

Goldman Sachs. 2014. "Are Mortgage Lending Standards Easing?" Goldman Sachs Credit Strategy Research, July 11.

Goodman, Laurie, Ellen Seidman, and Jun Zhu. 2014. "VA Loans Outperform FHA Loans. Why? And What Can We Learn?" Urban Institute working paper. https://www.urban.org/sites/default/files/publication/22776/413182-VA-Loans-Outperform-FHA-Loans-Why-And-What-Can-We-Learn-.PDF

Harvard Joint Center for Housing Studies. 2021. The State of the Nation's Housing 2021. https://www.jchs.harvard.edu/state-nations-housing-2021

Haughwout, Andrew, Richard Peach, and Joseph Tracy. 2008. "Juvenile Delinquent Mortgages: Bad Credit or Bad Economy?" *Journal of Urban Economics* 64(2): 246-57.

Herzog, John P., and James S. Earley. 1970. *Home Mortgage Delinquency and Foreclosure*. National Bureau of Economic Research General Series No. 91. New York: Columbia University Press. https://papers.nber.org/books/herz70-1.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti. 2021. "The Mortgage Rate Conundrum." Working paper, January. https://faculty.wcas.northwestern.edu/~gep575/mr3-8.pdf. Forthcoming in *Journal of Political Economy*.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. 2010. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." *Quarterly Journal of Economics* 125(1): 307–362.

Keys, Benjamin J., Amit Seru, and Vikrant Vig. 2012. "Lender Screening and the Role of Securitization: Evidence from Prime and Subprime Mortgage Markets." *The Review of Financial Studies* 25(7): 2071-2108.

Levitin, Adam J., and Susan M. Wachter. 2020. *The Great American Housing Bubble: What Went Wrong and How We Can Protect Ourselves in the Future*. Harvard University Press.

Li, Wei, and Laurie Goodman. 2014. "Measuring Mortgage Credit Availability Using Ex-Ante Probability of Default." Urban Institute working paper, November. https://www.urban.org/sites/default/files/publication/33581/2000018-Measuring-Mortgage-Credit-Availability-Using-Ex-Ante-Probability-of-Default.pdf

Little, Roderick J. A. 1988. "Missing-Data Adjustments in Large Surveys." *Journal of Business Economics and Statistics* 6(3): 287-96.

Mahoney, Peter E., and Peter M. Zorn. 1997. "The Promise of Automated Underwriting: Freddie Mac's Loan Prospector." *Housing Finance International* 11(4): 13-21.

Mayer Christopher, Karen Pence, and Shane M. Sherlund. 2009. "The Rise in Mortgage Defaults." *Journal of Economic Perspectives* 23(1): 27-50.

Mian, Atif, and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *Quarterly Journal of Economics* 124(4): 1449–96.

Mian, Atif, and Amir Sufi. 2014. *House of Debt: How They (and You) Caused the Great Recession, and How We Can Prevent It from Happening Again*. University of Chicago Press.

Mian, Atif, and Amir Sufi. 2017a. "Household Debt and Defaults from 2000 to 2010: The Credit Supply View." In *Evidence and Innovation in Housing Law and Policy*, edited by Lee Anne Fennell and Benjamin J. Keys, 257–88. Cambridge University Press.

Mian, Atif, and Amir Sufi. 2017b. "Fraudulent Income Overstatement on Mortgage Applications during the Credit Expansion of 2002 to 2005." *Review of Financial Studies* 30(6): 1832–64.

Mian, Atif, and Amir Sufi. 2018. "Finance and Business Cycles: The Credit-Driven Household Demand Channel." *Journal of Economic Perspectives* 32(3): 31-58.

Pacelle, Mitchell. 1991. "Quick Home Loans Have Quickly Become Another Banking Mess." Wall Street Journal. July 5.

Palmer, Christopher. 2015. "Why Did So Many Subprime Borrowers Default During the Crisis: Loose Credit or Plummeting Prices?" Working Paper, September 24. https://ssrn.com/abstract=2665762.

Rubin, Donald B. 1976. "Inference and Missing Data." *Biometrika* 63(3): 581-92.

Quercia, Roberto G., and Michael A. Stegman. 1992. "Residential Mortgage Default: A Review of the Literature." *Journal of Housing Research* 3(2): 341-79.

Sharpe, Steve A., and Shane M. Sherlund. 2016. "Crowding Out Effects of Refinancing on New Purchase Mortgages." *Review of Industrial Organization* 48(2): 209-239.

Sichelman, Lew. 1990. "Fannie Mae Sets 30% Down on Low-doc Loan." Chicago Tribune. July 14. https://www.chicagotribune.com/news/ct-xpm-1990-07-14-9002270602-story.html.

Smith, Scott, Debra Fuller, Alex Bogin, Nataliya Polkovnichenko, and Jesse Weiher. 2016. "Countercyclical Capital Regime Revisited: Tests of Robustness." *Journal of Economics and Business* 84: 50-78.

Smith, Scott, and Jesse Weiher. 2012. "Countercyclical Capital Regime: A Proposed Design and Empirical Evaluation." Federal Housing Finance Agency Working Paper 12-2, April. https://www.fhfa.gov/PolicyProgramsResearch/Research/Pages/Working-Paper-12-2.aspx

Stock, James H., and Mark W. Watson. 2015. *Introduction to Econometrics*, 3rd edition. Pearson Education Inc.

Urban Institute. 2018. Housing Credit Availability Index. https://www.urban.org/policy-centers/housing-finance-policy-center/projects/housing-credit-availability-index

Table 1: Share of Home Purchase and Refinance Loans with Imputed Risk Factors, 1990-2019 (percent)

Risk Factor	**Enterprise**	**Portfolio**	**PLS**	**FHA/VA**
DTI	9	65	50	59
Credit score	12	36	9	26
Documentation status	2	40	2	7
Amortization status	0	28	0	0
Occupancy status	0	13	0	0
Type of refinance loan	0	24	0	16
CLTV	0	0	0	2

Note: The shares pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. The imputed share for the type of refinance loan equals the number of refinance loans requiring such an imputation divided by the total number of refinance loans; the numerator and denominator for all risk factors include both home purchase and refinance loans. *Source*: Authors' calculations using data from Black Knight, Inc., CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Table 2: Default Tables

Fixed-rate mortgages	ARMs
Full doc / Standard amortization	Full doc / Standard amortization
Full doc / Non-standard amortization	Full doc / Non-standard amortization
Low or no doc / Standard amortization	Low or no doc / Standard amortization
Low or no doc / Non-standard amortization	Low or no doc / Non-standard amortization

Note: Non-standard amortization includes loans with interest-only payments, negative amortization, or a balloon payment at the end of the stated term.

.

Table 3: Stressed Default Rates for Enterprise 30-year Fixed-Rate Home Purchase Loans with Full Documentation and Standard Amortization (percent)

Credit scores of 720-769

DTI (%)	CLTV (%) 1-60	61-70	71-75	76-80	81-85	86-90	91-95	≥ 96
1-33	1.1	2.3	3.4	4.2	3.9	6.2	7.5	10.9
34-38	1.7	3.6	4.4	6.3	6.0	8.4	9.9	13.1
39-43	1.9	4.4	5.6	7.4	7.2	10.4	12.0	15.8
44-50	2.1	4.5	6.4	8.5	8.6	11.6	13.7	18.2
≥ 51	2.2	4.9	6.8	8.8	9.8	13.0	16.4	25.0

Credit scores of 660-689

DTI (%)	CLTV (%) 1-60	61-70	71-75	76-80	81-85	86-90	91-95	≥ 96
1-33	3.4	6.3	9.0	8.9	8.3	13.5	15.1	23.5
34-38	4.2	7.9	9.4	13.2	13.0	16.4	18.4	27.5
39-43	5.1	10.4	12.2	14.1	15.3	19.6	21.5	30.9
44-50	5.8	10.4	13.6	16.0	17.4	21.9	24.1	33.4
≥ 51	4.9	11.7	15.0	17.0	22.8	24.5	29.9	41.1

Note: The stressed default rates pertain to first-lien Enterprise home purchase mortgage loans secured by 1-4 unit properties. The default rates are calculated from the performance of Enterprise loans originated in 2006-2007 and reflect unweighted loan-level data. Default is defined as a loan ever being 180 days delinquent or terminating with less than full repayment. Shading: green for stressed default rates of 7% or less, orange for 7.01% to 14%, and red for more than 14%. *Source*: Authors' calculations using FHFA data.

Table 4: Influence of Risky Product Features on the Stressed Default Rate, All Loans

	Stressed default rate					
	Level (percent)			Change (ppts.)		
Series	**1994**	**2000**	**2006**	**1994-2000**	**2000-2006**	**1994-2006**
Baseline (Incl. risky product features)	15.6	24.0	36.3	8.4	12.3	20.7
Alternative (Excl. risky product features)	14.4	21.1	26.6	6.7	5.4	12.2

Note: Results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties and are not adjusted for changes in refinance volume. Risky product features include low/no documentation, non-standard amortization, and a loan term greater than 30 years. The stressed default rate excluding risky product features uses the same loans as the rate with risky product features but runs loans with those features through the default tables as full doc, standard amortization, 30-year loans. The changes shown may not equal the differences between the levels due to rounding. *Source*: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Table 5: State-level Shock Stressed Default Rate and Average CLTV, 2006-2007 (percent)

States	Shock Stressed Default Rate	Average Shock CLTV
Highest stressed default rate		
Nevada	61.9	200
Arizona	51.2	156
Florida	50.9	152
California	47.7	141
Maryland	35.1	120
Lowest stressed default rate		
South Dakota	15.5	97
North Dakota	16.8	102
Vermont	17.5	97
Nebraska	18.2	101
Kentucky	18.2	96

Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties and are not adjusted for the effect of changes in refinance volume.
Source: Authors' calculations using data from Black Knight, Inc. CoreLogic, and FHFA.

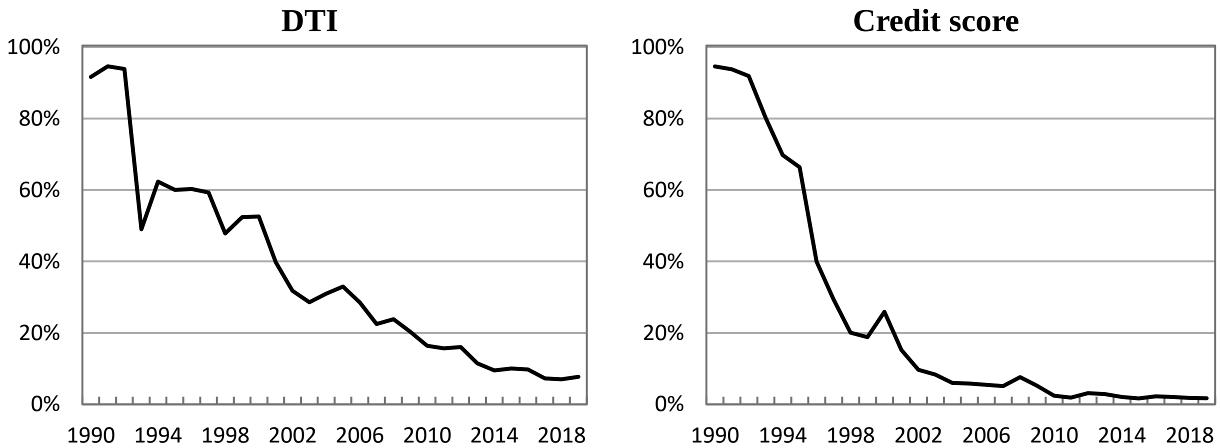
Table 6: COVID-19 and Features of Enterprise, FHA, and VA Loan Originations (all figures are in percent except for average credit score)

	2019	**2020**	**2021***
Baseline stressed default rate	13.3	10.7	10.4
Average credit score	728	744	741
Average DTI	37.6	34.9	34.4
Average CLTV	81.9	76.1	74.1
Enterprise share of home purchase loans	64.6	67.7	70.5
Average shock CLTV	96.6	90.2	NA

*The 2021 data are through Q3.

Note: The results shown are for the aggregate of Enterprise, FHA, and VA mortgage loan originations and are not adjusted for the effect of changes in refinance volume. *Source*: Authors' calculations using data from FHFA and Ginnie Mae data processed by the AEI Housing Center.

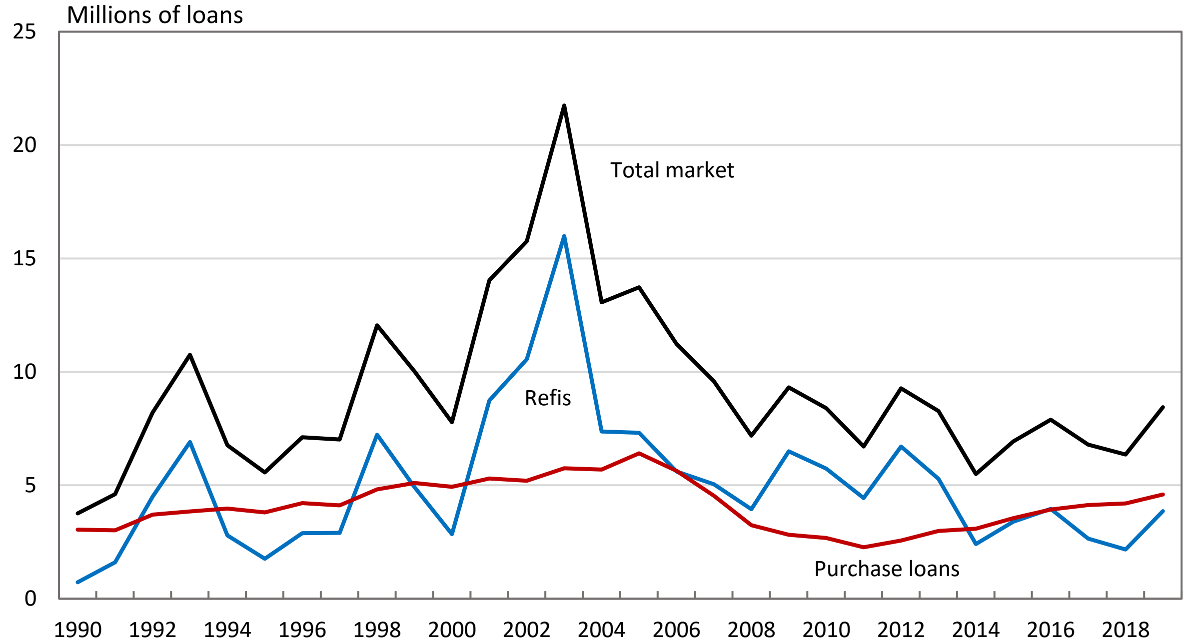
Figure 1: Imputation Share, Home Purchase and Refinance Loans



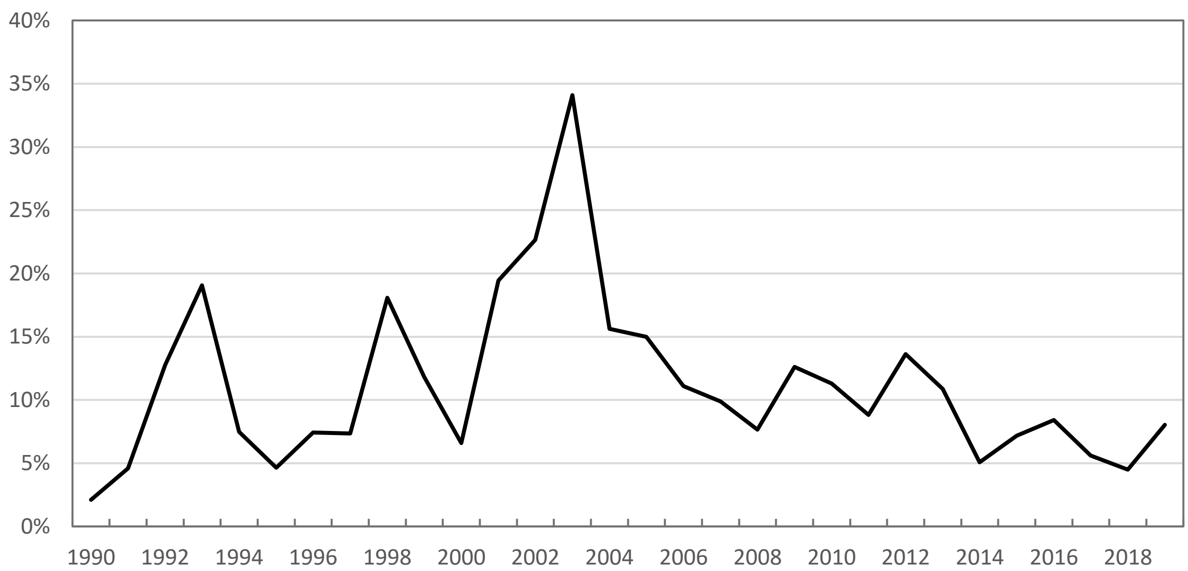
Note: Results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties and are based on weighted loan counts. *Source*: Authors' calculations using data from Black Knight, Inc., CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Figure 2: Home Mortgage Originations

Annual loan origination counts



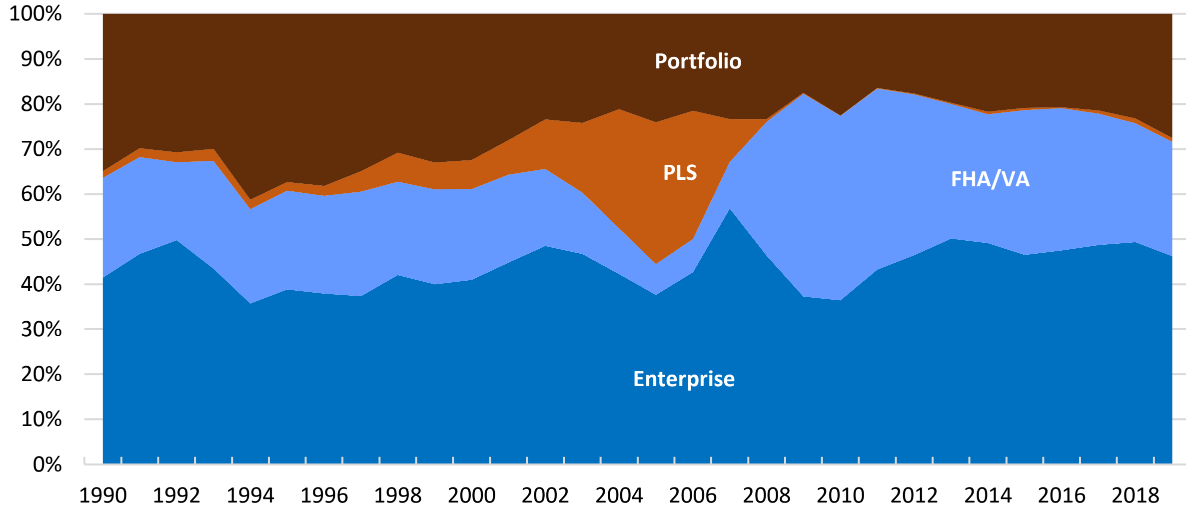
Refinance loan originations as a share of the prior-year stock of mortgages



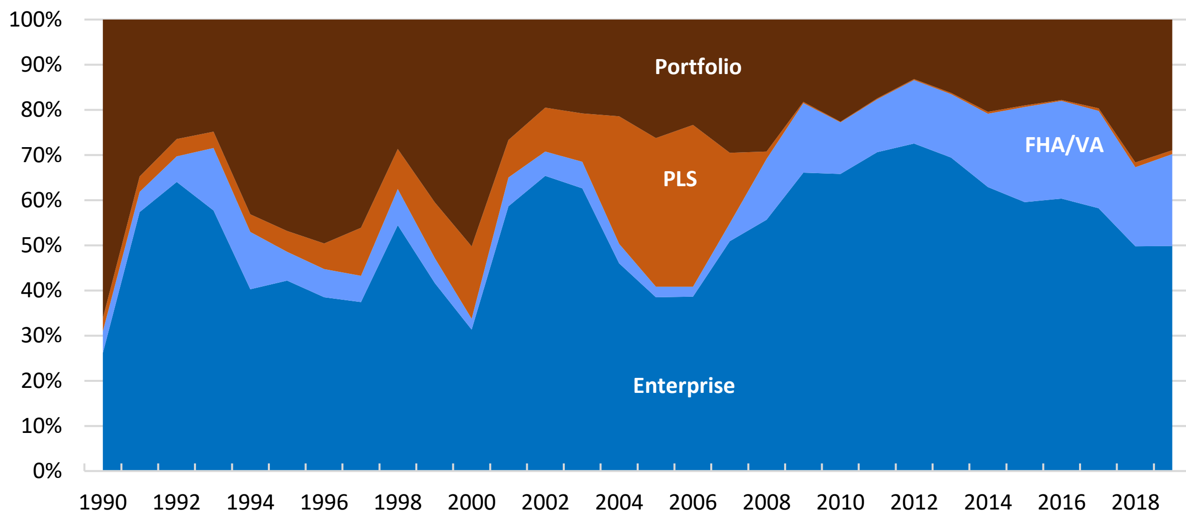
Note: Origination counts refer to home mortgages secured by 1-4 unit properties. The stock of mortgages comes from Census Bureau survey data and counts the number of owner-occupied units with a mortgage. There are small differences in definition between the origination counts and the stock of mortgages. *Source*: Authors' calculations using data from CoreLogic, FHFA, and HMDA (for loan counts) and the Census Bureau's American Housing Survey, Table 3-15, and the American Community Survey, Table S2506 (for the stock of mortgages).

Figure 3: Market Shares

Home purchase loans

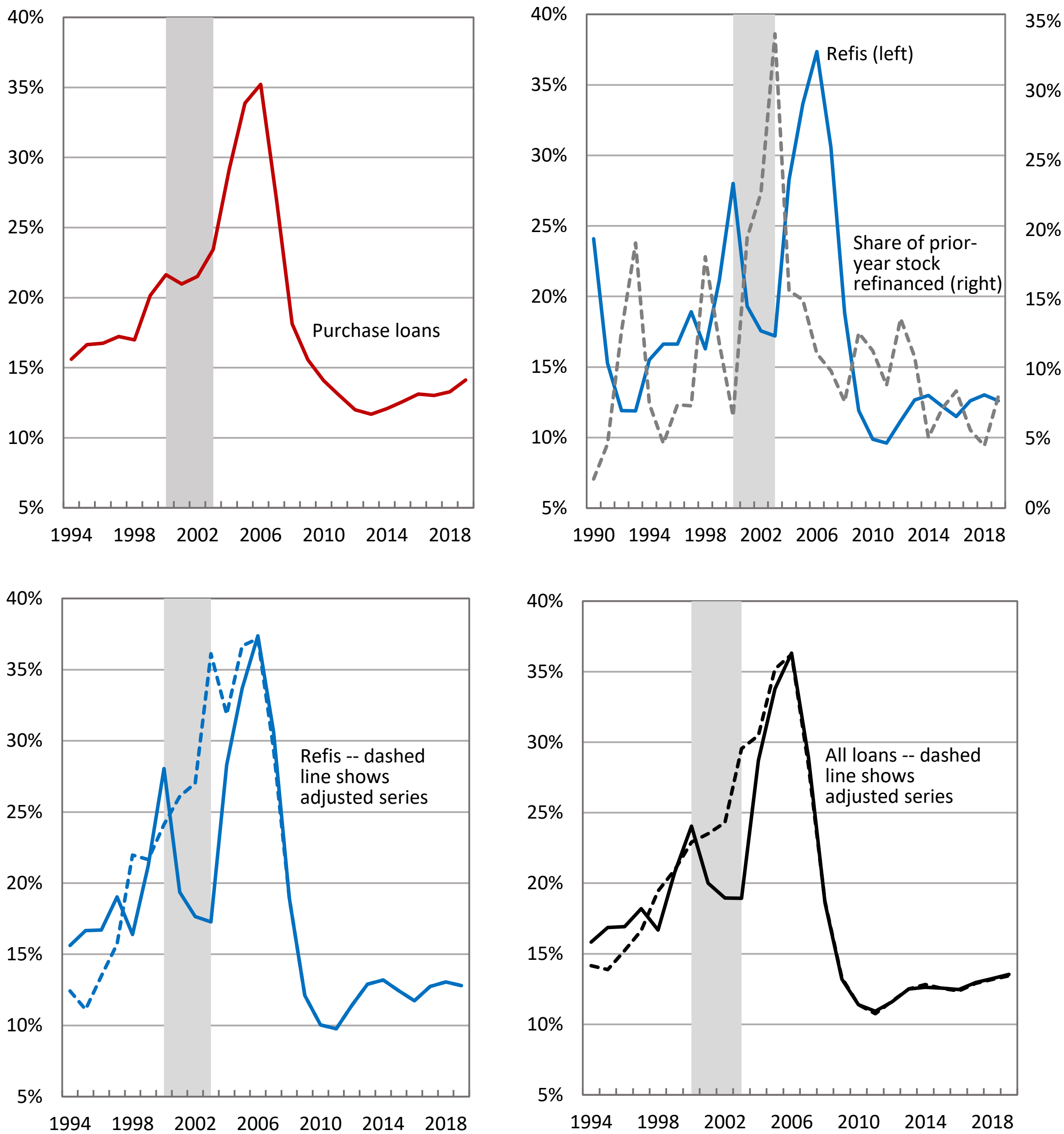


Refinance loans



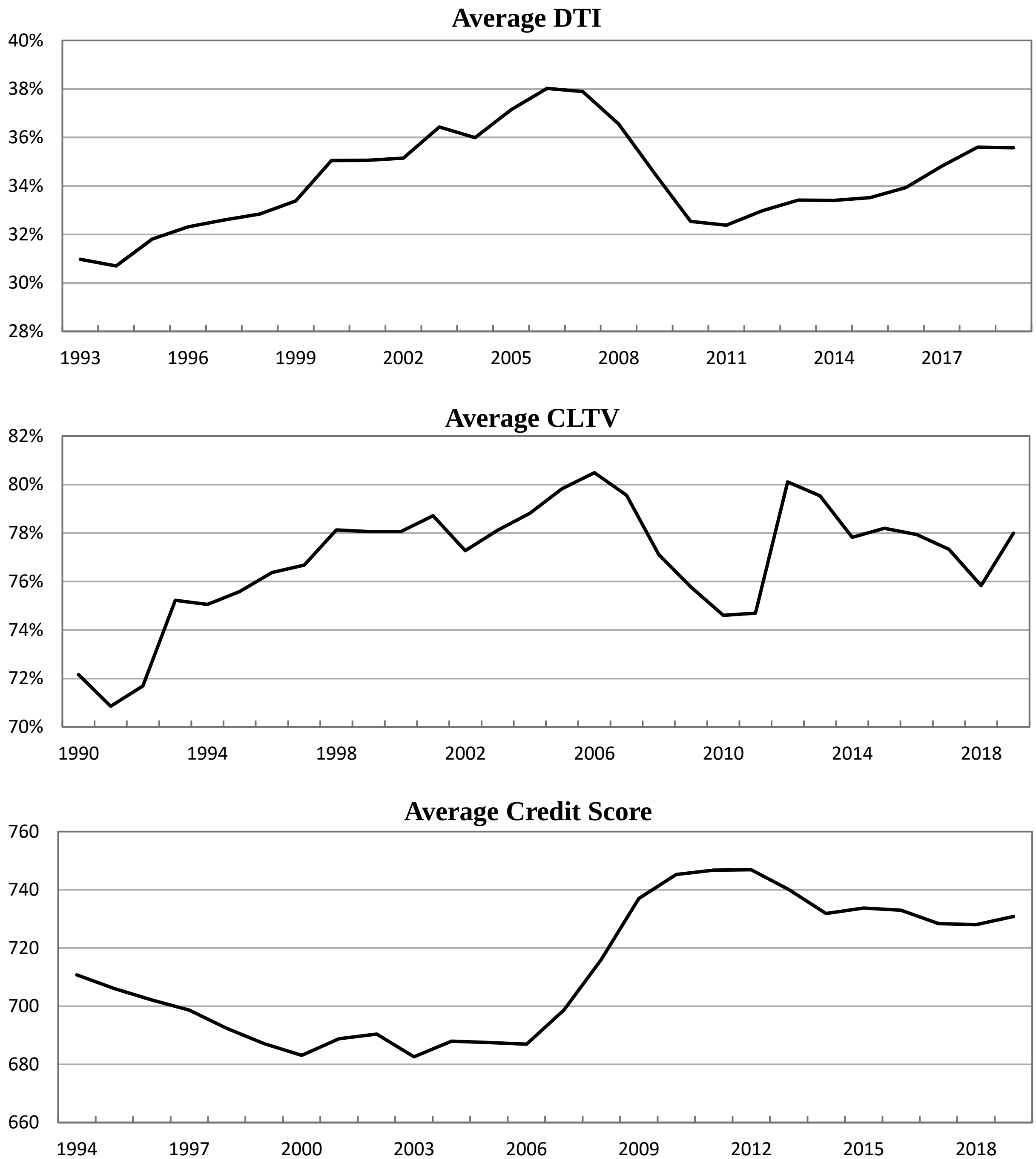
Note: Results pertain to home mortgages secured by 1-4 unit properties and are based on loan counts. The counts for the total market and FHA/VA loans are based on HMDA data and are grossed-up by an estimate of the undercount in HMDA. The Enterprise count comes from FHFA data and covers the universe of such loans. The PLS count from CoreLogic covers nearly the entire market; we gross up the count in our cleaned dataset only to add back the loans removed during the data cleaning process. The portfolio loan count is calculated as the HMDA count for conventional loans (grossed-up by an estimate of the HMDA undercount) minus the Enterprise and PLS counts. Because HMDA only began to collect information on lien status in 2004, the market shares for 1990-2003 include both first liens and subordinate liens. Starting in 2004, the shares include first liens only. *Source*: Authors' calculations using data from CoreLogic, FHFA, and HMDA.

Figure 4: Stressed Default Rates, by Loan Type



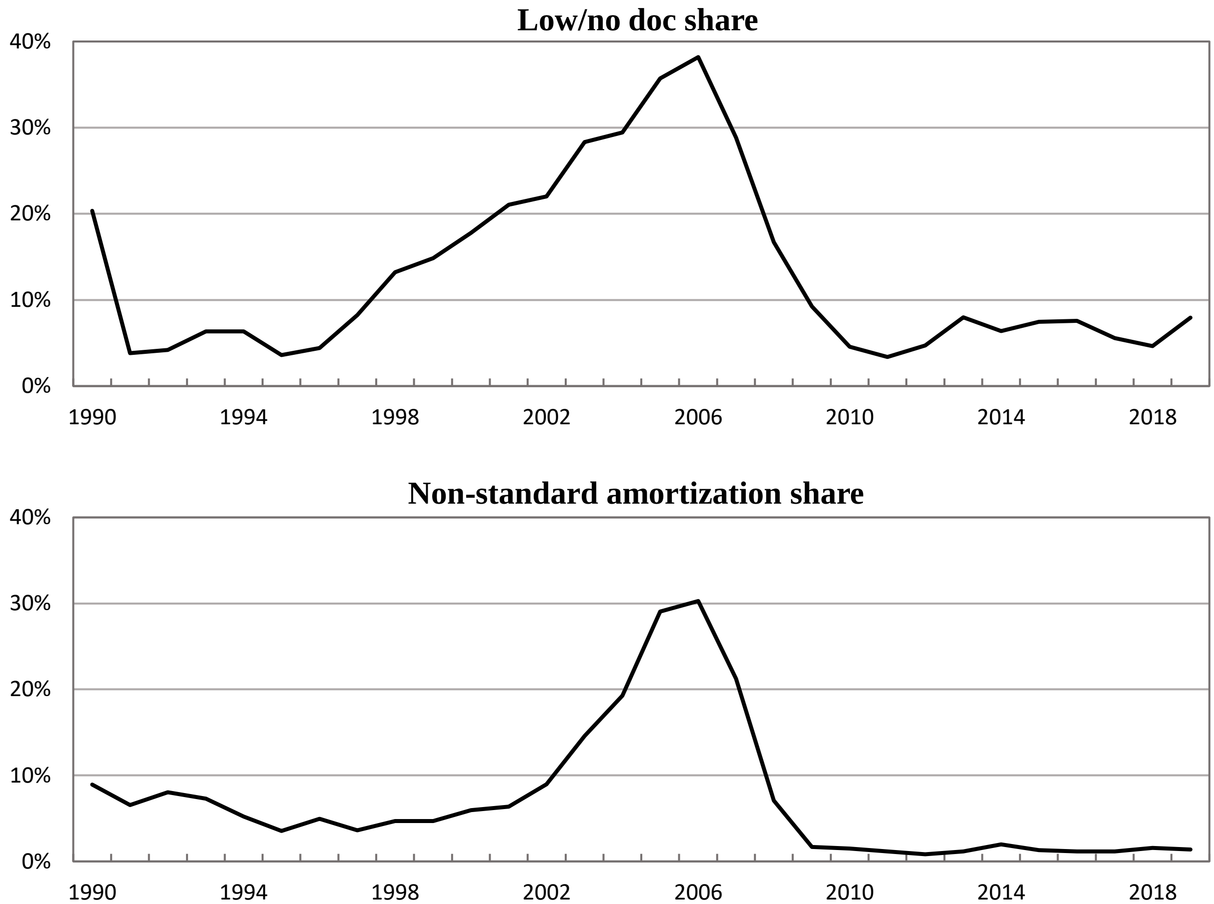
Note: The stressed default rates pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. The adjusted series for refis includes a regression-based adjustment for 1994-2007 to control for changes in refi volume; for 2008-2019, the adjustment is not significant at the 5% level and is omitted. The adjusted series for all loans is a weighted average of the purchase loan series for 1994-2019 and the adjusted refi series for 1994-2007 and the unadjusted series for 2008-2019; the weights in every year reflect the actual purchase loan count and the normalized refi count. Shading is for 2000-2003. *Source*: Authors' calculations using loan-level data from Black Knight, Inc., CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center, and AHS and ACS data for the stock of mortgage loans.

Figure 5: Average DTI, Average CLTV, and Average Credit Score, All Loans



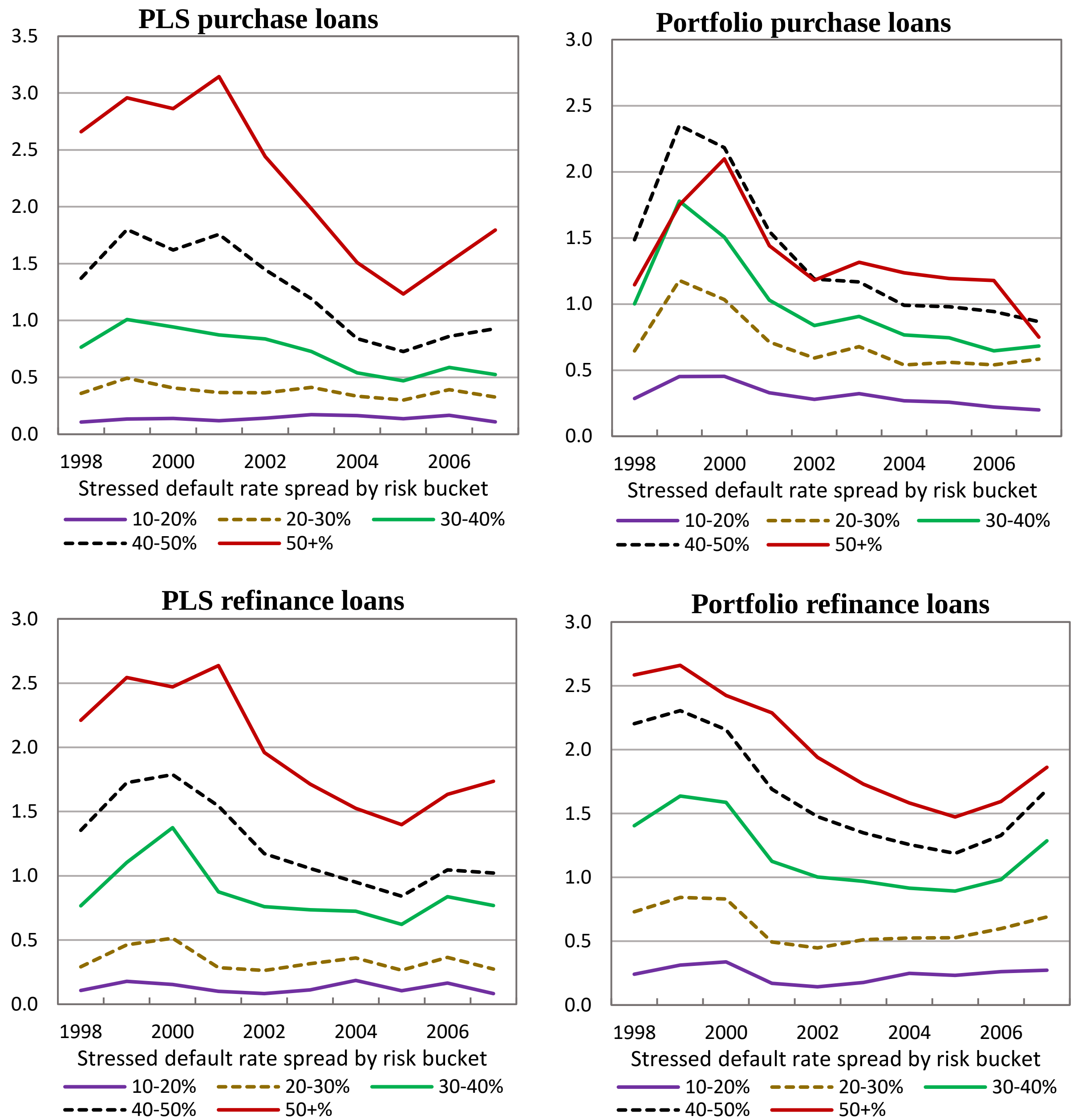
Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. All series include a regression-based adjustment that controls for changes in refi volume; see the text for details.
Source: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing

Figure 6: Share of Loans with Low or No Documentation or Non-standard Amortization, All Loans



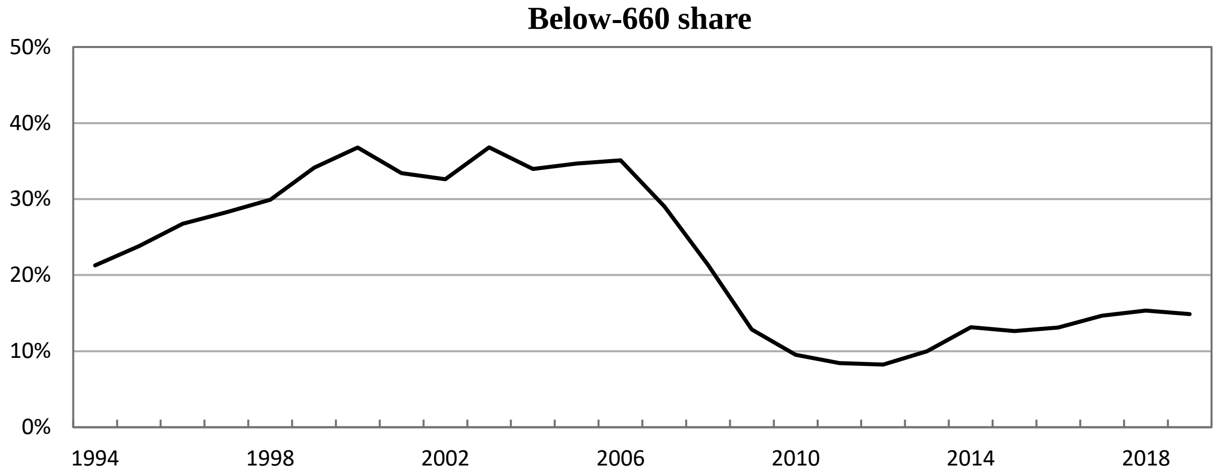
Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. Both series include a regression-based adjustment that controls for changes in refi volume; see the text for details. A loan is classified as having non-standard amortization if it has an interest-only period, negative amortization, and/or a balloon payment. *Source*: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

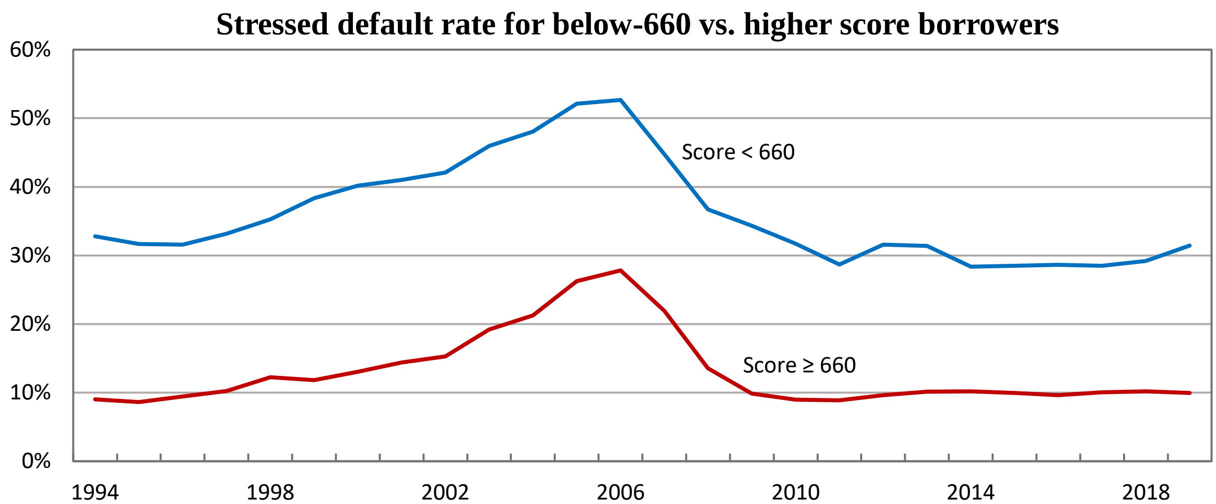
Figure 7: Average Mortgage Rate Spread, by Stressed Default Rate (in percentage points)



Note: The spreads in each panel pertain to first-lien 30-year fixed-rate mortgages secured by 1-4 unit properties and are calculated relative to the average mortgage rate for the lowest-risk group in that market segment (the group with stressed default rates of 0-10 percent). We exclude loans with private mortgage insurance to obtain spreads for lenders and investors that fully bear the credit risk. We also exclude loans with interest rates either above 20 percent or more than 100 basis points below the monthly 30-year FRM rate in Freddie Mac's Primary Mortgage Market Survey. *Source*: Authors' calculations using data from Black Knight, Inc. and CoreLogic.

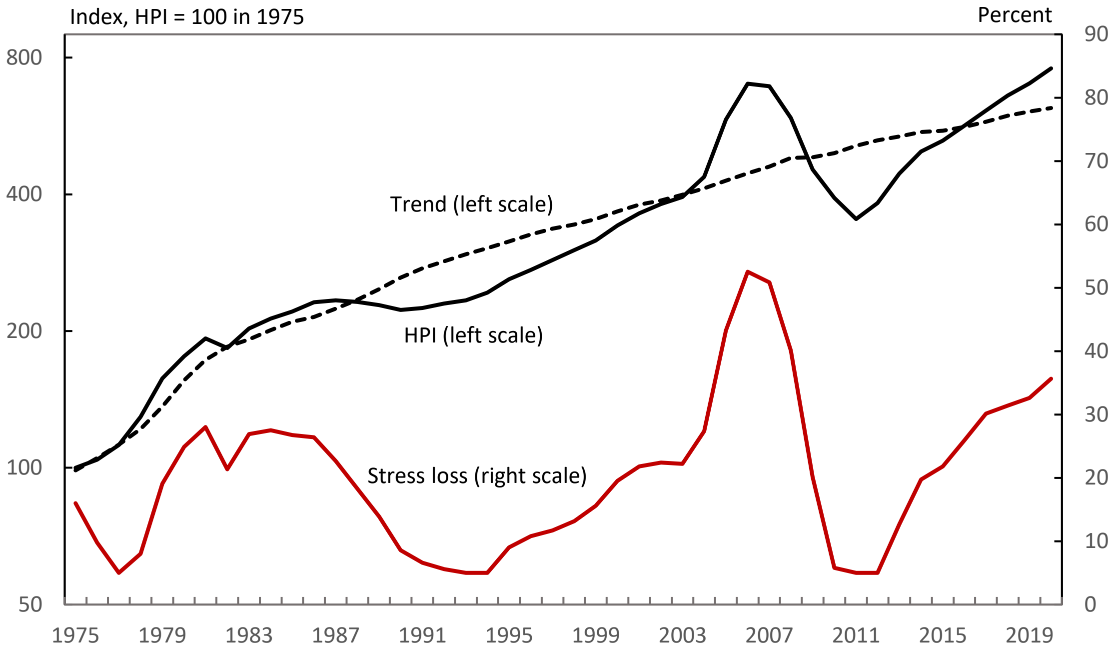
Figure 8: Below-660 Credit-Score Share and Stressed Default Rate for Below-660 vs. Higher-Score Borrowers





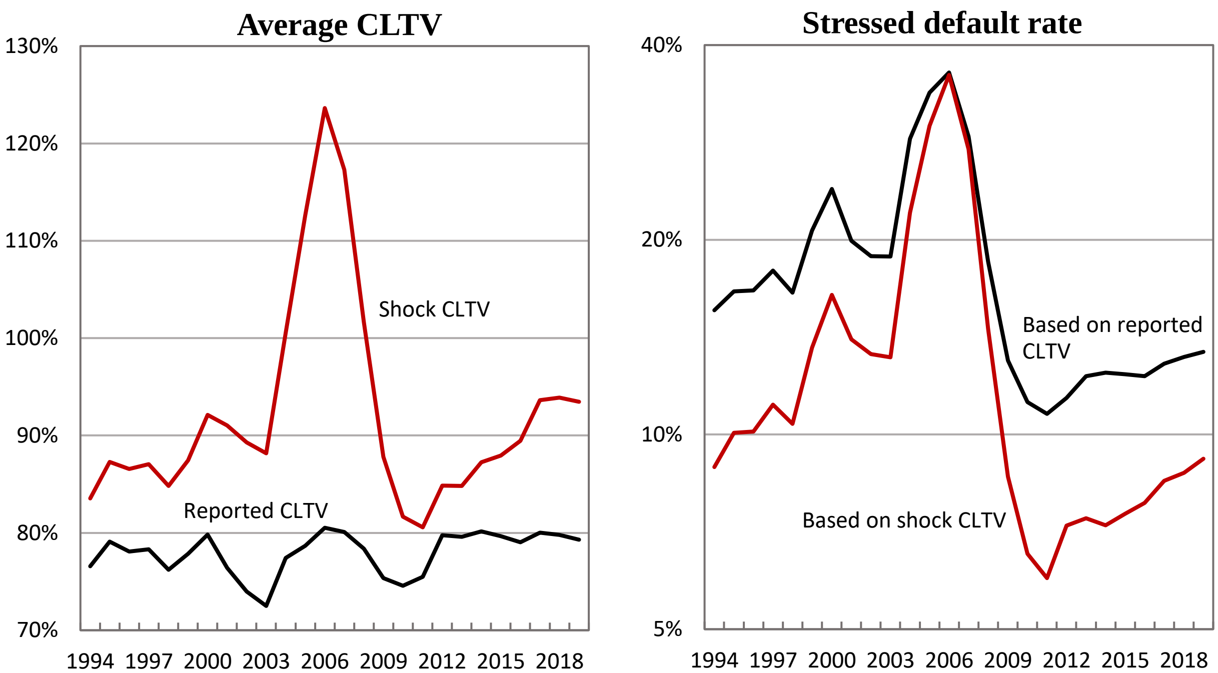
Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. All series include a regression-based adjustment that controls for changes in refi volume; see the text for details. *Source*: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Figure 9: Construction of House Price Shock Series for Phoenix, AZ MSA



Source: Authors' calculations using the FHFA house price index for MSA 38060 (Phoenix-Mesa-Scottsdale) in Bogin, Doerner, and Larson (2019).

Figure 10: Two Versions of the Average CLTV and Stressed Default Rate, All Loans



Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties and are not adjusted for changes in refi volume. The shock CLTV adjusts the reported CLTV to reflect an estimate of the decline in home value that would occur in a severely stressed scenario. The stressed default rate based on the shock CLTV is calculated from default tables that use buckets for the shock CLTV instead of the reported CLTV. *Source*: Authors' calculations using data from Black Knight, Inc., CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Appendix A: Data Considerations

Scope of the dataset

As noted in the main text, our analysis uses data on first-lien mortgages originated in 1990-2019 and secured by 1-4 unit residential properties in the 50 states and the District of Columbia. The dataset includes both home purchase loans and refinance loans for primary owner-occupied properties, second homes, and investment properties. It covers all major segments of the residential mortgage market, including loans guaranteed by the Enterprises, FHA, and VA as well as loans without a government guarantee that are bundled into private-label securities (PLS) or held in the portfolios of banks and other lenders.

Identifying portfolio loans

The portfolio loans used in our analysis are drawn from two servicer datasets, the Loan-Level Market Analytics (LLMA) dataset from CoreLogic and the McDash dataset from Black Knight, Inc. Both datasets identify conventional home loans (i.e., loans held in portfolio by lenders, sold to the Enterprises, or packaged into PLS). Both datasets also include investor codes that identify the loans acquired by the Enterprises, but these codes are not fully populated and there is no field for PLS loans. Thus the information in LLMA and McDash is not sufficient to completely identify portfolio loans.

As a starting point for this identification, we discard the conventional loans that are ever reported as having been acquired by the Enterprises. Among the "never-GSE" loans that remain, we keep those that are ever reported as privately held or that we classify as a jumbo loan (see below for details on this classification). We keep jumbo loans even without any affirmative information that the loan was privately held because jumbos are ineligible for purchase by the Enterprises. This initial set of potential portfolio loans still needs to be purged of PLS loans and duplicate copies of true portfolio loans. Before undertaking that task, we apply a set of data cleaning filters to the potential portfolio loans and to Enterprise, PLS, and FHA/VA loans.

Filtering out loans from the source datasets

Table A.1 summarizes the exclusion of loans from the source datasets. We first remove loans for which any of the following important characteristics are not reported: loan amount, LTV, interest rate, state in which the property is located, and ZIP code. We then remove duplicates of loans that appear more than once in the source data.[55] Duplicates are common for both the potential portfolio loans and FHA/VA loans because both types of loans are drawn from the combination of LLMA and McDash (except for FHA/VA loans originated after 2012, which come from a separate dataset with no duplicates). Many potential portfolio loans and FHA/VA loans appear in both LLMA and McDash, and some appear more than once within each dataset. We also remove loans from the potential portfolio dataset that we believe to be PLS loans (detailed provided below).

After these exclusions, we further clean the dataset by removing loans for which the term, product type (fixed rate versus adjustable rate) or property type is either not reported or reported with incomplete or inconsistent information. We remove the very small share of Enterprise and PLS loans with missing occupancy status.[56] We also remove loans with LTVs below 25 percent, which we suspect are second liens incorrectly reported as first liens, and loans with CLTVs above 135 percent if these loans are not in the Enterprises' Home Affordable Refinance Program (HARP), as such high CLTVs outside HARP likely are erroneous.

A final cleaning step applies to the PLS data. Some PLS loans were securitized long after they were originated, which raises questions about whether the reported loan characteristics represent those at origination or when the loan was securitized.[57] Because our analysis focuses on loan risk at the time of origination, this uncertainty introduces the potential for significant measurement error. Accordingly, we remove PLS loans that were securitized more than one year after origination.

As shown in the row labelled "Total percent excluded," these exclusions taken together remove less than 2 percent of Enterprise loans and about 10 percent of PLS loans. Larger fractions of potential portfolio and FHA/VA loans are excluded, but most of these are duplicate

---

[55] Loans with missing information on loan amount, LTV, interest rate, and property location are excluded before we check for duplicates because those loan characteristics are needed for the matching algorithm that identifies duplicates.

[56] To avoid unnecessary reductions in sample size, we assume that the substantial number of FHA and VA loans with missing occupancy status are owner-occupied given that this is the norm for FHA/VA loans, and for portfolio loans, we impute occupancy status when this information is missing.

[57] These delays, which sometimes exceed two or three years, could arise for various reasons. First, a loan may have become delinquent soon after origination and could not be placed in the market until the payment problem was cured or the loan had been modified in some way. Another source of delay, which was common during the financial crisis, is that market demand dried up for subprime loans and other loans with risky features. In this situation, the deal arranger would end up holding the loans until market conditions became more receptive. Finally, there are loans that a lender originally intended to hold in portfolio but were subsequently sold due to poor performance or because the lending institution failed.

loans or potential portfolio loans we believe to be PLS loans. After netting out these exclusions, the data filters remove about 10 percent of potential portfolio loans and about 5 percent of FHA/VA loans.

Detail on removing duplicate loans from the servicer datasets

To detect duplicates among FHA and VA loans in LLMA and McDash, we attempt to match every FHA loan with every other FHA loan, and do the same for VA loans. If the loans in a given pair do not match on specified characteristics (described below), we declare them not to be duplicates. Because some data are erroneously reported, we allow some minor differences across the match fields. When a pair of loans satisfies the match criteria, we deem that pair to be duplicates and exclude one of the loans from the dataset.

This computationally intensive matching algorithm proceeds in two steps:

1. Declare a non-match for a given pair of loans when any of the following is true: the 3-digit ZIP code is different, the loan type is different, the origination date differs by more than one month, the interest rate differs by more than five basis points, or the rounded original loan amount differs by more than $1,000. These non-matching pairs are deemed not to be duplicates.
2. For all remaining pairs of loans, score their similarity across a given set of fields. The fields used to score the loans are LTV, CLTV, DTI, credit score, product type, occupancy type, loan term, loan performance, origination month, original loan amount, 5-digit ZIP code, and interest rate. We declare pairs that match perfectly to be duplicates, and use a scoring algorithm to declare that additional pairs are duplicates if they match on a large majority of the fields, prioritizing the key match variables used in step 1.

The procedure for removing duplicates from the dataset of potential portfolio loans is very similar to that for FHA and VA loans. We repeat step 1 described above but modify step 2 to include the presence of private mortgage insurance as an additional matching field.

When we remove a duplicate from a pair of loans, we use all available information on loan characteristics across the two loans. For example, if one loan in the pair has a reported DTI but the other does not, we keep the reported DTI. In other cases, duplicate loans can have different values for one or more loan characteristics. The differences generally are small – otherwise the two loans would not have had a high match score in step 2 above. In these cases, we have a set of decision rules for which value to keep. When the value for a particular characteristic has been rounded in one loan but not the other, we keep the more specific value. In addition, we keep the highest CLTV as some servicers may not record all liens against the property. Otherwise, the decision rules involve essentially arbitrary selection.

<u>Detail on removing PLS loans from the potential portfolio dataset</u>

Although the previous steps have removed duplicate loans from the potential portfolio dataset, the dataset still contains PLS loans. Thus, the next step is to remove PLS loans as fully as possible. We do this by using loan-level PLS flags supplied by CoreLogic and Black Knight in auxiliary data files. Because the PLS flags do not appear to identify all PLS loans, we performed a secondary match of the unflagged loans to the separate CoreLogic PLS dataset, using the same algorithm described in the previous section to remove duplicate loans. Loans that match to the PLS dataset are removed from the potential portfolio dataset. This procedure is meant to ensure that we do not double-count PLS loans.

Second liens and measured CLTVs

A possible issue with the CLTVs in our dataset is that they may not fully capture the second liens taken out in connection with home purchases (so-called "piggyback" loans). We examined this question by comparing the prevalence of such loans in our dataset to the estimates in Adelino, McCartney, and Schoar (2020), Bhutta and Keys (2018), and LaCour-Little, Calhoun, and Yu (2011), who show that the use of piggybacks increased during the housing boom and then plummeted during the bust. Although our dataset does not explicitly identify piggyback loans, we deem a piggyback to exist whenever the CLTV at origination exceeds the LTV by at least 5 percentage points.[58] For 2004 and later years, the occurrence of piggybacks in our data closely tracks the other estimates. Thus, the average CLTVs shown in Figure 5 in the paper appear to be accurate from 2004 forward. Before 2004, however, the reporting of piggybacks in our dataset is incomplete. Our rough estimate is that the true CLTV for home purchase loans could be understated by ½ to ¾ percentage point in 2003 and by perhaps twice

[58] This assumption is based on the observation that piggybacks tend to occur at 5 percentage point intervals; our histograms have major spikes at 10, 15, and 20 percentage points, with smaller bumps at 5 and 25 percentage points. Accordingly, we set the minimum threshold at a 5 percentage point difference between the LTV and CLTV to infer the existence of a piggyback loan.

that amount over 1997-2002. Correcting for this understatement would not materially change the contour shown in the figure.[59]

Weighting of loan data

Because our FHA, VA, and portfolio datasets do not constitute the universe of such home mortgage loans, we construct weights so that the data will be representative at the MSA and state level using adjusted counts of originated loans from the Home Mortgage Disclosure Act (HMDA) Loan Application Register. Our national results are then aggregations of the state-level results.

For FHA and VA loans, we partition both our dataset and the HMDA data by year, loan type (FHA purchase, FHA refi, VA purchase, VA refi), loan amount buckets, and geography (state or MSA). Because HMDA does not include all mortgage loans, we increase the HMDA loan counts to approximate the universe of loans.[60] For FHA and VA loans, we use the estimate of the undercount through 1995 in Scheessele (1998). We then assume HMDA coverage increases monotonically from 93 percent in 1996 to 100 percent in 2010 where it remains through 2019, based on the coverage results for FHA loans in Scheessele (1998) and Szymanoski et al. (2011) and the results for FHA and VA loans together in Williams (2015). For each cell in the partition described above, we use the ratio of the adjusted HMDA loan count to the count in our dataset as the weight.[61]

For loans held in lenders' portfolios, we use the same partition as for FHA and VA loans. We then estimate the universe of portfolio loans through a series of steps as the estimated universe count of conventional loans minus the counts of Enterprise and PLS loans. The key component of these steps is estimating the universe count of conventional loans as a gross-up of the conventional loans reported in HMDA, where the gross-up factor is based on the HMDA undercount of Enterprise loans.

[59] In addition to piggyback loans, second liens can be taken out after the purchase date by borrowing against accumulated home equity. We do not track these subsequent second liens. Using data for one of the Enterprises, Leventis (2014) estimated that approximately 5 percent of the first liens originated in 2006-2007 had a subsequent second lien. Thus, the performance of the 2006-2007 loans used to calculate our default tables would reflect the CLTV at origination and, to a modest extent, unobserved subsequent claims against home equity.

[60] For information about the regulatory changes affecting the undercount of mortgages reported in HMDA, see https://www.ffiec.gov/hmda/history2.htm

[61] For a small number of cells, the count in our dataset exceeds the adjusted HMDA count. In those instances, we use our loan count rather than the HMDA-based count, in effect setting the weight to one.

To begin, note that loans sold to the Enterprises after the year of origination are not counted as Enterprise loans in HMDA; this occurs because HMDA does not require respondents to amend their prior-year filings. Accordingly, when comparing to HMDA counts, we omit the loans in the FHFA dataset that were acquired by the Enterprises after the origination year. We then calculate the ratio of this same-year FHFA count to the HMDA count for each cell in the partition, separately for purchase loans and refis. The resulting ratio is the cell-specific gross-up factor for Enterprise loans in HMDA. We then assume that these cell-specific factors based on Enterprise loans apply for all conventional loans. With these factors, we calculate the universe count of portfolio loans for each cell in the partition as the grossed-up HMDA count of conventional loans minus the Enterprise count from the FHFA data and the PLS loan count (prior to cleaning) from the CoreLogic data. Our weight for each cell in the partition equals the ratio of the estimated universe count for portfolio loans to the count in our dataset.

For PLS loans, the underlying source data represent close to a universe count. Our data cleaning removes about 10 percent of these loans. To restore the original near-universe counts, we weight the loans in each cell of the partition by the ratio of the pre-cleaning to post-cleaning count. For Enterprise loans, we dispense entirely with weighting because the source data represent a full universe count and only a tiny fraction of the loans are removed through cleaning.

Imputation of missing values for risk factors

As described in Section 2 of the paper, we use the multiple imputation approach (Rubin, 1976) to impute missing data. A relevant question is why impute at all: dropping observations with missing information is conceptually much simpler. The key issue is whether the missing information is correlated with other variables. If missing information is not correlated with any observed or unobserved variables, it is classified as missing completely at random (MCAR), in which case observations with missing information can be safely dropped without introducing any bias. However, if missing information is correlated with other observable variables, then dropping observations will result in biased statistics.[62]

[62] The two alternative missing data structures to MCAR are missing at random (MAR) and missing not at random (MNAR). Under MAR, the missing values for a variable *x* are correlated with other observed variables, but whether an observation of *x* is missing does not depend on the value of *x* itself. In contrast, under MNAR, the missingness of *x* does depend on the value of *x*. Multiple imputation is suitable for the MAR structure, but not for MNAR, which requires the researcher to model the process that generates missing values. It is not possible to test whether the data

If the data were MCAR, we should observe no systematic variation in the likelihood of missing values across market segments (Enterprise, PLS, etc.) or risk characteristics (DTI, credit scores, etc.). However, we do find substantial evidence that missing information in our mortgage data is not MCAR. For instance, Enterprise loans with missing credit scores tend to have lower reported incomes. Based on this evidence, we proceed with an imputation strategy rather than dropping observations to avoid a known bias in the reported statistics. We impute credit score, DTI, documentation status, occupancy status, and the three components of amortization status (balloon status, negative amortization status, and interest-only status) for both purchase loans and refinance loans. For refis, we also impute the type of refinance (rate-and-term or cash-out) and the CLTV (only for FHA/VA loans in the Ginnie Mae source data).[63]

Within the multiple imputation framework, we use predictive mean matching (PMM) as developed in Little (1988) and Rubin (1986). The PMM procedure uses regression-based predictions to find the nearest "donors" of a given variable for loans with missing values. We randomly select one of the eight closest donors to provide the missing value. We use the PMM procedure for two reasons. First, PMM is well suited for imputing variables that are not normally distributed, such as credit scores and DTIs. And second, PMM avoids the convergence problems that can affect other methods, such as logit models, for predicting categorical variables. For the imputation of occupancy status, we use a conditional imputation model that first imputes if a loan is primary owner-occupied, and then imputes whether it is an investor loan or a second home loan only if it is found not to be primary owner-occupied. To allow the estimated parameters to vary over time, every PMM model is run over the shortest time windows that we judged to have sufficient data for reliable estimation. When there are insufficient observations for a monthly window, we use a mix of quarterly, half-yearly, annual, and grouped-year windows depending on loan counts. See Table A.2 for details.

We use multiple imputation chained equations (MICE) to iteratively impute all of the variables with missing data in our dataset. An iterative procedure is necessary because the

are MAR versus MNAR because that would require the missing data to be known. We assume missing data in our datasets are MAR in order to implement multiple imputation. Under MAR, it is important to point out that the fraction of missing values only affects the uncertainty of the imputation.

[63] The theory of multiple imputation is based on having access to infinite imputations, but according to Rubin (1987), an average of just two imputations has 90% asymptotic efficiency with a 50% rate of missing values. For higher missing rates, the consequence is reduced efficiency but not bias. To assess the accuracy of the imputations for the first and second moments, we have performed a cross-validation exercise for PLS imputations and found them to be closely aligned with observed moments.

structure of our missing data is not monotone (i.e., the variables with missing data cannot be ordered such that the loans with missing data for the first variable are a subset of the loans with missing data for the next variable, and so on.) The imputation equations include a large set of explanatory variables: dummy variables for the state, product type, property type, loan term, documentation status, occupancy status, balloon status, negative amortization status, interest-only status, default status, month of origination, and market segment, as well as fourth-degree polynomial functions of the interest rate, loan amount, DTI, credit score, reported CLTV, shocked CLTV, and borrower income. We perform ten iterations of the model and then draw the set of imputed values.[64]

To assess the robustness of the imputed values, we repeated the imputation procedure 40 times with the PLS, portfolio, and FHA/VA datasets (we did not attempt the same exercise with the much larger Enterprise dataset due to computational limitations). We used these results to construct 95 percent confidence bands for the risk factors imputed for those loans and for the implied stressed default rate. The confidence bands pertain to the full respective sets of loans, including those with reported values for the variable shown. Consequently, for risk factors that require few imputations, the confidence bands will be very tight by definition. Conversely, the band would be wide for a year in which a risk factor is largely imputed and the imputed values vary widely across the 40 repetitions of the imputation procedure.

As an illustration of the results, Figures A.1 and A.2 present the confidence bands for the average DTI and average credit score – the two most heavily imputed risk factors – for PLS and portfolio loans, respectively. For PLS loans, the confidence bands for the average DTI are wide in the early 1990s and again during the financial crisis (the latter only for purchase loans) but are tight for other years; the bands for the average credit score are tight over the entire period. For portfolio loans, the confidence bands for both risk factors are narrow even during the 1990s. Although not shown, the analogous bands for FHA and VA loans are generally narrow as well, except during the early 1990s. Recall that we do not report results in the paper for the average DTI before 1993 and the average credit score before 1994. Overall, the results concerning the confidence bands, combined with the 1993 and 1994 start dates, indicate that the model

[64] In the FHA and VA datasets from 2013 onward, we use number of units rather than property type and the loan's origination channel rather than default status, as property type and default status are not available in the Ginnie Mae data we use for those years. Borrower income is available only in the Enterprise dataset.

uncertainty associated with loan-level imputations has little effect on the summary statistics we report in the paper.

However, as we discuss in the main text, even when there is little model uncertainty, the resulting summary statistics still could be biased if the available source data used to estimate the imputation models are not a random sample. Because of this concern, we elected not to report results for DTIs or credit scores until these risk factors have good coverage in the Enterprise data, which helps anchor the results for the market as a whole. We also make use of external benchmark data to correct potential biases, as described in the next section.

External benchmarking

The National Mortgage Database (NMDB) serves as our primary source of external benchmark information. The NMDB is a 5 percent random sample of credit files back to 1999 provided by Equifax, one of the national credit repositories; these credit files are linked to other data sources to fill out loan and borrower characteristics beyond those contained in the credit files. The NMDB is still undergoing internal quality checks at FHFA. We use only the data for credit scores, which are pulled directly from the credit files, and the data for DTIs, which – when missing – are calculated by NMDB staff based in part on loan information in the credit files. Even for these risk factors, we benchmark to the NMDB only when the data line up with other data sources known to be reliable or accord with well documented features of the mortgage market.

In addition to the NMDB, we rely on a few other sources of external benchmark data. These sources include data on the distribution of credit scores for all mortgage loan originations in 1992 and 1998, commentary on the low/no doc share of mortgage originations in the 1990s, and information on DTIs in An et al. (2007) and credit scores in Newberger (2011) that is specific to FHA purchase loans. Table A.3 provides a “score card” for the benchmarking, which affects portfolio, FHA, and VA loans.

Portfolio loans

We benchmark to the annual NMDB distributions of credit scores for portfolio purchase loans and refinance loans from 1999 to 2019. This benchmarking is accomplished by reweighting the portfolio loans in our dataset to replicate the annual NMDB distribution of credit scores across the buckets used for our stressed default tables (i.e., less than 580, 580-619, and so on). We do not utilize the NMDB data on portfolio loan DTIs because of concerns that the NMDB may overstate the share of loans with DTIs above 43 percent in recent years.

Before 1999, we benchmark credit scores for portfolio purchase and refinance loans using data on the credit score distribution for all mortgage originations from Equifax for 1992 and FICO for 1998.[65] The 1998 data provide less detail on the score distribution than the 1992 data, but both sources shows the share of loans with scores below 660. We also use the data to approximate the average credit score in both years, and then phase-in the change in the below-660 share and the average score for years between 1992 and 1998. With these annual benchmark values for all mortgage loans, we adjust the score distribution for portfolio loans so as to approximate the benchmarks. Portfolio loans are the most logical market segment for adjustment because we already have good data on credit scores for Enterprise loans from FHFA, and portfolio loans are by far the largest of the other market segments before 1999.

Our final adjustment to the portfolio loan data involves the share of loans with low or no documentation. The information on doc status in the 1990s is very sparse in McDash, and for the loans with reported data, the low/no doc share is unrealistically high, often above 50 percent. The LLMA data are somewhat better on both counts – there is less missing data and the low/no doc share is lower. As a result, we set aside the information on doc status in McDash until 2002, the first year in which the low/no doc shares for portfolio loans in McDash and LLMA converge. For 1990-2001, we use the LLMA data to impute doc status for all McDash loans and for LLMA loans with missing information.

Although this procedure reduced the low/no doc share for portfolio loans in the 1990s, the share still appears to be too high given the near-disappearance of low/no doc loans for the Enterprises in the early 1990s and the slow re-emergence of these loans in the late 1990s.[66] We believe the issue is that our dataset contains portfolio loans classified as low/no doc that have LTVs above the limits that were typical for these loans. Pinto (1991) indicates that low/no doc programs in the late 1980s and 1990 generally required LTVs to be 75 percent or less. When these loans re-appeared after the lull through 1996, a contemporaneous press account suggests that lenders were limiting LTVs to 60 or 70 percent (American Banker, 1997). Accordingly, we

[65] We thank Edward Pinto for providing the Equifax and FICO data.
[66] See American Banker (1997) for evidence that the low/no doc share for portfolio lenders largely echoed the pattern for the Enterprises.

impose LTV limits on portfolio loans classified as low/no doc. Loans with LTVs above the following limits by origination year were reclassified as full doc: 75 percent for 1990, 60 percent for 1991-1996, 70 percent for 1997, 75 percent for 1998, 80 percent for 1999, and 90 percent for 2000. No limits are imposed for loans originated in later years.

FHA purchase loans

We benchmark our DTI results for FHA purchase loans to the NMDB for 1999-2012; we do not benchmark DTIs for 2013-2019 because our source data from Ginnie Mae for those years are very complete. For years before 1999, we benchmark to the DTI distributions from An et al. (2007) for 1992 and 1996. For 1993-1995, we linearly interpolate the 1992 and 1996 DTI bucket distributions and benchmark to the interpolated distributions. For 1997 and 1998, we follow the same interpolation method using the An et al. distribution for 1996 and the NMDB distribution for 1999.

Regarding credit scores, we use the annual distributions for 2004-2009 for FHA purchase loans in Newberger (2011).[67] For those years, we combine the credit score buckets in Newberger with the DTI buckets in the NMDB to create joint annual benchmark distributions for both risk factors. We reweight our FHA data for those years to match the combination of the credit score and DTI distributions.

VA purchase loans

Consistent with our treatment of FHA purchase loans, we benchmark the DTIs for VA purchase loans to the NMDB for 1999-2012, but do not benchmark to the NMDB credit score data because these data differ from the distribution of scores in the Ginnie Mae data for 2013-2019. We found no benchmark information for VA loans prior to 1999. The only adjustment we make before 1999 is to the DTI distribution for VA purchase loans for 1997-1998. For these years, the share of loans with high DTIs falls substantially, which seems unlikely to have been the case in reality. We interpolate the DTI distribution for these years using the results for 1996 and 1999 as endpoints.

[67] We do not use the NMDB credit score data for FHA purchase loans because we found these data did not match the distribution of scores in the Ginnie Mae data for 2013-2019. This difference does not necessarily imply any error in the NMDB data. It could be, for example, that the NMDB convention for calculating the credit score for loans with more than one borrower differs from that in the administrative data provided by Ginnie Mae.

FHA and VA refinance loans

We benchmark rate-and-term refis and cash-out refis separately to the NMDB data. For 1999-2012, we benchmark both types of refis to the joint distribution of DTIs and credit scores in the corresponding NMDB data. For 2013-2019, the Ginnie Mae data that we use have very few missing values for credit scores or DTIs for cash-out refis, so there is no need to benchmark these loans to the NMDB. However, there are many rate-and-term refis with missing values in the Ginnie Mae data; these are streamlined refinance loans that did not require the lender to verify credit scores or calculate DTIs. Consequently, for rate-and-term refis, we benchmark both credit scores and DTIs to the NMDB for 2013-2019.

Table A.1: Loan Counts and Exclusions

	Type of Loan			
	Enterprise	Potential Portfolio	PLS	FHA/VA
Initial count (millions)	140.08	24.41	21.97	51.34
Percent excluded for reason shown				
Loan amount not reported	0.04	<0.01	0.02	<0.01
LTV not reported	<0.01	4.43	0.00	3.23
Interest rate not reported	<0.01	0.68	1.73	0.88
State not reported	0.08	1.31	0.12	0.36
ZIP code not reported or not valid	<0.01	1.11	1.47	0.19
Duplicate loans	<0.01	8.41	0.00	30.95
Removal of PLS loans	NA	7.60	NA	NA
Term not reported	0.18	0.19	0.39	0.03
Product type not reported or reported inconsistently	0.11	0.61	0.30	0.35
Property type not reported	0.02	0.02	0.31	<0.01
Occupancy status not reported	<0.01	NA	0.18	NA
LTV < 25 percent	1.14	1.30	0.82	0.06
CLTV > 135 percent if not a HARP loan	0.01	0.62	0.02	0.08
PLS loans securitized more than one year after origination	NA	NA	5.22	NA
Total percent excluded	1.58	26.28	10.59	36.12
Final count (millions)	137.87	18.00	19.64	31.97

Note: The counts represent first-lien home purchase and refinance mortgages originated in 1990-2019 and secured by 1-4 unit properties. The initial count of potential portfolio loans excludes conventional loans that (1) the investor codes in the servicer datasets ever show as GSE or (2) have initial amounts below the conforming loan limits with no investor information. The percent of loans excluded in each column is rounded to the nearest 0.01. The total percent excluded will be less than the sum of the percent excluded in the preceding rows if some loans are excluded for more than one reason. NA in the row for occupancy status indicates that we did not exclude these loans when occupancy status was missing. As noted in the text, we assumed all FHA/VA loans with missing data were primary owner-occupied, and we imputed missing occupancy status for potential portfolio loans. *Source*: Authors' calculations using data from Black Knight, Inc., CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Table A.2: Estimation Periods for Imputation Regressions

Risk factors	Estimation periods
Enterprise (all loan types)	
Credit score, DTI, documentation status	Monthly: 1995:Jan.-2019:Dec.
	Grouped months: 1990-1994:Jan., 1990-1994:Feb., … , 1990-1994:Dec.
PLS purchase loans	
Credit score, DTI, documentation status	Monthly: 1999:Jul.-2007:Jun.
	Quarterly: 1998:Q3-1999:Q2
	Half-yearly: 1997:H1-1998:H1, 2007:H2
	Yearly: 1994
	Grouped years: 1990-1993, 1995-1996, 2008-2012, 2013-2014, 2015-2017, 2018-2019
PLS rate-and-term refinance loans	
Credit score, DTI, documentation status	Monthly: 1998:Apr.-2007:Jun.
	Quarterly: 1997:Q1-1998:Q1, 2007:Q3, 2007:Q4
	Yearly: 1994, 2008
	Grouped years: 1990-1993, 1995-1996, 2009-2012, 2013-2017, 2018-2019
PLS cash-out refinance loans	
Credit score, DTI, documentation status	Monthly: 2001:Apr.-2007:Jun.
	Quarterly: 1997:Q1-2001:Q1
	Half-yearly: 2007:H2
	Yearly: 1996, 2012-2014
	Grouped years: 1990-1993, 1994-1995, 2008-2011, 2015-2017, 2018-2019
Portfolio purchase loans	
Credit score, DTI, documentation status, amortization status, occupancy status	Monthly: 2002:Oct.-2007:Sep.
	Quarterly: 1995:Q1-2002:Q3, 2007:Q4-2008:Q2, 2014:Q3-2019:Q4
	Half-yearly: 1993:H1-1994:H2, 2008:H2, 2011:H1-2014:H1
	Yearly: 1990-1992, 2009-2010

Table A.2: Estimation Periods for Imputation Regressions (continued)

Portfolio refinance (rate-and-term and cash-out)	
Credit score, DTI, documentation status, amortization status, occupancy status, type of refinance loan	Monthly: 1998:Jan.-1998:Dec., 2001:Jan.-2002:Mar., 2002:Jul.-2007:Sep.
	Grouped months: 2018:Jul.-2019:Dec.
	Quarterly: 1996:Q1-1997:Q1, 1997:Q4, 1999:Q1-1999:Q2, 2000:Q1-2000:Q4, 2002:Q2, 2007:Q4-2013:Q4, 2015:Q1-2016:Q4
	Half-yearly: 1993:H1-1995:H2, 1997:Q2-1997:Q3, 1999:H2, 2014:H1, 2014:H2, 2017:H1-2018:H1
	Yearly: 1990, 1991, 1992
FHA/VA purchase loans	
Credit score, DTI, 1990-1996	For credit score: Quarterly: 1992:Q1-1996:Q4; Yearly: 1990, 1991
	For DTI: Grouped years: 1990-1996*
Credit score, DTI, 1997-2012	Monthly: 1998:Jan.-2012:Dec.
	Quarterly: 1997:Q1 to 1997:Q4
Credit score, DTI, CLTV, 2013-2019	Yearly: 2013-2019
FHA/VA refinance loans (rate-and-term and cash-out)	
Credit score, DTI, type of refinance loan, 1990-1998	For credit score: Quarterly: 1992:Q1-1998:Q4; Yearly: 1990, 1991
	For DTI: Grouped years 1990-1998*
Credit score, DTI, type of refinance loan, 1999-2012	Monthly: 2003:Jan.-2012:Dec
	Quarterly: 2002:Q1-2002:Q4
	Half-yearly: 2001:H1, 2001:H2
	Yearly: 1999, 2000
Credit score, DTI, type of refinance loan, and CLTV, 2013-2019	Yearly: 2013-2019

* For FHA and VA purchase loans, a large share of the reported DTIs had implausible values. As a result, we estimated the DTI imputation regression with data for 1996 and used the 1996 loans as potential donors to impute DTIs for loans originated in 1990-1995. For FHA and VA refinance loans, we estimated the imputation regression with data for 1996-1998 and used loans originated in those years as the donors.

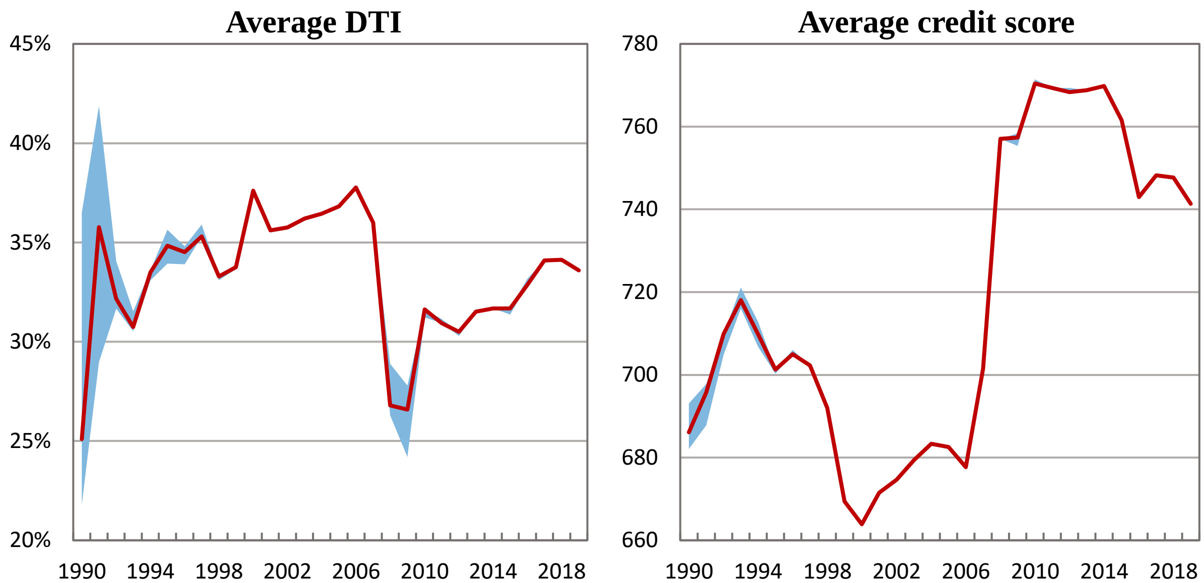
Table A.3: Summary of Benchmarking Data for Credit Scores and DTIs

Market Segment	Loan Type	Credit score (years and source)	DTI (years and source)
Portfolio	Purchase	1999-2019: NMDB Pre-1999: Equifax, FICO	NA
	Refinance	1999-2019: NMDB Pre-1999: Equifax, FICO	NA
FHA	Purchase	2004-2009: Newberger (2011)	1999-2012: NMDB Pre-1999: An et al. (2007), NMDB
	Refinance	1999-2012: NMDB (cash-out) 1999-2019: NMDB (rate-and-term)	1999-2012: NMDB (cash-out) 1999-2019: NMDB (rate-and-term)
VA	Purchase	NA	1999-2012: NMDB
	Refinance	1999-2012: NMDB (cash-out) 1999-2019: NMDB (rate-and-term)	1999-2012: NMDB (cash-out) 1999-2019: NMDB (rate-and-term)

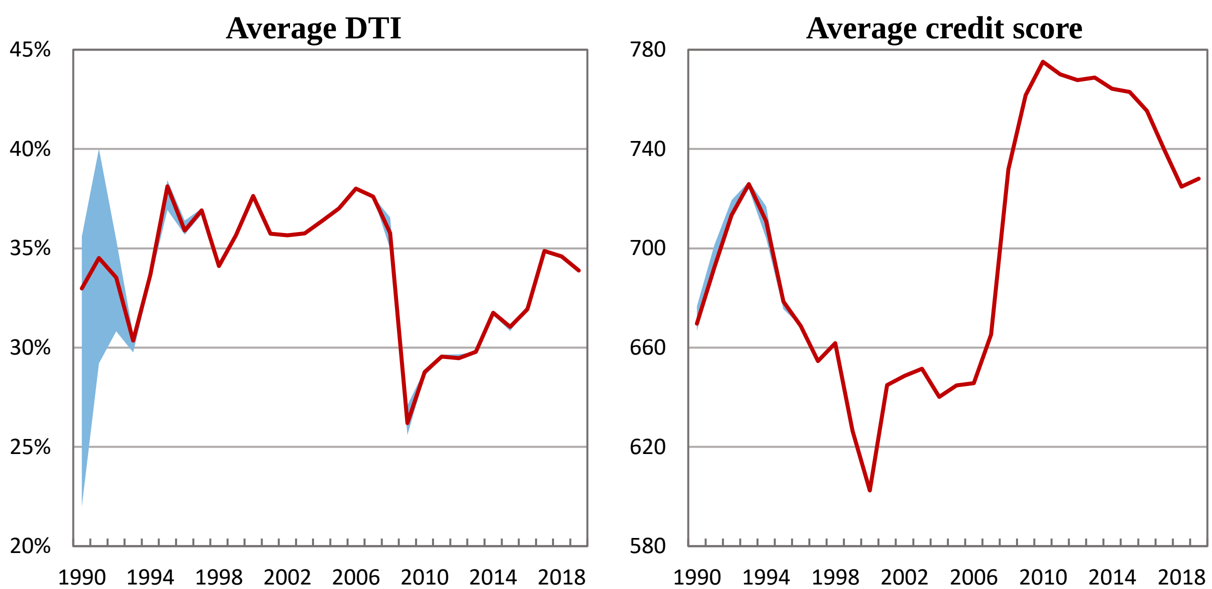
Note: NA indicates that no benchmarking was done. The benchmarking to the NMDB for refinance loans was done separately for rate-and-term refis and cash-out refis. For the pre-1999 period, the reference to the NMDB indicates that the 1999 NMDB data were used as an endpoint for interpolating earlier years.

Figure A.1: Confidence Bands for Selected Characteristics of PLS Loans

Purchase Loans

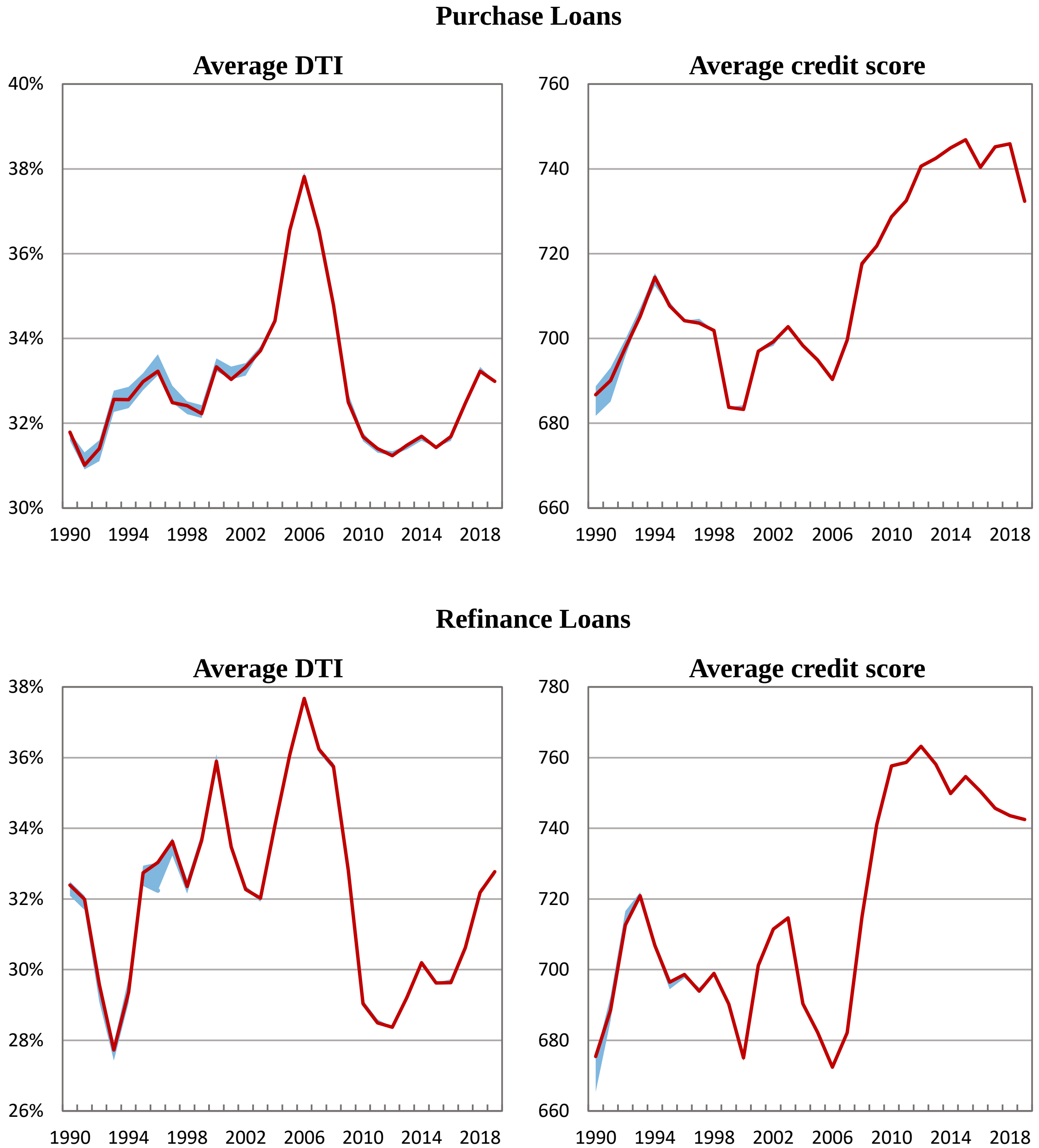


Refinance Loans



Note: Results pertain to mortgages secured by 1-4 unit properties. Area in blue represents the 95% confidence band that accounts for the imputation of missing values for risk factors.
Source: Authors' calculations using data from CoreLogic.

Figure A.2: Confidence Bands for Selected Characteristics of Portfolio Loans



Note: Results pertain to mortgages secured by 1-4 unit properties. Area in blue represents the 95% confidence band that accounts for the imputation of missing values for risk factors.
Source: Authors' calculations using data from Black Knight, Inc. and CoreLogic.

Appendix B: Machine Learning Models

To test the robustness of our baseline default-table methodology, we construct additional indicators of mortgage risk for Enterprise and PLS loans using two machine learning models.[68] For both models, we use the same set of risk factors as in the default tables and implement the models using loans originated in 2006 and 2007. The risk factors include the credit score, CLTV, DTI, loan term, loan type (fixed rate versus adjustable rate) occupancy status, amortization status, and income documentation status. The definition of default is the same as in the baseline default-table approach: we deem a loan to have defaulted if it was ever 180 days' delinquent or was terminated with less than full payment of principal.

The first machine learning model builds a regression tree based on the idea first proposed in Breiman et al. (1984).[69] The model, which performs functionally like a decision tree, compares splits of the data on all eight risk characteristics to find which one produces the most productive split and then repeats this step until only 500 loans are left in each bucket.[70] Then, the model builds 20 additional trees using 95 percent of the data, with the remaining 5 percent kept for cross-validation. The results from the 20 cross-validations determine the amount of tree complexity associated with the maximum predictive power. We then apply this complexity parameter to the initial decision tree, which "prunes" branches from the tree that are caused by spurious correlation. In the end, the final regression trees built for the six loan categories we analyze (purchase loans, rate-and-term refis, and cash-out refis for the Enterprise and PLS market segments) have between 250 and 700 terminal nodes. These terminal nodes are the counterpart in the regression tree to the 2,560 cells in the default tables.

Figure B.1 depicts the first several splits in the regression tree for PLS purchase loans to illustrate the results. The ordering of the splits provides information about the relative importance of the various risk factors for explaining defaults. The initial split is based on the loan's credit score, indicating that it is the single most informative risk factor. In all, credit scores account for 41 percent of the tree's fit, followed by CLTVs at 29 percent, amortization status at 9

[68] We focus on Enterprise and PLS loans because our data provide nearly complete coverage of loans in these market segments and thus do not require loan-level weighting to be nationally representative. This simplifies the estimation of the machine learning models.
[69] The package rpart was used to create these trees. For more on recursive partitioning, see https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf.
[70] For an overview of productive splits, see chapter 3 of the link in the previous footnote.

percent, income documentation status at 8 percent, with the remaining risk factors making up the rest.

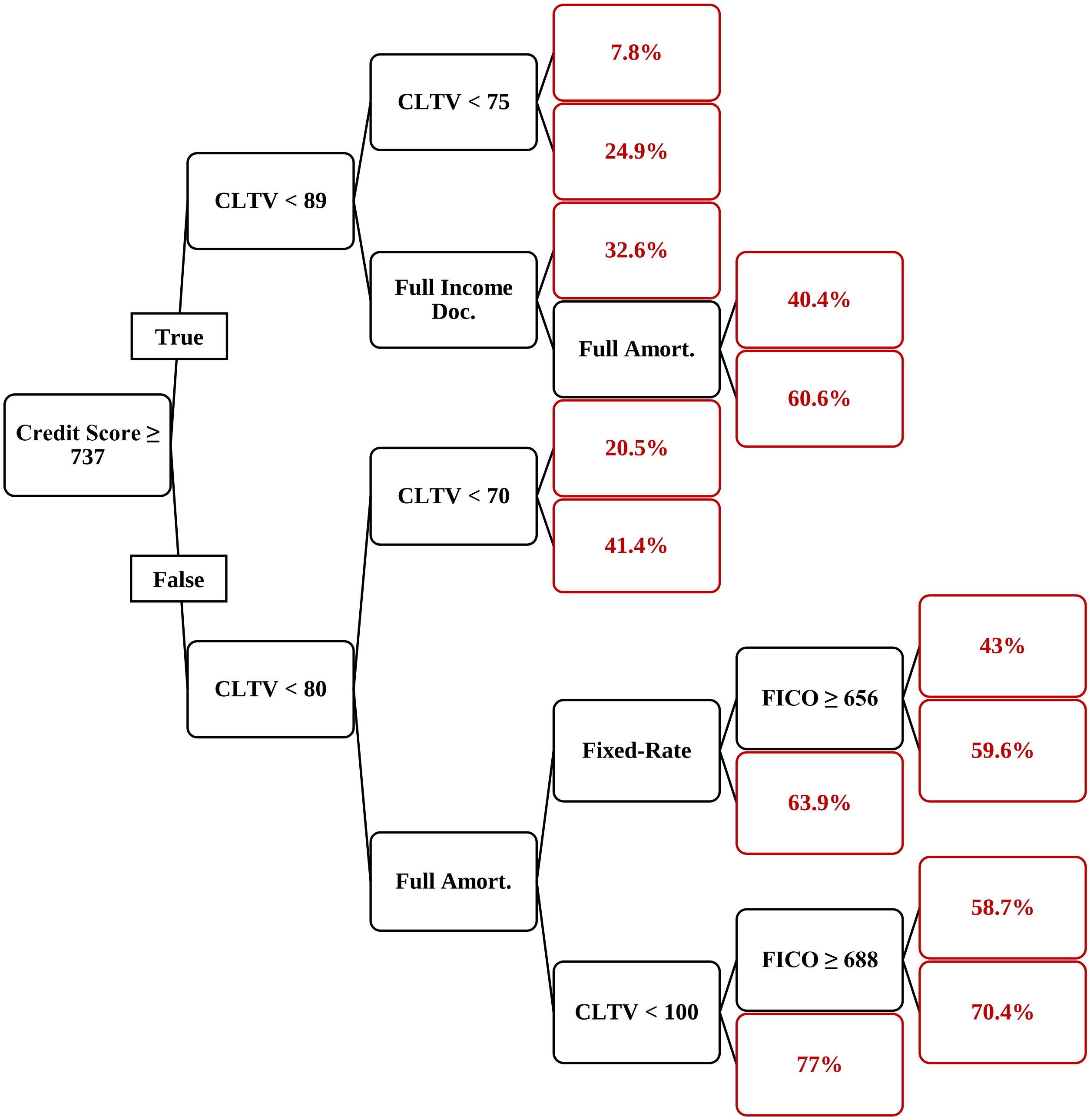
We also build a high-performance random forest to capture further relationships between variables in our dataset.[71] Random forests, proposed in Breiman (2001), are an extension of regression trees. We construct our forests as follows: first, we randomly select three of our eight risk factors; second, we construct a training sample by picking observations with replacement such that approximately two-thirds of the total sample is drawn to construct a given tree; and finally, we build a regression tree using those observations and variables. We repeat these steps 500 times to construct a proverbial "forest" of decision trees. By using only a subset of our risk factors and observations in every tree, each one captures a unique relationship among the selected risk factors. Additionally, selecting only a portion of loans to build each decision tree allowed us to validate our model using the out-of-sample observations.

For both the regression tree and random forest models, we generated stressed default rates that can be compared to those from our baseline default-table methodology. To create the alternative stressed default series, we calculate the predicted default rate for every loan originated from 1994 to 2019 using the results for each model. We then compute the average predicted default rate for purchase loans and the aggregate of both types of refinance loans in each origination year, just as in our baseline methodology.

Figures B.2 and B.3 present the results from this comparison for purchase loans and refis respectively. The figures show an extremely high correlation between our baseline results and the machine learning alternatives. Indeed, for most years, the stressed default series in each panel of Figures B.2 and B.3 are almost indistinguishable from one another. The bottom line from this exercise is that our baseline results are robust to the use of a machine learning approach.

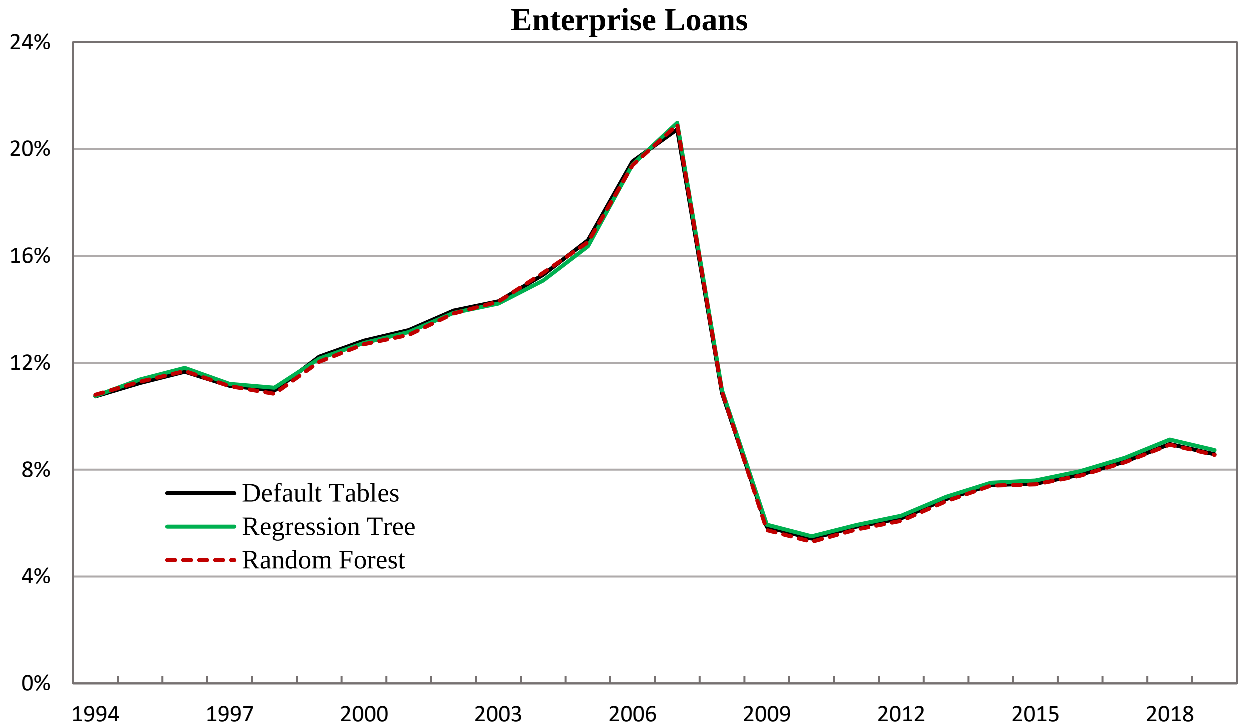
[71] The package Rborist was used to implement this model. For more on random forests, see https://cran.r-project.org/web/packages/Rborist/index.html.

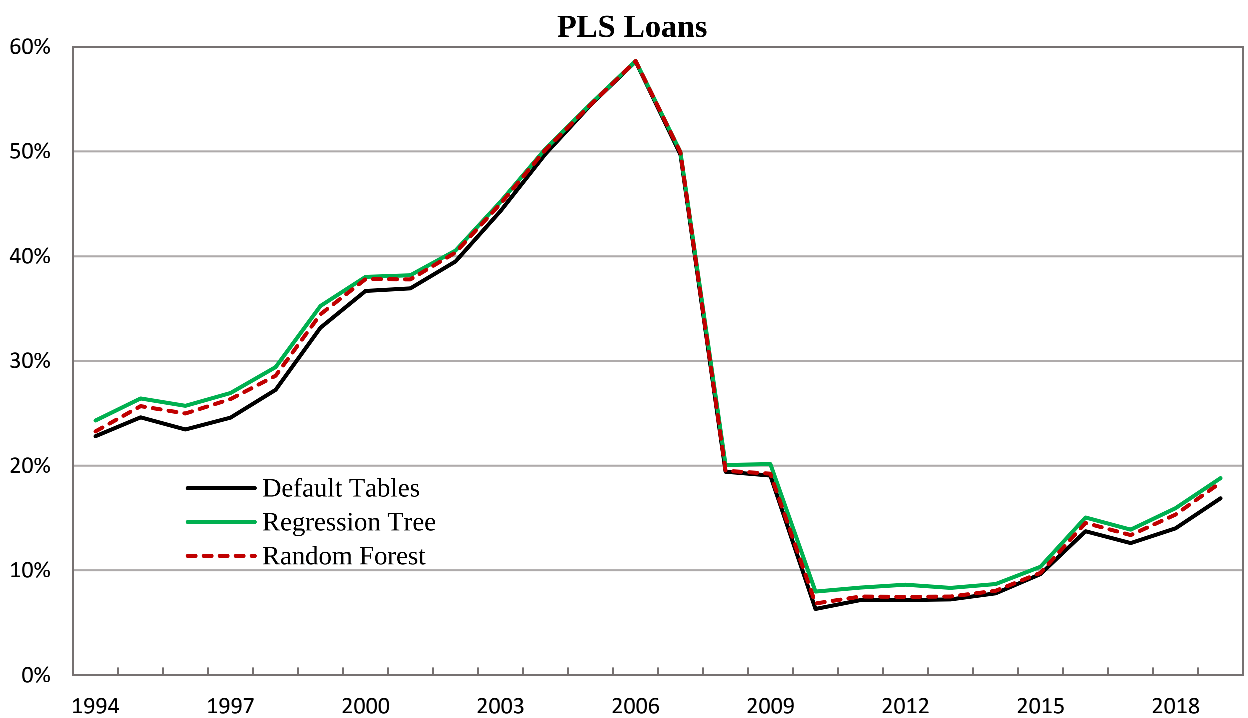
Figure B.1: Visualization of the First (Most Important) Splits in the Regression Tree for PLS Home Purchase Loans



Note: Results pertain to first-lien PLS home purchase loans originated in 2006-2007 and secured by 1-4 unit properties. For each split, the upward arm indicates that the condition shown in the decision node is true, while the downward arm indicates that the condition is false. Black represents decision nodes, while red represents terminal nodes. The value in every terminal node represents the default rate for loans in that node.
Source: Authors' calculations using data from CoreLogic.

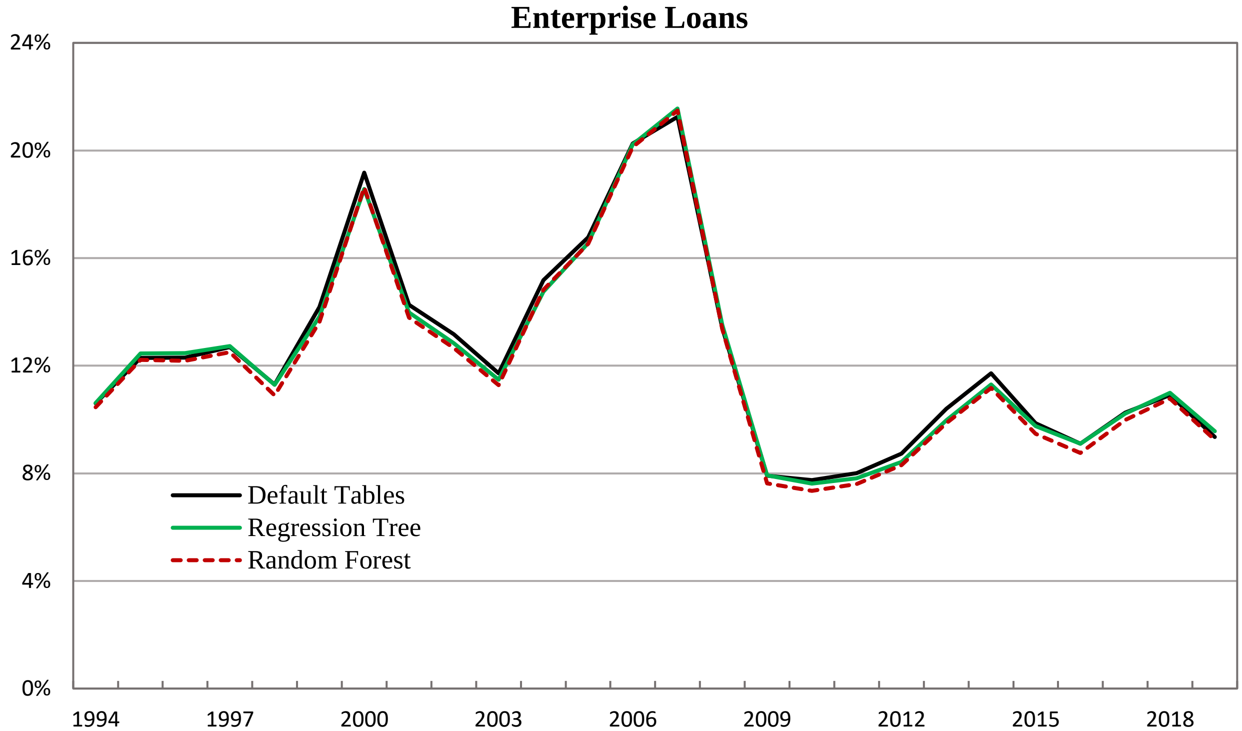
Figure B.2: Stressed Default Rates from Different Models, Enterprise and PLS Home Purchase Loans

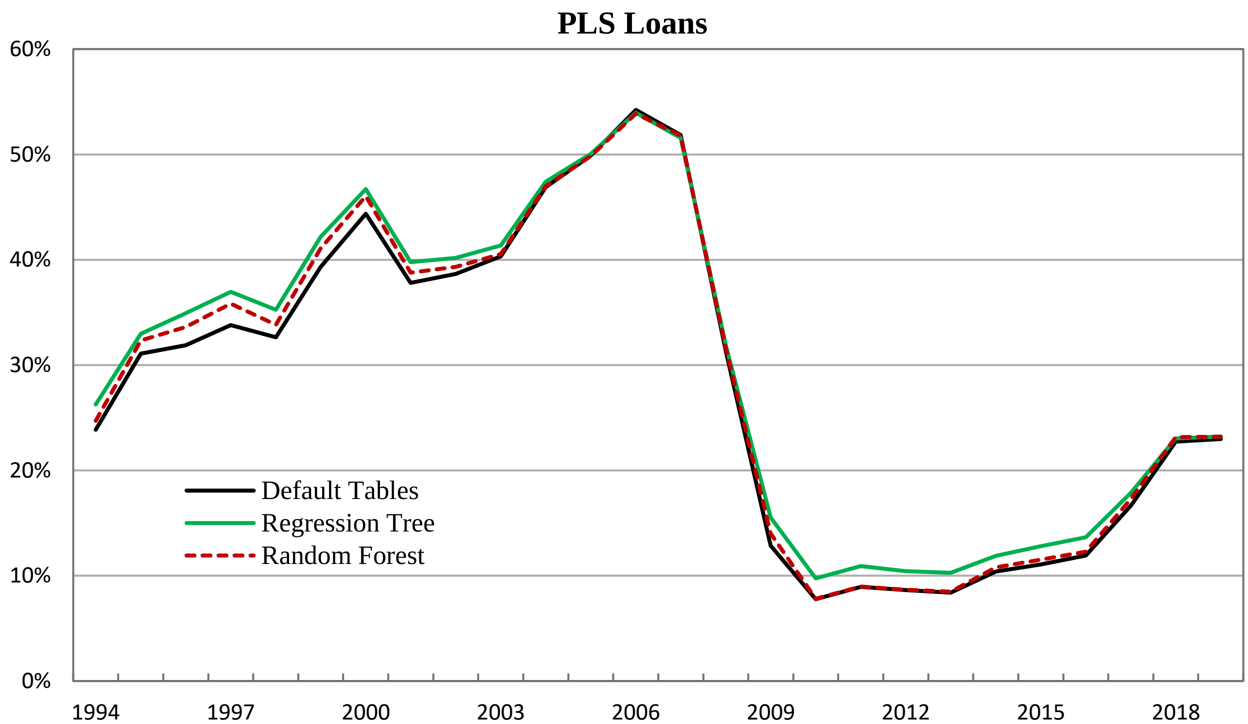




Note: Results pertain to first-lien home purchase loans secured by 1-4 unit properties.
Source: Authors' calculations using data from CoreLogic and FHFA.

Figure B.3: Stressed Default Rates from Different Models, Enterprise and PLS Refinance Loans



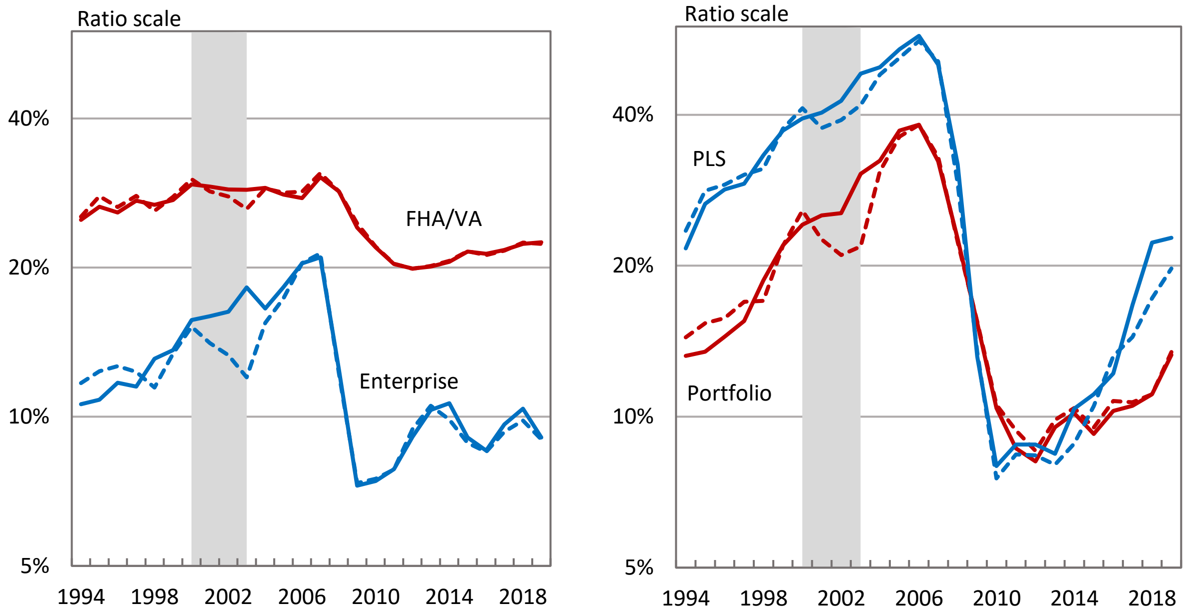


Note: Results pertain to first-lien refinance loans secured by 1-4 unit properties.
Source: Authors' calculations using data from CoreLogic and FHFA.

Appendix C: Additional Figures

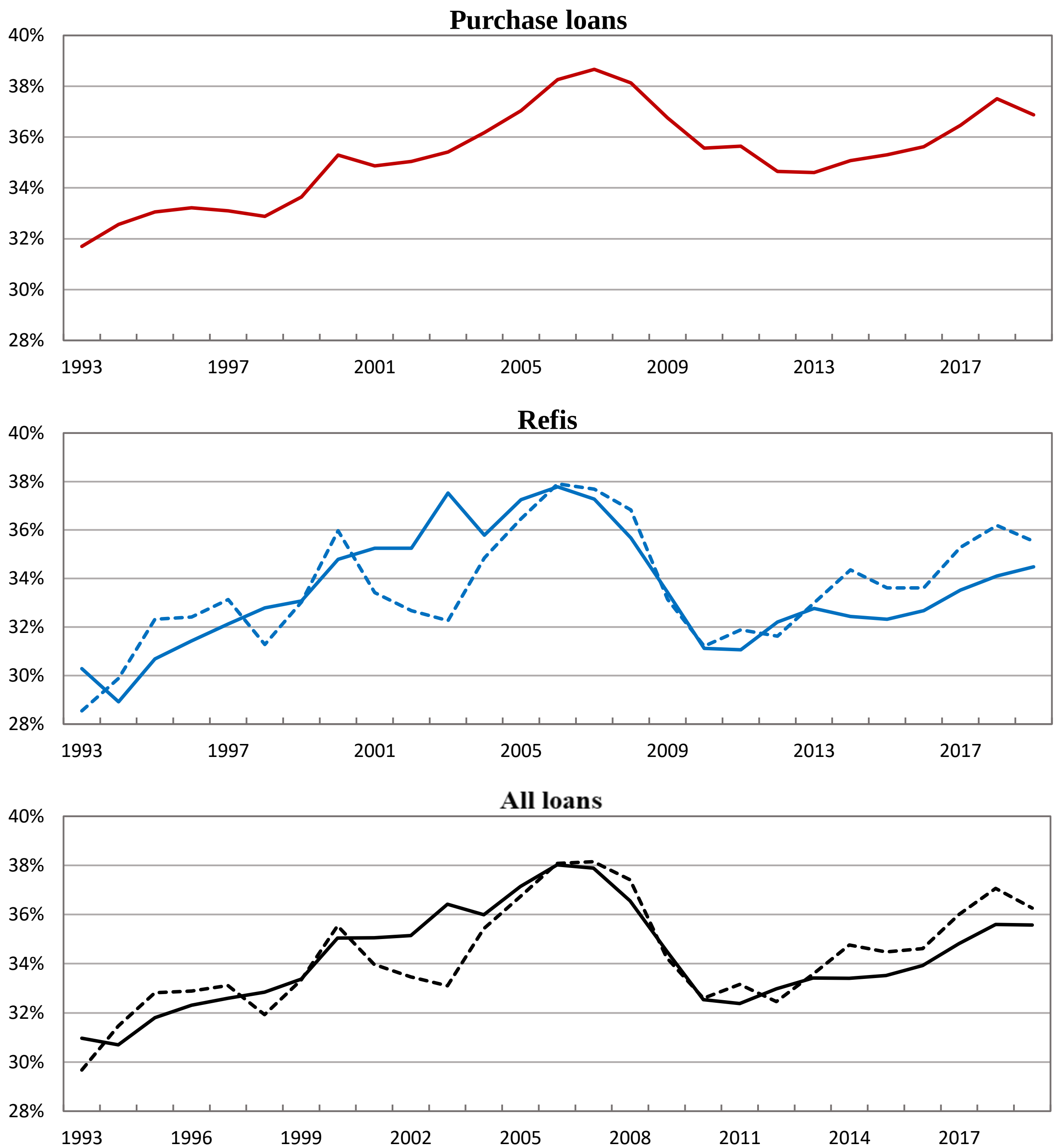
These figures provide detail on stressed default rates and risk factors that supplements what is presented in the main text.

Figure C.1: Stressed Default Rates for All Loans, by Market Segment
(Solid lines adjust for changes in refi volume; dashed lines show the unadjusted data)



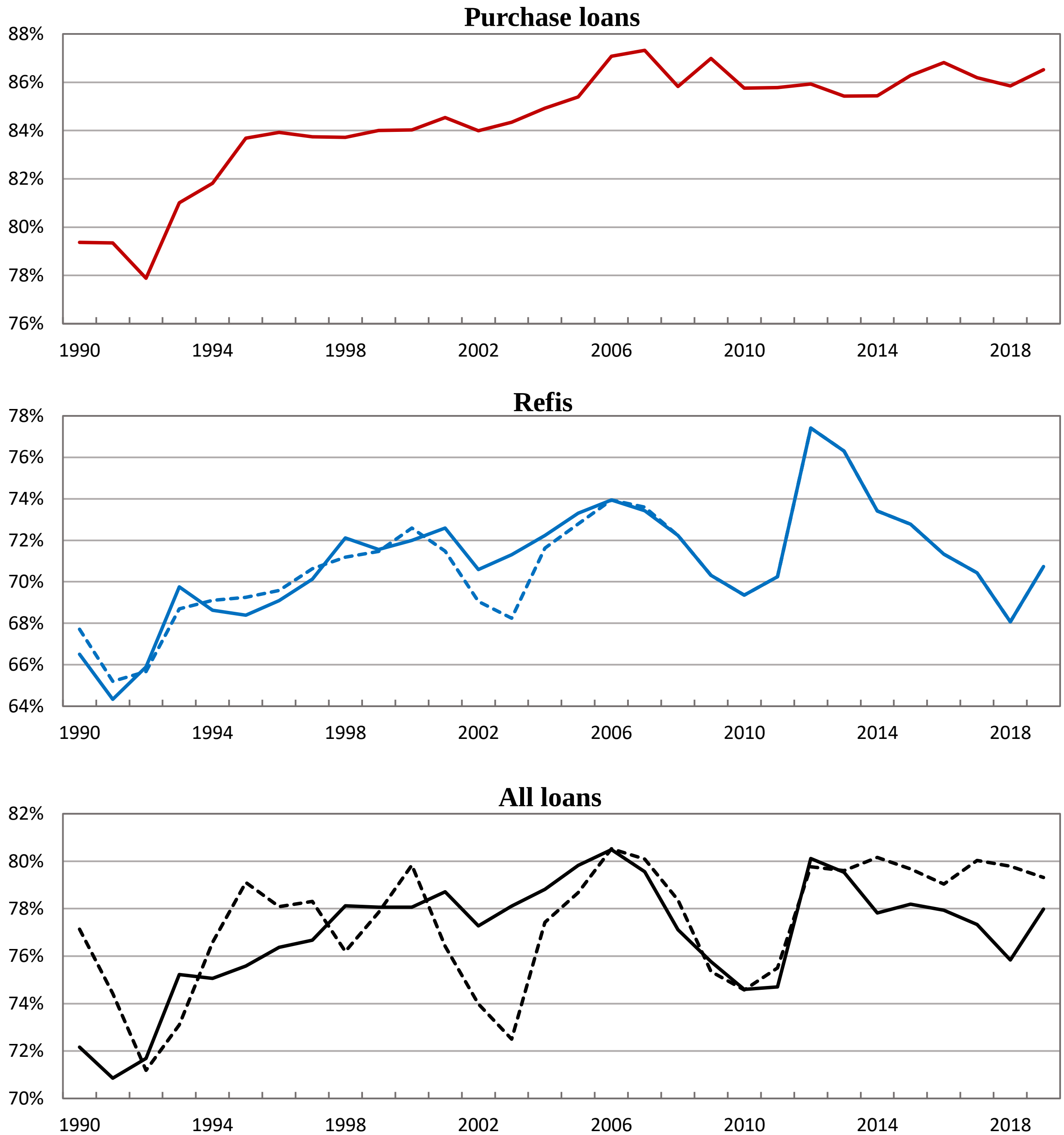
Note: The results pertain to home purchase and refinance mortgage loans secured by 1-4 unit properties. The adjusted series include a regression-based adjustment that controls for changes in refi volume; see the main text for details. Shading is for 2000-2003. *Source*: Authors' calculations using loan-level data from Black Knight, Inc., CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Figure C.2: Average DTI, by Loan Type
(For refis and all loans, solid lines adjust for changes in refi volume; dashed lines show unadjusted data)



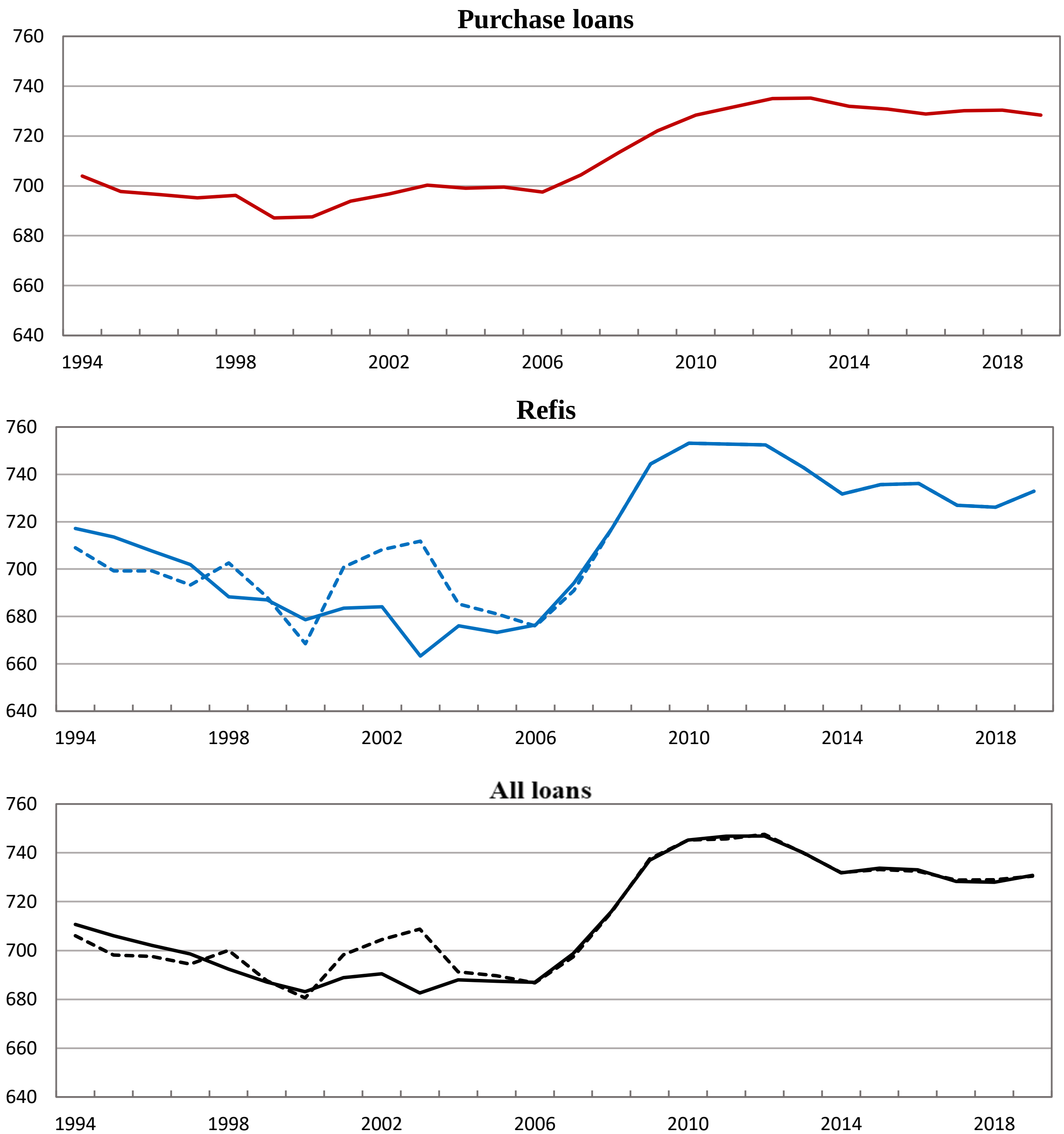
Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. The adjusted series include a regression-based adjustment that controls for changes in refi volume; see the main text for details. *Source*: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Figure C.3: Average CLTV, by Loan Type
(For refis and all loans, solid lines adjust for changes in refi volume; dashed lines show unadjusted data)



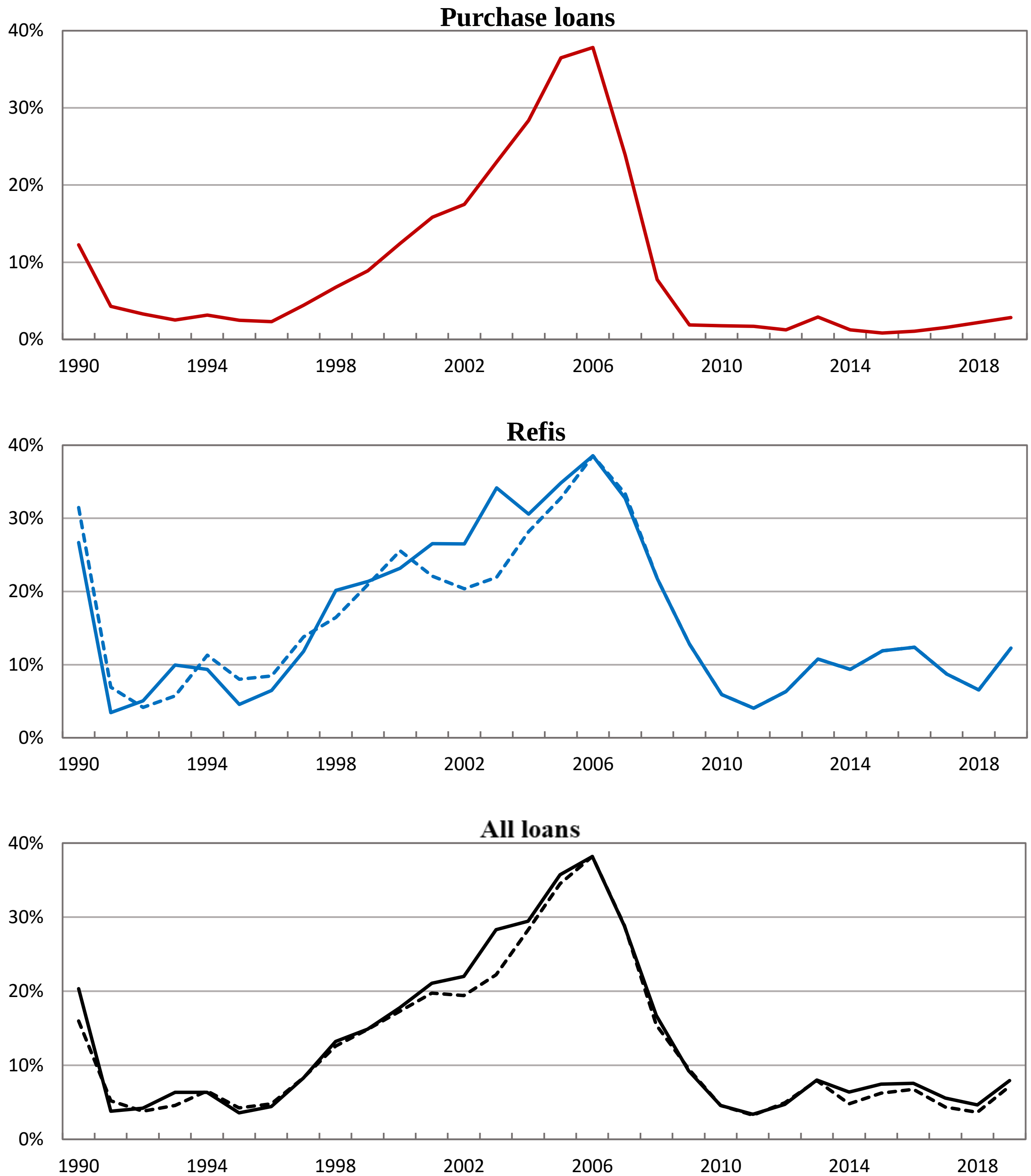
Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. The adjusted series include a regression-based adjustment that controls for changes in refi volume; see the main text for details. *Source*: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Figure C.4: Average Credit Score, by Loan Type
(For refis and all loans, solid lines adjust for changes in refi volume; dashed lines show unadjusted data)



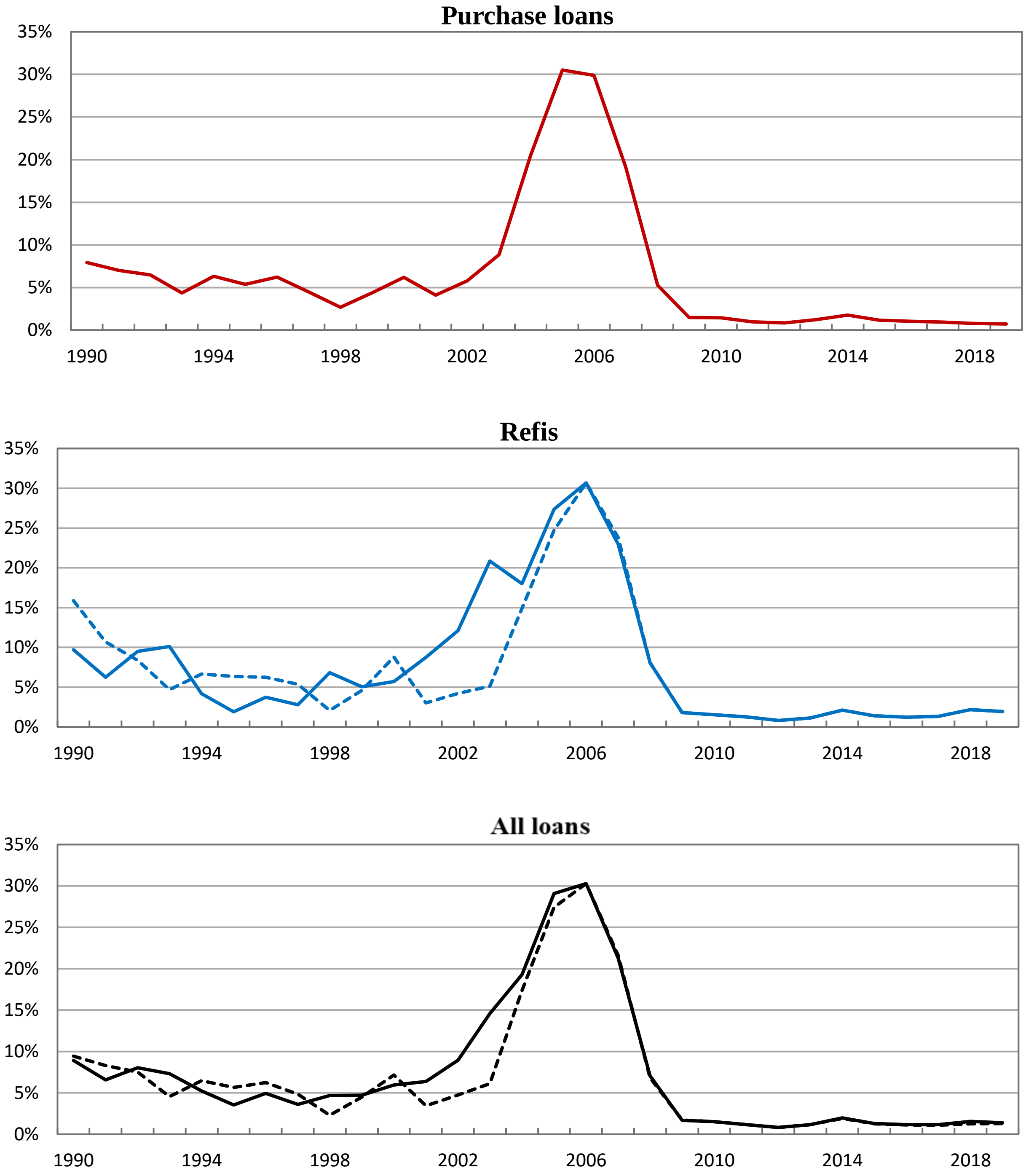
Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. The adjusted series include a regression-based adjustment that controls for changes in refi volume; see the main text for details. *Source*: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Figure C.5: Share of Loans with Low or No Documentation, by Loan Type
(For refis and all loans, solid lines adjust for changes in refi volume; dashed lines show unadjusted data)



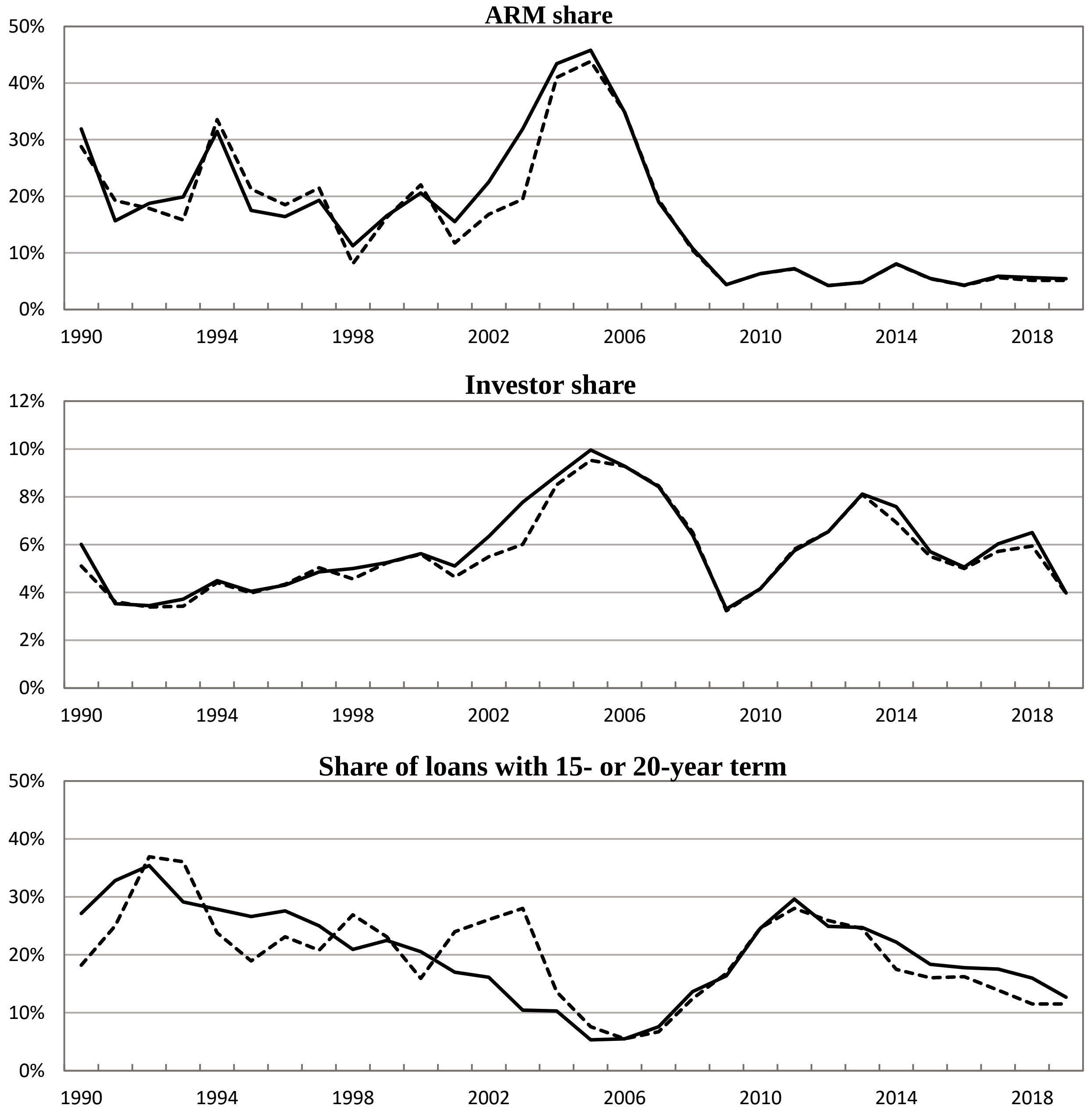
Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. The adjusted series include a regression-based adjustment that controls for changes in refi volume; see the main text for details. *Source*: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Figure C.6: Share of Loans with Non-standard Amortization, by Loan Type
(For refis and all loans, solid lines adjust for changes in refi volume; dashed lines show unadjusted data)



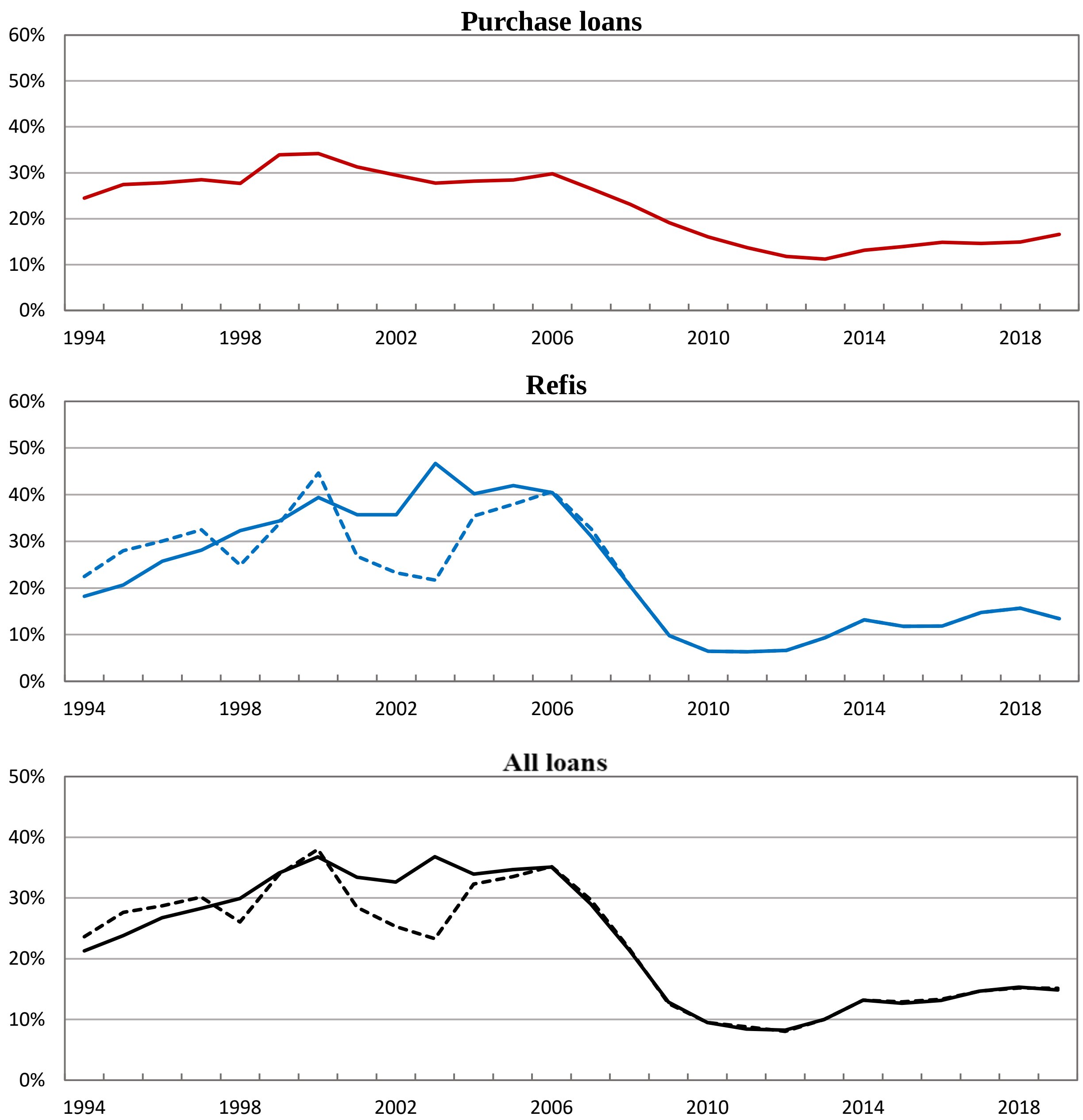
Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. The adjusted series include a regression-based adjustment that controls for changes in refi volume; see the main text for details. A loan is classified as having non-standard amortization if it has an interest-only period, negative amortization, and/or a balloon payment. *Source*: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Figure C.7: Other Characteristics, All Loans
(Solid lines adjust for changes in refi volume; dashed lines show the unadjusted data)



Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties. The adjusted series include a regression-based adjustment that controls for changes in refi volume; see the main text for details. *Source*: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Figure C.8: Below-660 Credit-Score Share, by Loan Type
(Solid lines adjust for changes in refi volume; dashed lines show the unadjusted data)

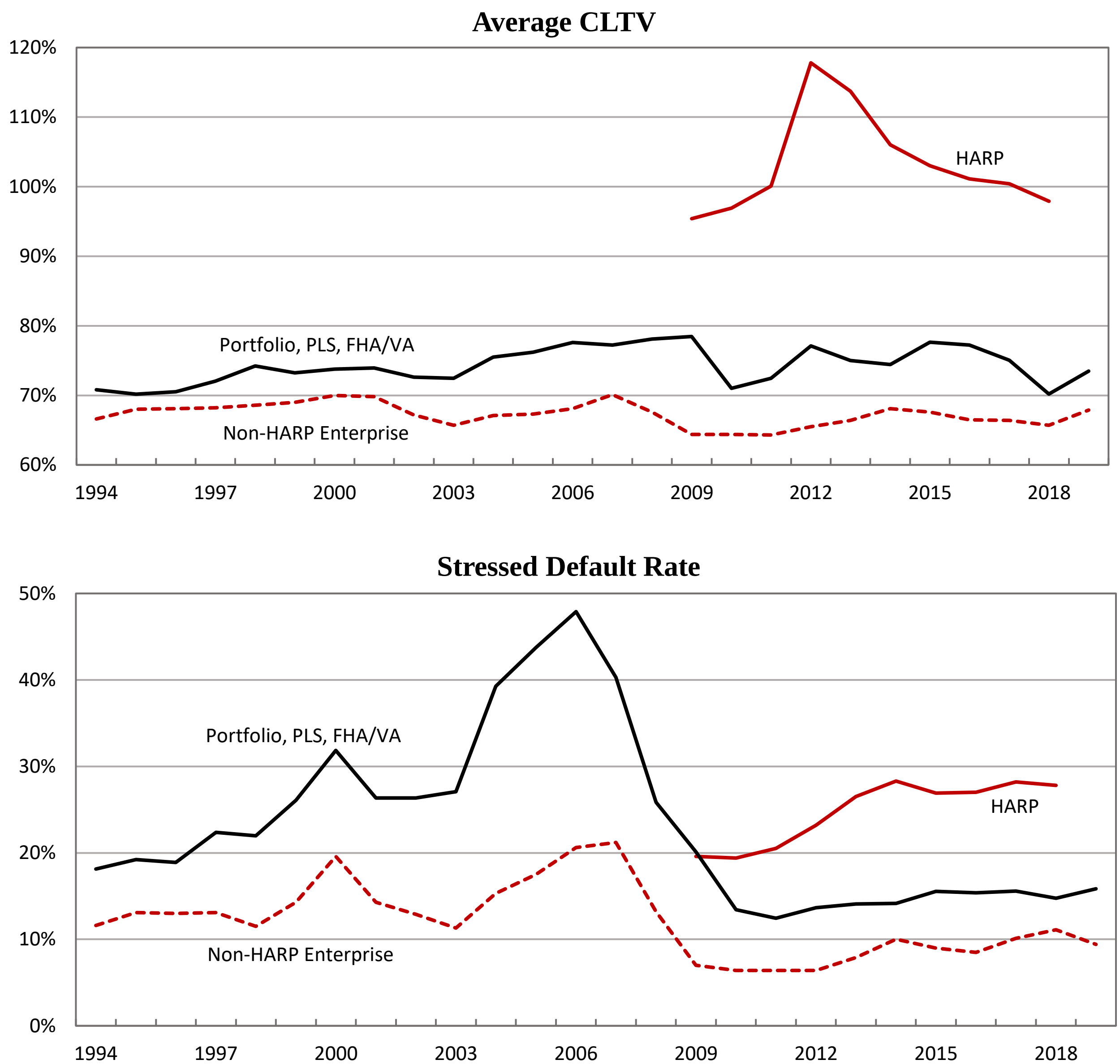


Note: The results pertain to first-lien home purchase and refinance mortgage loans secured by 1-4 unit properties The adjusted series include a regression-based adjustment that controls for changes in refi volume; see the main text for details. *Source*: Authors' calculations using data from Black Knight, Inc. CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

Appendix D: HARP versus Other Refinance Loans

The figure below compares HARP loans to other refinance loans along two dimensions: the average CLTV and the stressed default rate. The HARP series begin in 2009 with the inception of the program and end in 2018, when the program expired. As shown, HARP loans had much higher average CLTVs than both non-HARP Enterprise refinance loans and the aggregate of portfolio, PLS, and FHA/VA refinance loans in every year from 2009 to 2018. HARP loans also had much higher stressed default rates than non-HARP Enterprise refinance loans during this entire period and higher stressed default rates than the aggregate of other refinance loans in every year except 2009.

Figure D.1: HARP Loans versus Other Refinance Loans



Note: The results pertain to first-lien refinance loans secured by 1-4 unit properties and are not adjusted for changes in refinance volume. The HARP series begin in 2009 with the inception of the program and end in 2018, when it expired. *Source*: Authors' calculations using data from Black Knight, Inc., CoreLogic, FHFA, and Ginnie Mae data processed by the AEI Housing Center.

References for Appendices

Adelino, Manuel, William B. McCartney, and Antoinette Schoar. 2020. "The Role of Government and Private Institutions in Credit Cycles in the U.S. Mortgage Market." NBER Working Paper 27499.

American Banker. 1997. "Notorious No-Documentation Loans Paying Off for Some Mortgage Lenders." May 9.

An, Xudong, Raphael W. Bostic, Yongheng Deng, and Stuart A. Gabriel. 2007. "GSE Loan Purchases, the FHA, and Housing Outcomes in Targeted, Low-Income Neighborhoods." *Brookings-Wharton Papers on Urban Affairs*: 205-40.

Bhutta, Neil, and Benjamin J. Keys. 2018. "Eyes Wide Shut? The Moral Hazard of Mortgage Insurers during the Housing Boom.' NBER Working Paper 24844, July. Forthcoming, *Review of Financial Studies*.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45(1): 5-32.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees*. Chapman and Hall/CRC: Boca Raton, FL.

LaCour-Little, Michael, Charles A. Calhoun, and Wei Yu. 2011. "What Role Did Piggyback Lending Play in the Housing Bubble and Mortgage Collapse?" *Journal of Housing Economics* 20: 81-100.

Leventis, Andrew. 2014. "The Relationship between Second Liens, First Mortgage Outcomes, and Borrower Credit: 1996-2010." FHFA Working Paper 14-3, September. https://www.fhfa.gov/PolicyProgramsResearch/Research/PaperDocuments/WP_14-3_Second_Liens.pdf

Little, Roderick J. A. 1988. "Missing-Data Adjustments in Large Surveys." *Journal of Business Economics and Statistics* 6(3): 287-96.

Newberger, Harriet. 2011. "FHA Lending: Recent Trends and Their Implications for the Future." Federal Reserve Bank of Philadelphia, Community Development Studies and Education Discussion Paper, December. https://www.philadelphiafed.org/-/media/community-development/publications/discussion-papers/discussion-paper_fha-lending-trends-and-implications.pdf

Pinto, Ed. 1991. "Low Doc Takes Its Toll on Loan Performance." Mimeo, May 15.

Rubin, Donald B. 1976. "Inference and Missing Data." *Biometrika* 63(3): 581-92.

Rubin, Donald B. 1986. "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations." *Journal of Business & Economic Statistics* 4: 87-94.

Rubin, Donald B. 1987. Multiple Imputation for Nonresponse in Surveys. New York: Wiley.

Scheessele, Randall M. 1998. "HMDA Coverage of the Mortgage Market," U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Working Paper No. HF-007, July. https://www.huduser.gov/portal/Publications/pdf/wkgpapr_7.pdf

Szymanoski, Edward, Randall Scheessele, Christian Malagon, and Ishmail Mohammed. 2011. "A Look at the FHA's Evolving Market Shares by Race and Ethnicity," *U.S. Housing Market Conditions, 1st Quarter 2011*, May: 6-12. U.S. Department of Housing and Urban Development, Office of Policy Development and Research. https://www.huduser.gov/portal/periodicals/ushmc/spring11/ch1.pdf

Williams, Michael. 2015. "HMDA Coverage of the Mortgage Market," MPP Professional Paper, University of Minnesota, May 18.