Correlation and Residential Mortgage Defaults

Simone Varotto, Chiara Maria Ventura

July 31, 2023

Abstract

We utilize a sample of 25 million US mortgages originated between 1999 and 2017 to calculate pairwise mortgage correlations inferred from mortgage defaults. Our findings reveal that the flat correlation of 15% adopted by bank regulators inadequately considers the substantial variability present in US mortgage portfolios. This variability is primarily driven by the borrower's debt-to-income and loan size, while other obligor and loan characteristics also exert significant influence. Moreover, our study identifies that this heterogeneity leads to distortions in mortgage underwriting. As the 15% correlation is highly conservative, lenders lack a regulatory incentive to mitigate concentration risk. In fact, we demonstrate that some banks assign a negative price to correlation risk, resulting in concentrated portfolios. Finally, we estimate that borrowers could make mean savings of 4.41% on total interests for a standard mortgage by "shopping around", as lenders may charge economically different rates to the same customer based on (1) differences in their mortgage portfolio composition and (2) variations in their consideration of correlation risk.

Keywords: mortgages, default risk, correlation, Great Financial Crisis

JEL Classification: G01, G11, G21

1 Introduction

The US mortgage market has historically played a crucial role in major financial crises throughout the last century, including the Great Depression of the 1930s, the Savings and Loans crisis of the 1980s and 1990s, and the Great Recession of 2007-2009.

These crises were characterized by a high degree of correlation among borrower behavior, which resulted in a significant increase in mortgage defaults. This study aims to analyze the factors that contribute to the rise in pairwise correlations in mortgage portfolios by utilizing a comprehensive loan-level database that encompasses the period of the Great Recession. Our research contributes to the existing literature in the following ways.

First, to our knowledge, we are the first to use granular mortgage loan level data with extensive coverage of the US market to study empirical correlations segmented by borrower and loan characteristics. We find that mortgage correlations appear to be highly sensitive to such characteristics. This is important because, current international bank capital regulation is based on a flat unconditional correlation in mortgage portfolios of 15%. While we observe, in line with previous studies, that 15% is a conservative upper bound, (Botha and van Vuuren (2010), Chernih et al. (2006), Crook and Bellotti (2009)), our results indicate that ignoring the variability of portfolio correlation and its dependence on loan's and borrower's factors, effectively penalises portfolios that are more diversified, i.e. with a lower average correlation. As a result, current regulation could create incentives for banks to increase portfolio concentration which could lead to greater fragility in the banking system.

Second, our methodological approach is novel. Previous studies calculated correlations among mortgages either from residential mortgage-backed securities (RMBS) prices (Geidosch (2014)) or time series data available (Nickerson and Griffin (2017), Botha and van Vuuren (2010), Stoffberg and Vuuren (2015)), or from specialised lending institutions (Cowan and Cowan (2004)). Instead, we employ loan level data that enables us to condition our analysis on loan and borrower characteristics from a more extensive sample that represents the U.S. mortgage market. Our estimation approach employs the intuition that correlation is what drives the difference between long run default probabilities ($PD_{LongRun}$) and default probabilities in a crisis (PD_{Crisis}). Utilizing a logit model, we estimate both probabilities by exploiting the Great Recession as a benchmark crisis period. Then, we leverage the correlation-driven relationship between $PD_{LongRun}$ and PD_{Crisis}, based on a popular model adopted by bank regulators (BCBS (2005), Blumke (2018)), to ascertain the implied average pairwise correlation of individual borrowers within the lender's mortgage portfolio. This approach allows us to identify how borrower and loan characteristics influence mortgage portfolio correlations in a crisis. Our findings indicate that mortgage correlations are primarily affected by the borrower's loan size, debt-to-income ratios, and loan-to-value ratios.

Third, we examine whether banks price the risk of correlation in the interest rates charged to mortgage borrowers, while controlling for all other characteristics. New borrowers who exhibit higher (lower) correlation with existing borrowers in a bank's portfolio should be imposed a higher (lower) interest rate by the bank to compensate for the increased (decreased) risk of joint default in its mortgage portfolio during a crisis. Our findings indicate that while some lenders charge a positive premium for correlation risk (US Bank, Sun Trust, Provident), others apply a negative premium (JP Morgan Chase, Citi, Bank of America and Wells Fargo). We conjecture that a negative premium may be the result of (1) intense market competition that pushes interests down and disconnects them from portfolio concentration considerations, (2) an aggressive expansion strategy pursued by lenders to gain market share in a specific market segment, which yields a similar outcome as in point (1), and (3) the neglect of portfolio correlation risk because as mortgages would be securitised, skin-in-the-game provisions fail to generate the incentive for some banks to align mortgage prices to correlation risk (Fuster et al. (2022) and Krahnen and Wilde (2022)). Such correlation-price connection may also not be justified as Freddie Mac (Federal Home Loan Mortgage Corporation) and other agencies combine in the same securitised transaction mortgages from different banks. This potentially increases diversification of the underlying pool of loans relative to diversification in originators portfolio. Nonetheless, correlation patterns need to be considered in the design of residential mortgage-backed securities (RMBS) pools because, in such cases, correlation risk is ultimately transferred to investors without being adequately priced.

A corollary of the above is that the sensitivity toward correlation risk exhibits a significant variation amongst banks. Our findings indicate that borrowers have the potential to save an average $13,688 on a standard mortgage by "shopping around". This is attributed to the fact that lenders may apply varying interest rates to the same customer, not only to account for variations in how they consider correlation risk but also due to differences in their mortgage portfolio composition and the marginal impact of a new borrower on concentration in the bank's portfolio.

Although there is extensive literature on correlation (Adams et al. (2017), Driessen et al. (2009), Longin and Solnik (2001), Chernih et al. (2006), Gordy (2000), Blumke (2018)), few studies have examined correlations in retail portfolios. The absence of market prices for this asset class means that correlations among mortgages at the loan level must be calculated using default/loss data. Cowan and Cowan (2004) were the first to adopt this approach. We build upon their analysis by utilizing a more comprehensive database that covers the period of the Great Recession and by employing an alternative methodology. Our dataset includes 25 million mortgages issued between 1999 and 2017 across the United States. The data is publicly available through Freddie Mac.

The paper is organized as follows. In Section 2, we provide a review of the relevant literature. Section 3 presents a description of the data. Section 4 outlines the methodology employed. In Section 5, we discuss our findings. Finally, in Section 6, we present the conclusion of the paper.

2 Literature Review

This section provides a comprehensive review of the literature on mortgage correlation, beginning with the broader research in the corporate loan market. Additionally, we discuss existing studies that have identified specific mortgage features as triggers for default contagion.

The hit of Great Financial Crisis raised questions on the validity of correlation values and on the methodological assumptions set by either BCBS (2005) or alternative risk assessment frameworks. Literature and studies on this topic has grown bigger, with a particular focus on corporate asset classes or securities, leading to a widespread consensus on the lack of understanding of correlation risk (Nickerson and Griffin (2017), Chamizo et al. (2019), Chernih et al. (2006), Adams et al. (2017)). Nickerson and Griffin (2017) revise the assessment of default correlation for structured portfolios, finding that even estimating their model on pre-crisis data, the correlations used by rating agencies for CLOs were lower than those obtained by their model. Additionally, the authors argue that a commonly assumed lesson from the financial crisis is the lack of understanding of default correlations, and despite a significant period of massive defaults, limited academic work has been carried out to understand default correlations for structured products. Similarly, Chamizo et al. (2019) points out that a deficient modelling of correlation under stress could have been the cause of the failure of pre-crisis stress tests to detect the vulnerabilities of the financial system. A comprehensive work was also done by Chernih et al. (2006), who compare asset correlations calculated on monthly asset value with both Basel II and previous literature. The authors find that their results align with previous literature, while a notable discrepancy emerges when compared with Basel II and major software providers. Adams et al. (2017) explore correlation breaks among daily returns and argue that correlations are constant over time, but financial shocks lead to breaks that cause a shift in correlation level. All these studies highlight the necessity to better explore the role of correlation across different asset classes, as the Great Financial Crisis highlighted a flaw in risk assessment frameworks to correctly measure contagion effect. Nonetheless, mortgage correlation studies are quite limited in the current literature despite the relevance of this asset class in banking books and securitised markets.

The majority of the cited literature focuses on corporate portfolios, whereas little investigation has been conducted on mortgages. In general, a common misconception regarding residential mortgages correlation is that it remains relatively stable, leading to the assumption that the value set by BCBS (2005) can be applied to any capital calculation, including internal capital allocation. Assuming that conservativism is well-proven (despite being questioned by Hull (2015)) there is not much evidence in previous studies about the flat nature of correlation for residential mortgages. This is also attributed to the challenge of measuring mortgage correlation, since its asset value cannot be directly quantified. A study of Duellmann et al. (2010), for example, examines if it is better to estimate asset correlation from stock prices or default rates. The authors conclude that when market price time series are available, it is advisable to utilize stock prices rather than default rates, as the latter tend to underestimate and are often characterised by scarce data. However, only one option is possible for mortgage exposures (i.e. rely on default data).

Studies that explore mortgage correlation generally employ this approach, primarily focusing on assessing the adequacy of Basel assumptions. As mentioned previously, correlation value set in the regulatory framework is a flat value equal to 15% for residential mortgages (this value is derived by Calem and Follain (2003)) and to 4% for credit cards, consistent with BCBS guidelines (BCBS, 2021). A significant portion of the existing literature is dedicated to testing the accuracy of these values, and it often establishes that they are relatively conservative. For example, Botha and van Vuuren (2010) studies charge-off information loss data derived from the 100 largest US banks, while Crook and Bellotti (2009) analyses UK credit cards, finding consistent results with a piece of research from Rösch and Scheule (2004) on US credit cards. Geidosch (2014) examines correlation of residential mortgages using RMBS data, even including toxic RMBS deals. The author employs different estimation methodologies (SFGC, methods of moments, maximum likelihood estimation, parametric approach) and once again finds that the inferred correlation is remarkably low compared with Basel parameter, even after incorporating extremely low-quality deals. On the other hand, Neumann (2018) uses UK and US loss data to infer residential mortgage correlation using multiple estimators and instead conclude that Basel 15% parameter is at the appropriate level. Unlikely Geidosch (2014) and Neumann (2018), our methodology relies on popular copula models to derive correlations from default data (as in Lee et al. (2021)). Despite the criticism of these models in some literature (Egami and Kevkhishvili (2017)), we acknowledge their limitation in computing correlation. Nevertheless, we use them to extract a correlation indicator that shows heterogeneity and sensitivity to portfolio composition.

The global financial crisis has significantly impacted mortgage performance, serving as a catalyst for further research. A substantial amount of academic literature following the crisis has emphasized the importance of specific characteristics in explaining not only the increase in mortgage delinquency but also default contagion during economic downturns, which we eventually quantify through correlation.

A set of studies (Gupta and Hansman (2022), Goodstein et al. (2017) and Guiso et al. (2013)) analyse borrowers' choice and focus on strategic default determinants and its clustering. Guiso et al. (2013) and Gupta and Hansman (2022) find an important connection between leverage and default. In particular, Gupta and Hansman (2022) investigate defaulting behaviour of highly leveraged borrowers when house prices fall, separating moral hazard (i.e., leverage increases the probability of default) from adverse selection (i.e. risky borrowers prefer high-leverage mortgages). Although we cannot separate these two triggers, we also emphasize the effect of Updated LTV on default contagion. Additionally, we also corroborate the relevance of other factors, including as Balance, Income and FICO scores. Moreover, while only non-agency options ARMs are deployed in Gupta and Hansman (2022), we utilize a more representative sample of the US mortgage market. Similarly, Goodstein et al. (2017) analyse contagion effect among strategic defaulters resulting from increasing delinquency within the same ZIP code area. Once again, negative equity is identified as a significant driver, similar to the findings in Gupta and Hansman (2022). However, in this case, Goodstein et al. (2017) explicitly isolate strategic defaulters. Likewise, we investigate mortgage contagion implied by default experience, albeit without focusing on strategic and not-strategic behaviour, as we adopt the lenders' perspective that is blind to this aspect. Moreover, we incorporate a broader range of covariates to estimate correlation simultaneously. For instance, we also discover that borrowers in non-recourse exhibit higher sensitivity to economic shocks, aligning with the conclusions in Ghent and Kudlyak (2011).

In their attempt to explain the reasons behind default increase during the Global Financial Crisis (GFC), Mian and Sufi (2009) and Arentsen et al. (2015) establish the role of heightened lending to high-risk borrowers as a primary trigger. While Arentsen et al. (2015) attribute the surge of subprime mortgage defaults to the increased issuance of CDS, Mian and Sufi (2009) link the surge in mortgage defaults to disproportionate lending in subprime ZIP code areas. Both findings support the conception that 2009 economic downturn can be also explained by credit expansion to risky borrowers. Our research expands this result by pointing out that current regulation could have generated an incentive to increase banks' portfolio correlation (and risk) to make more efficient use of capital, by expanding credit allowance to risky borrowers while remaining compliant to international standards. In contrast to Mian and Sufi (2009), who aggregate default rates by Zip codes (similar to Goodstein et al. (2017)), we refrain from any data aggregation and, instead, preserve the unique combination of mortgage characteristics at borrower level. Secondly, the authors do not quantify the difference/discrepancy conditional on other drivers, which instead we address in our study by estimating correlation patterns.

Additional research has examined the dependence of correlation on firm characteristics. For instance, Lopez (2004) investigates the empirical relationship between average asset correlation, a firm's probability of default, and asset size. Although their focus is on the corporate sector, their findings are relevant to our study. The empirical findings indicate that the average asset correlation increases with asset size. In other words, as firms increase the book value of their assets, the correlation with the economic environment also increases. Although our research focus is different, we demonstrate that mortgages with larger balances are more sensitive to the systemic risk factor and experience higher contagion. Similar findings are reported by Duellmann and Scheule (2003), who explore asset correlation and its dependence on firm size and probability of default, finding a significant relationship with both factors.

3 Data

In 2021, within the US commercial banking sector, residential mortgages accounted for 23.01% of total the assets, evenly distributed between mortgage-backed securities (12.6%) and residential real estate loans (10.4%), totalling 5.27 trillion dollars (of Governors of the Federal Reserve System Data (2023)). However, the size of US residential mortgage market stretches well beyond the numbers just reported, as the largest part of originated residential mortgages is then securitised and sold to Government Sponsored Enterprises (GSE) like Freddie Mac, Fannie Mae and Ginnie Mae (66%, according to Fuster et al. (2022) and Banking Strategist (2022)). Overall, US single family residential mortgage market volume was close to $13 trillion in Q3 2022 (Banking Strategist (2022)).

This study employs loan-level and borrower-level data on 25 million fully amortizing fixed-rate, single-family mortgages. The dataset includes mortgages originating from the first quarter of 1999 through the end of 2017. These mortgages were issued by over 100 lenders and subsequently acquired by Freddie Mac for securitization purposes. The active and default statuses of the loans are tracked until the second quarter of 2018. Consistent with the demographic distribution in the United States, states such as California (with over 3 million mortgages), Florida, Texas, and Illinois (each with over 1 million mortgages) have a larger representation within the sample (Figure 1).

Data on both origination and performance is collected for each mortgage. Origination data includes borrower-, property- and mortgage-related characteristic measured at time of issuance. Table 1 presents the distribution of selected variables, including *Credit Score*, *Original Loan-to-Value*, *Debt-to-Income*, *Interest Rate* and *Balance*. The *Credit Score* is the FICO score, ranging from 300 to 850, with higher scores indicating a lower expected default rate. Scores below 669 are typically associated with a subprime status. Scores below 669 are typically indicative of a subprime status. *Original Loan-to-Value* (LTV) is calculated as the ratio of the original mortgage loan amount to the appraised value of the property at the time of purchase, ranging from 6% to 105%. The *Debt-to-Income* (DTI) ratio represents the sum of the borrower's monthly debt payments, including housing expenses related to the underwritten mortgage, divided by the total monthly income used to underwrite the loan. DTI ranges from 0% to 65% with 65%, with higher values indicating a higher debt burden relative to income. The introduction of stricter underwriting standards following the Great Recession is evident in the average increase/decrease of *Credit Score* and *Debt-to-Income*, respectively. This structural break in eligibility criteria is supported by previous studies (see Furfine (2020), Floros and White (2016)). Similarly, the average *Loan-to-Value* experienced a decrease after 2009 during the Great Recession, but there has been a recent reversal in this trend, primarily due to the implementation of support schemes for homebuyers.

Table A1 shows that the majority of borrowers purchase primary residences, while a smaller proportion buy investment or second homes. In contrast, the *Loan Purpose* exhibits an interesting increase in refinance mortgages immediately after the Great Recession, which can be attributed to the declining interest rate environment. On the other hand, the *Channel* variable experiences a significant decline in *Third-Party-Originations* (TPOs) [1], due to enhanced transparency and stricter reporting criteria mandated by Freddie Mac after the crisis. With the exception of *Property Type*, which shows an increasing share in the *Planned Unit Development* (PUD) segment[2], all other mortgage characteristics are evenly distributed based on the year of origination.

Within every quarterly vintage cohort, loans performance is monitored with monthly frequency since origination date. *Delinquency Status*, *Interest Rate* and *Unpaid Bal-*

[1]The Channel field is set to the data value of "TPO" (i.e., Third Party Originator Not Specified) for all loans which do not specify whether they are Broker ("B"), Correspondent ("C"), or Retail ("R"). Note that prior to 2008, Freddie Mac did not collect granular information on the types of origination channels. In 2008, Freddie Mac began collecting the granular information necessary to disclose whether a Broker or Correspondent was involved in the origination of each loan Federal Home Loan Mortgage Corporation (FHLMC (2022)

[2] A Planned Unit Development (PUD) is a real estate project in which each unit owner holds title to a lot and the improvements on the lot, and the home-owners association holds title to the Common Elements. The unit owners have a right to the use of the Common Elements and pay a fee to the home-owners association to maintain the Common Elements for their benefit. See Mandelker (2018) and David (2015)

ance are regularly updated throughout the entire loan lifetime. The availability of performance variables helps us to determine the evolution of each mortgage's credit performance and collateral information. For example, by knowing *Property State* (i.e. the state or territory where the property securing the mortgage is located) we can track the changes in state-level House Price Index and thus derive *Updated Loan-to-Value* from *Updated Appraisal Value* and outstanding *Unpaid balance* [3]. Likewise, we can calculate the *Loan Age* and follow mortgage lifecycle from origination to the latest available observation.

Amongst performance variables, repayment information is crucial in determining the default status of the mortgage. Two indicators are available to monitor the repayment performance of each loan. The first indicator is *Zero Balance Code*, which marks the reason why the loan balance has been reduced to zero, including charge-off, real estate owned (REO) acquisition [4], repurchase prior to property disposition and third-party sale. The second indicator is *Delinquency Status*, which refers to the number of days a borrower has been delinquent. Both variables are deployed to identify high-risk customers and trigger default status at the first occurrence of either 90-days delinquency or *Zero Balance Code* being populated. This aligns with the recently updated regulatory definition of default (for International Settlements (BIS)). We consider the occurrence of first default as an absorbing state, and thus, we exclude any observations after the initial default occurs. Figure 2 and Figure 3 depict two complementary aspects of the evolution of mortgage defaults during our sample period. Figure 2 shows that the peak in defaults occurs after the onset of the Great Recession. Therefore, we identify the mortgage crisis period as the years from 2009 to 2011 that capture the bulk of default events. Figure 3 displays the number of mortgages by the year of origination, highlighting that mortgage originated just before the crisis are more prone to default. This phenomenon is a result of the combined impact of the Great Recession

[3] While *Original Loan-to-Value* is the ratio between original loan amount on the note date and mortgaged property's purchase price, *Updated Loan-to-Value* is the ratio between outstanding balance at time t and updated appraisal value, where this latter is calculated based on state-level change in house prices from origination to observation date t.

[4] Real Estate Owned (REO) acquisition refers to foreclosed properties that are owned by the lender and were not sold at an auction

and the natural lifecycle of mortgages, characterised by a hump-shaped pattern with default rates peaking within the first 5 years from origination. Both of these factors are controlled for in our models. Overall, based on the default definition outlined earlier, 4.68% of the mortgages in our sample experienced default during the observation window.

Tables 2 through A4 provide a breakdown of the annual foreclosure rates based on borrower and loan characteristics. The default rates exhibit an inverse relationship with the *Credit Score*, with subprime borrowers (scores below 669) being approximately 20 times riskier than super-prime borrowers. The default rates for *Original Loan-to-Value* and *Updated Loan-to-Value* (Table A3) align with economic intuition, showing an increase in delinquency rates as leverage increases. While *Original Loan-to-Value* is a static field, meaning that accounts within a particular bucket do not migrate, *Updated Loan-to-Value* is dynamic. This means that accounts belonging to bucket i at time t can migrate to bucket j at $t+1$, depending on the ratio between the amortized balance and the updated appraisal value. The updated appraisal value is influenced by variations in the House Price Index at the state level, while the outstanding balance follows the amortization schedule. Lastly, Table A4 breaks down the default rate based on the main categorical variables, revealing variability in default rates across different segments that will be appropriately controlled for in our estimation.

In addition to ranking default rates based on the implied risk of variables segmentation, it is interesting to observe the heterogeneous change in delinquency rates between the *Long-Run* and the *Crisis* periods. This is one of the main objectives of our study, and it is clearly illustrated in Figure 4, which presents the ratio of average yearly default rate before and during the GFC for each state. While most states experienced a twofold increase in default rates, states like California, Nevada, Florida, Arizona witnessed a sixfold default rate rise during the *Crisis* compared to the *Long-Run*. Notably, among these states, only California and Arizona are non-recourse states, suggesting that strategic defaulters may not be the primary factor contributing to the significant change in defaults observed in the Great Recession (see Ghent and Kudlyak (2011) and Guiso et al. (2013)).

In contrast to Cowan and Cowan (2004), who focused on a single subprime lender, our study utilizes a larger dataset encompassing loans originated from multiple lenders. TTo evaluate the representativeness of our sample in relation to the US market, we compare it with data from the Home Mortgage Disclosure Act (HMDA) (Consumer Finance Protection Bureau (CFPB (2022)). The HMDA database is the most comprehensive source of publicly available information on the U.S. mortgage market. Enacted by Congress in 1975, the Home Mortgage Disclosure Act (HMDA) requires many financial institutions to maintain, report, and publicly disclose loan-level information about mortgage applications. Although HMDA data does not provide complete coverage of the US mortgage market, it remains the most extensive publicly available source of loan-level mortgage data. Table A5 displays the number of applications and originated loans over time. Of the 187 million mortgage applications received from 2007 to 2017, 48.1% resulted in originated mortgages. The majority of these applications correspond to Conventional loans (69.1%), which are the most common loan type in the US mortgage market. Conventional mortgages are not directly insured by the US Government, unlike FHA-insured [5], FSA/RHS-guaranteed [6], and VA-guaranteed mortgages [7]. Instead, they are retained on banks' balance sheets or acquired by GSEs (Freddie Mac and Fannie Mae), which are the primary participants in this segment.

[5]A Federal Housing Administration (FHA) loan is a home mortgage that is insured by the government and issued by a bank or other lender that is approved by the agency. FHA loans require a lower minimum down payment than many conventional loans, and applicants may have lower credit scores than is usually required. The FHA loan is designed to help low- to moderate-income families attain home-ownership. They are particularly popular with first-time homebuyers.

[6]FSA/RHS loans are a type of financing provided or guaranteed by the Farm Service Agency (FSA)/Rural Housing Service (RHS) of the U.S. Department of Agriculture (USDA). FSA provides direct and guaranteed farm loans for farmers and ranchers of all kinds. RHS lends directly to low-income borrowers in rural areas and guarantees loans issued by approved lenders that meet RHS requirements.

[7]VA-guaranteed mortgages are loans available through a program established by the U.S. Department of Veterans Affairs (VA) (previously the Veterans Administration). With VA loans, veterans, service members, and their surviving spouses can purchase homes with little to no down payment and no private mortgage insurance and generally get a competitive interest rate.

Fannie Mae and Freddie Mac primarily acquire conventional loans not insured by the government (46.1%) and establish guidelines (conformity rules) that depository and non-depository lenders must adhere to when securitizing loans under GSEs. Conformity rules require loan size, minimum credit score, down-payment, debt-to-income ratios, mortgage insurance to be within specific ranges, even though there are lot of exceptions and compensating factors whenever some criteria are not met. While the conformity rules established by Freddie Mac and Fannie Mae do not completely overlap, they significantly impact the acceptance/rejection mechanism of mortgage applications in the broader mortgage market. Although there is no explicit market division between Fannie Mae and Freddie Mac, it is well-known that historically, Freddie Mac has targeted smaller banks and thrifts, while Fannie Mae has predominantly acquired mortgages from larger commercial banks. However, the post-Great Recession mortgage market witnessed numerous mergers and acquisitions among lenders, blurring the boundaries between the originators served by each agency. While Fannie Mae has a larger volume of mortgages compared to Freddie Mac, Table A5 illustrates that Freddie Mac still maintains a significant share of approximately 25% for conventional loans, which is noteworthy within the scope of our study on the US mortgage market.

4 Empirical Methodology

4.1 Correlation

Our loan-level estimates are derived using on a panel-logit discrete hazard model, which allows us to calculate the long-run ($PD_{LongRun}$) and a downturn PDs (PD_{Crisis}) for each loan. These *PDs* and then used to compute correlation. We employ annual data so that the model produces 12-Month *PDs* that can be directly used to extract implied correlations from the internal rating-based approach of current bank capital regulations model (BCBS (2005)). The performance of each loan is tracked annually, and a binary 0/1 dependent variable is computed each year to flag default based on loan's delinquency at the end of the respective year. Default is triggered according to the previously defined definition. The explanatory variables for each loan include time-invariant characteristics at origination (e.g. *Credit Score*, *Purpose*, *Region*) and time-varying characteristics (e.g. *Loan Age*, *Updated Loan-to-Value*), as well as and macroeconomic variable. Only *Credit Score* at origination is available. Once the panel loan-level dataset is built, multi-period logit model [8] is estimated as per Equation 1.

$$PD_{W_{it}} = \frac{1}{1 + exp(-W_{it})} \tag{1}$$

with

$$\begin{aligned} W_{it} = \alpha + \sum_{b}^{N} \beta_b LoanCharacteristics_{b,it} + \sum_{z}^{Z} \zeta_z Macro_{z,t} + \gamma Crisis_t + \\ \sum_{b}^{N} \delta_b Crisis_t \times LoanCharacteristics_{b,it} \end{aligned} \tag{2}$$

Where α is the intercept, β_b in $(1, \cdots, N)$ capture mortgage characteristics throughout the entire development sample period, γ captures the crisis effect on overall default rates, while instead δ_b in $(1, \cdots, N)$ measure mortgage characteristics behaviour in response to the crisis. The same mortgage features are used during non-crisis and crisis. In doing so, we are able to separate the downturn predicted *PDs* (PD_{Crisis}) from long-run values ($PD_{LongRun}$). The coefficients ζ_z capture the effect of macroeconomic trends. The dummy $Crisis_t$ is activated for the years running from 2009 to 2011 included, as we have observed that the effect of financial crisis on the mortgage market was not immediate. In the regressions we do not use *Channel*, as it is not available consistently throughout the sample period, nor *Original Interest Rate* due to its non-stationary trend over the observation window. All regression models are validated based on a set of criteria which includes rank-ordering measured by GINI and AUROC coefficients (See Yang et al. (2023) and Zeng and Zeng (2019)). PD_{Crisis} and a $PD_{LongRun}$ are estimated by switching on and off, respectively, the crisis dummy. Then, for each combination of mortgage characteristics b in $(1, \cdots, N)$, we are able to feed PD_{Crisis} and a $PD_{LongRun}$ into the asymptotic single risk factor model (ASRF) used by regulators that links PD_{Crisis}, $PD_{LongRun}$ and correlation as in Equation 3:

[8]Multiperiod logit/probit models are often deployed for default estimation. In fact, a similar model specification is adopted by Arentsen et al. (2015), Ghent and Kudlyak (2011) and Lee et al. (2021)

$$PD_{Crisis,i} = \varphi\left(\frac{\varphi^{-1}(PD_{LongRun,i}) + \varphi^{-1}(0.999)\sqrt{\rho}}{\sqrt{1-\rho}}\right) \tag{3}$$

where $\varphi(x)$ denotes the cumulative distribution function for a standard normal random variable, $PD_{Crisis,i}$ is the downturn PD for mortgage i, while $PD_{LongRun,i}$ is the long-run PD for mortgage combination i. Once $PD_{Crisis,i}$ and a $PD_{LongRun,i}$ are estimated, the relevant correlation can be obtained by inverting Equation 3 numerically.

4.2 Excess Interest Rate

The second step in our study involves defining Excess Interest Rate at the mortgage level. We aim to investigate whether and how lenders incorporate the non-flat correlation ρ_i when pricing newly issued mortgages. To ensure independence between the sample used for estimating the correlation and the sample used for Excess Interest Rate model, we calculate correlation ρ from a model run on a reduced sample comprising mortgages originated up to 2011 (inclusive). Conversely, Excess Interest rate model is estimated using the remaining data, which includes mortgages originated immediately after 2011.

Hence, Excess interest Rate δ is defined as the difference between each mortgage *Original Interest Rate* and the average *Original Interest Rate* of all loans issued in the same vintage. Excess Interest Rate can be referred as δ_i for each mortgage i in $(1, \cdots, N)$. Two approaches have been tested to calculate δ_i, as outlined in Equation 4 and Equation 5

$$\delta_i = OriginalIR_i - \frac{\sum_{j=1}^{N_J^{T_Q}} OriginalIR_j}{N_J^{T_Q}} \tag{4}$$

$$\delta_i = OriginalIR_i - \frac{\sum_{j=1}^{N_J^{T_Y}} OriginalIR_j}{N_J^{T_Y}} \tag{5}$$

where Equation 4 calculates δ_i as the difference with quarterly average *Original Interest Rate* and Equation 5 calculates δ_i as the difference with yearly average *Original*

Interest Rate. By adopting this definition, we want to isolate the pure effect of the premium charged by lenders relative to the average portfolio average rate, without any noise coming from interest rate trends. We rely on Equation 4 to calculate Excess Interest rate, as it yields more accurate results. We do not introduce any further spacial segmentation following Hurst et al. (2016) findings, who establish that despite large regional variation in predictable default risk, GSE mortgage rates for otherwise identical loans do not vary spatially. Spread over 30Yr-Freddie Mac mortgage rate is not considered because it might also reflect non-stationary trends in the economy. Nevertheless, the large sample size deployed and the wide coverage makes us confident of the approach used.

Excess Interest Rate δ_i is then linearly regressed against the explanatory drivers, which also include correlation ρ_i as in Equation 6:

$$\begin{aligned}\delta_i = \alpha + \sum_b^N \beta_b LoanCharacteristics_{b,i} + \sum_{m=1}^N \mu_m Macro_{m,t} + \\ \omega \times \rho_i + \sum_f^N \varphi_f Bank_{f,i} + \sum_p^N \psi_p Bank_{p,i} \times \rho_i + E_i\end{aligned} \tag{6}$$

where δ_i is the estimated Excess Interest Rate at origination for mortgage i in $(1, \cdots, N)$. Besides the intercept α, β_i coefficients (i in $1, \cdots, N$) capture loan characteristics at origination, ensuring correlation impact is not biased by omitting mortgage drivers; μ_m (m in $1, \cdots, N$) capture the state of economy at time of origination. The rest of the equation includes bank and correlation coefficients; φ_f (f in $1, \cdots, N$) are bank fixed-effects, ω measures the pure effect of correlation ρ, while ψ_p (p in $1, \cdots, N$) captures the interaction between correlation and banks. Lastly, E_i is the error term.

As Equation 6 introduces an interaction term between correlation ρ and bank-fixed effects, we also test a second set of regressions by separating the sample by banks and by estimating each model separately as in Equation 7:

$$\delta_{ij} = \alpha + \sum_b^N \beta_b LoanCharacteristics_{b,i} + \sum_{m=1}^N \mu_m Macro_{m,t} + \omega_j \times \rho_{ij} + E_{ij} \tag{7}$$

where, differently from previous Equation 6, δ_{ij} is the Excess Interest Rate at origination for mortgage i in $(1, \cdots, N)$ issued by lender j in $(1, \cdots, N_j)$, while ω_j is the bank-specific correlation coefficient.

Excess interest rate models are estimated on a cross-sectional sample, considering each mortgage only once, specifically at the time of origination. To correctly incorporate correlation ρ, which has been estimated using a multiperiod logit model on panel data (Equation 1), a set of assumptions is made for time-varying variables such as *Updated Loan-To-Value*, *Loan Age*, and macro drivers. First, the *Updated Loan-To-Value* is substituted with the *Original Loan-To-Value*. Although this choice leads to lower correlations, further corrections, such as stressing the *Original Loan-To-Value*, are not preferred. The second assumption relates to *Loan Age*, where the optimal value is selected based on the observed peak in both PD_{Crisis} and $PD_{LongRun}$. Thirdly, macroeconomic variables are incorporated by calculating Long-Run and downturn averages at the state level. All models undergo validation based on a set of criteria, including goodness-of-fit measured by R^2, robust standard errors, the correctness of coefficient signs, and the stability of estimates across different interactions.

5 Results

The primary finding of our study is that the correlation in residential mortgage portfolios exhibits a non-flat pattern, with significant variability observed across different mortgage characteristics. Additionally, we demonstrate that, in the aftermath of Great Financial Crisis, lending institutions have priced correlation differently, potentially influenced by regulatory compliance biases. In the following sections, we illustrate how our results support all these claims.

5.1 Correlation

The starting point before drawing any conclusion on correlation patterns is the correct assessment of PD_{Crisis} and $PD_{LongRun}$, as estimated by multi-period logistic model introduced in the empirical methodology. Selection criteria have been broadly defined in the methodology section. We now discuss in detail the steps undertaken to obtain the final model (*Model 5*) presented in Table 3.

Since logistic regression establishes a non-linear relationship between model drivers and target variable through the interacted Dummy *Crisis*, interpreting the sign of interacted terms from raw regression outputs can be difficult (Ai and Norton (2003)). To overcome this, we compute average marginal effects to clearly unfold the interaction effect and the contribution of each driver on the target variable. Initially, this simple test guided us to eliminate any non-stationary variables like *Original Interest Rate* and *Channel*, that were otherwise yielding counter-intuitive results. Second, the use of marginal effects helps to verify economic significance of explanatory drivers. Table 3 shows the stepped approach adopted, which we are now going to discuss.

The first set of models (*Model 1* and *Model 1a)* only incorporates static variables, i.e. measured at origination and not changing over time. Having as a reference the observed default rates presented in Table 2, economic significance of *Credit Score*, *Debt-to-Income* and *Excess Interest Rate* on default probabilities is verified [9]. The average marginal effects (and therefore, the underlying model coefficients) remain stable even in the subsequent model specifications, although they decrease due to the inclusion of additional factors. *Model 1* is further augmented with Dummy *Crisis*, as shown in column *Model 1a*, where we observe that the switch from 0 to 1 yields an increase of 1.5% in yearly default probability, as empirically observed in Table 2.

However, although origination variables as *Credit Score*, *Debt-to-Income* are key in the rank ordering of risk, mortgage default is also influenced by changing factors over time. As a second step, *Loan Age* and *Balance* are therefore included in the model, notably improving rank ordering compared to initial estimations (*Model 2*). Given

[9] A decrease of *Credit Score* from 799 to 739 yields an increment in default probability of 65 bps, in line with average default rates observed in Table 2. An increase of *Debt-to-Income* from 40 to 55 yields an increment in default probability of 42 bps. Finally, an increase of *Excess Interest Rate* from 0 to 0.5 yields an increment in default probability of 26 bps

the non-linear relationship between *Loan Age* and Default rate, dummy variables are deployed to capture loan age in years. This choice is preferred over using splines or first/second order variables due to frequency of observations, which does not grant sufficient granularity. On the other hand, *Balance* is transformed using a natural logarithm function and enters the model with a positive, significant effect. The strength of Dummy *Crisis* in *Model 2a* slightly decreases because mortgage lifecycle peak partly overlaps with the years of great financial crisis for many loans in the sample. Thus, to separate the effect of mortgage lifecycle from *Crisis*, the correct relationship is eventually re-established by interacting *Loan Age* with Dummy *Crisis* in the final *Model 4*.

A third key component in modelling yearly default rate is the economic cycle. Estimations are progressively enhanced by including yearly change in Unemployment Rate at State level (Ump_{12}) and *Updated Loan-to-Value*, which is a function of State-level House Prices. Both variables are statistically and economically significant [10]. Economic-cycle dependent variables expectedly decrease Dummy *Crisis* marginal contribution on yearly default rate (*Model 3a*), as they also contribute to explain GFC effect on mortgage defaults.

At last, *Model 4* interacts all previous model drivers with Dummy *Crisis*, according to our empirical methodology to calculate correlations. First, we verify that the marginal effect of each driver is always greater in *Crisis* than in *Long-Run*. Again, this is proved in Table 3. Figure 6 further breaks-down average marginal effects at different levels of *Credit Score*, *Updated Loan-to-Value*, *Debt-to-Income*, *Balance* and *Excess Interest Rate* on default probabilities. In all cases, the variable increase (decrease for *Credit Score*) bears a higher impact on probability of default during *Crisis* compared with *Long-Run*. In line with expectations, we are reassured that the model correctly captures the contribution of Dummy *Crisis*. The interaction with Dummy *Crisis* also helps to better separate mortgage dynamics from *Long-Run* to *Crisis*. For example, *Balance* gains its correct marginal contribution after is made interacted

[10] An increase of Ump_{12} from 1 to 2 yields an average increment in default probability of 12 bps, while an increase of 10% (e.g. from 60 to 70) for *Updated Loan-to-Value* yields an increment in default probability of 34 bps.

with Dummy *Crisis* (see Figure 5), which would have been otherwise neglected.

Robust standard errors are calculated and presented in Table 3. Moreover, as the sample is a panel dataset with repeated observations for each mortgage, clustered standard errors are also computed by grouping on loan identifier. Even in this case, model coefficients remain highly significant. In addition to economic soundness and significance of the estimates, rank ordering and predictive power are achieved with high levels of AUROC (87.97%) and Gini (75.94%). We are then confident that the model underpinning correlation inference is correctly specified.

Development sample data is then deployed to assess in-sample correlations. For each mortgage, $PD_{LongRun}$ and PD_{Crisis} are calculated based on loan's characteristics. Correlation ρ_i is then computed numerically for each data point by minimizing the quadratic difference between PD_{Crisis} and the transformation of $PD_{LongRun}$ (see Equation 3), where the only unknown is correlation ρ. We obtain a complete distribution of correlations ρ_i, whose variability is driven by the unique combination of each mortgage i features within the portfolio. Figure 7 displays ρ distribution by Balance and Lenders, while Table 4 complements the analysis with additional statistics.

The non-flat nature of correlation ρ, fully ranging from 0% to a maximum value of 13.07%, and the diverse sensitivity to mortgage characteristics are remarkably evident. Despite its variability, though, we observe that correlation never exceeds the 15% value set by BCBS (2005) (at least in our sample). This is in line with previous research from Chamizo et al. (2019), Chernih et al. (2010) and Botha and van Vuuren (2010). Therefore, the benchmark set by BCBS (2005) proves once again to be sufficiently conservative even when deploying data covering GFC, regardless of the doubts expressed by Hull (2015). Such level of conservativism opens up to additional considerations that will be expanded in the second part of the results section, when touching on Excess Interest Rate Analysis. Table 4 reports average, standard deviation and upper quantiles of correlation distribution by most relevant mortgage features. Average correlations in Freddie Mac seem aligned with Cowan and Cowan (2004), even if slightly higher on average for those drivers that at least we both consider in our analysis. Although the variation in data and methodology may explain the differences, it is reassuring that our results are consistent with existing literature. However, unlike Cowan and Cowan (2004) who examine each dimension separately by deriving correlation from aggregated time series, our methodology incorporate all mortgage dimension simultaneously. This enables us to effectively observe trends that are specifically driven by each variable, eliminating potential bias from omitted controls. Furthermore, we examine correlation distribution for additional characteristics (e.g. *Updated Loan-to-Value*, *Balance*, *Seller*).

Figure 7 shows that mortgage correlation is proportional to loan balance. A standard deviation increase yields 39% positive change in correlation. This result is particularly significant because it draws a parallelism with correlation firm-size adjustment for SME exposures set by BCBS (2021), where correlation is an increasing function of firm-size. Also Lopez (2004) detects that average asset correlation is an increasing function of asset size (which is generally used as a proxy for firm size). As firms increase the book value of their assets, they become more correlated with the general economic environment and the common factor. Although this has never been factored in residential mortgages, such outcome points out that a loan-size adjustment could be a sound approach for residential mortgages too, or at least worth considering. Somehow, it seems that under adverse economic conditions, borrowers with higher balances are more exposed to contagion effect, regardless of the equity held on the property. In fact, even though we find a positive relationship between *Updated Loan-to-Value* and correlation, such relationship is not as strong as for Balance (see Figure A3 and Table 4). A standard deviation increase in Loan-to-Value results in 1% upward change in correlation. Therefore, even though leverage clearly increases borrowers' default risk (see Gupta and Hansman (2022), Campbell and Cocco (2015), Elul et al. (2010)), the impact on correlation is not as sensitive as with Balance.

Credit Score exhibits a similar pattern to *Updated Loan-to-Value*. Figure A2 and Table 4 highlight that correlation variability is also influenced by the borrower's *Credit Score*, e although the effect is relatively homogeneous, similar to that observed with

Updated Loan-to-Value. In fact, a standard deviation change results in a 2% increase in correlation ρ. This suggests that borrowers' sensitivity to correlation during a downturn is fairly consistent across risk grades. However, a distinct hump-shaped trend is observed, with borrowers in the mid-range experiencing higher correlation compared to those in the low or high segments [11]. Although it is well established that the Credit Score ranks default risk (see Table 2 and Demyanyk et al. (2011)), this result suggests that in the midst of an economic downturn, borrowers with mid-range creditworthiness react more adversely to a single risk-factor, possibly due to the sway and weakening of available resources caused by financial instability (Adelino et al. (2016)).

In addition, *Debt-to-Income* provides valuable insights and proves to significantly impact the correlation in residential mortgages. Specifically, an increase in its standard deviation results in an 11% rise in correlation. Alongside *Balance*, *Debt-to-Income* emerges as one of the strongest determinants of contagion. This finding aligns with previous research (see Linn and Lyons (2020) and Quercia and Stegman (1992)) that emphasizes the influence of income on default. Furthermore, it suggests that not only foreclosures but also correlation ρ are positively associated with borrowers' debt relative to their income. This result is consistent with the findings of Cowan and Cowan (2004), although our observations reveal a clearer pattern.

Variability in correlation ρ is ascertained across other portfolio dimensions too. For example, Figure A3 shows variability across U.S. regions, where Far-West, Rocky Mountains, New England and Mid-East stand out over the other territories. This was already anticipated by observing the ratio between average yearly default rate during Great Financial Crisis (GFC) and average yearly default rate before the GFC in Figure 4, although it remains an interesting result that well aligns with research made by Hurst et al. (2016) and Mian and Sufi (2009). Similarly, discrepancy in contagion is discovered when splitting the sample by Recourse and Non-Recourse States (see Figure A4 and last row of Table 4). Borrowers in non-recourse states experience higher contagion, which is likely linked to state policies that prevent lenders to pursue borrower's other assets in the event of foreclosure. This finding contributes to the ex-

[11] This is also consistent with Cowan and Cowan (2004)

isting research conducted by Ghent and Kudlyak (2011), as we provide evidence that verifies the hypothesis they were unable to reject. Specifically, we demonstrate that borrowers in non-recourse states are not significantly prone to default under normal circumstances. However, they tend to cluster in adverse scenarios, even when their loans are securitized by government-sponsored enterprises (GSEs). Differently from Ghent and Kudlyak (2011), we control for a wider array of mortgage features. This result aligns with the existing stream of strategic default literature (Goodstein et al. (2017), Guiso et al. (2013)), even though our analysis does not aim at discriminating between strategic and non-strategic defaulters.

Finally, the correlation ρ among lenders exhibits clear variability, as depicted in Figure 7. However, it is challenging to establish the underlying causes of this phenomenon, as it may depend on risk appetite and lending strategies, which are beyond the scope of our assessment. Additionally, it is noteworthy that not all lending institutions exhibiting higher correlation experienced a significant surge in default rates, as illustrated in Figure A1, when compared to other banks with lower contagion. Nevertheless, it is crucial for the second phase of our study to consider the contribution of lenders to correlation since we aim to comprehend how they price this risk differently, or not at all.

5.2 Excess Interest Rate

We have ascertained that correlation is a non-flat value, and that it can be highly dependant on specific mortgage characteristics. We now assess how financial institutions evaluate correlation risk when pricing through-the-door mortgages, and whether variability in contagion might influence lenders' risk appetite. To achieve this objective, Excess Interest Rate is linearly regressed by the usual mortgage factors and segment-varying correlation ρ. Differently from panel-logit discrete hazard model, the frequency of observations is now quarterly, and Excess Interest rate is only measured at origination by design, as the sample is composed of fixed-rate mortgages. We are not anyway interested in interest rate resets, as our interest is to quantify if and how lenders price correlation risk at mortgage application.

Regression results are reported in Table 5, where Bank-specific effects have been progressively included in the estimations, alongside with correlation ρ and loan level factors. To keep independence between Correlation and Excess Interest Rate samples, the correlations in Excess Interest Rate regression are derived from the same model specification as in Table 3, albeit after excluding Excess Interest Rate as independent variable and including only mortgages originated up to December 2011. The sample used for Excess Interest Rate regression, instead, includes mortgage originated from January 2012. *Model 1* does not include any Bank-effect, *Model 2* incorporates Bank fixed effects only, and *Model 3* finally accounts for the interaction between correlation and Lenders. All other drivers are kept identical, and none is dropped throughout. As with panel-logit discrete hazard model, significance of the coefficients is ensured by robust standard errors, and goodness-of-fit is measured by R^2, $AdjustedR^2$ and AIC. Given the high number of observations (more than 7.3 million), R^2 and $Adjusted - R^2$ are almost identical.

First, it is ensured that the most common drivers for mortgage pricing are correctly estimated in terms of statistical and economic significance, reflecting the perceived risk. A *Credit Score* increase of 100 lowers Excess Interest Rate of 19 bps, while joint borrowers yield a 4 bps decrease compared to single borrowers. On the other hand, an *Original Loan-to-Value* increase by 10 determines an 6.3 bps increment, while the same increase in *Debt-to-Income* yields a 2.4 bps change. Macroeconomic drivers, used as a proxy of economic activity at time of origination, correctly measure mortgage premium when economy is expanding. Ump_{12} and HPI_{12} are significant and negative/positive respectively in all regressions. Mortgage level coefficients remain stable, significant and sound across all model specifications, despite the progressive inclusion of bank fixed-effects and interaction terms.

We now draw our attention to the role of correlation ρ on Excess Interest Rate. In *Model 1* and *Model 2* of Table 5, correlation is priced positively. Lenders price contagion effect upward because it is correctly perceived that as correlated segments bear higher risks, interest rates should be adjusted accordingly. While this finding is initially intuitive and reasonable, we are currently unable to determine whether lenders consistently exhibit this behaviour or if the results are primarily driven by the majority of observations in the "Other Sellers" category [12].

Model 3 sheds light on such behaviour and unfolds one of the most interesting results of our study. When making correlation ρ interacted with Lenders, the sign does not remain consistently positive, and instead the net effect is negative for the following lenders: Bank of America [13], JP Morgan Chase, Citi and Wells Fargo [14]. It is important to remind that not all these lenders exhibited the highest in-sample correlations, nor those having a positive coefficient necessarily belong to the low-correlation group.

Additionally, a second analysis is conducted and presented in Table 6 to further support the results. In this analysis, a linear regression is performed individually for each lender, utilizing the same set of predictors. The objective is to ascertain robustness of the findings when assessing the correlation impact on Excess Interest Rate for each lender independently. The results remain consistent, with correlation continuing to be negatively priced for the same lenders as reported in Table 5.

This finding prompts significant considerations, revealing one of the most important results of our paper. It has been already seen that in-sample correlations never exceed the regulatory benchmark set at 15%. At the same time, *Model 3* in Table 5 and in Table 6 shows that banks do not consistently assign the same weight to correlation. In fact, the most important lenders (Bank of America, JP Morgan Chase, Citi and Wells Fargo) price mortgage correlation negatively. These financial institutions

[12]According to Freddie Mac Federal Home Loan Mortgage Corporation (FHLMC (2022): *Seller Name will be disclosed for sellers with a total Original UPB representing 1% or more of the total Original UPB of all loans in the Dataset for a given calendar quarter. Otherwise, the Seller Name will be set to "Other Sellers"..* We also incorporate in this category those sellers we could not observe throughout the sample period, either because of merging or closure.

[13]Bank of America is the reference category, hence its correlation coefficient corresponds to the coefficient of ρ, which is negative and equal to -0.9464

[14] Whilst Wells Fargo coefficient is positive, the actual effects has to be calculated relative to the reference category. As Wells Fargo coefficient is lower in absolute value than the reference category, its impact is effectively negative.

are the only ones in our sample that belong to the list of Global Systemically Important Banks (G-SIBs) (Financial Stability Board (2022)), which need to comply with international capital requirements set by Basel accords. As such, this result highlights that implementation of Basel standards might perversely encourage banks in perilous lending, whether aware or not. As these institution need to implement the conservative 15% correlation value for capital reporting, they might be pushed to increase profitability by lending towards those correlated, yet profitable, segments, feeling anyway backed-up by regulatory compliance. We thus point out that current regulation could generate the incentive for banks to increase portfolio correlation (and risk) in order to make more efficient use of capital. Whilst we acknowledge that such claim can be strong and influenced by model error, it is also true that previous literature has already pointed out the rise of perilous lending to increase profitability, motivated by either market expectations (Mian and Sufi (2009)) or hedging (Arentsen et al. (2015)). Likewise, we add a new perspective in the risks posed by bad-usage of regulation in residential mortgage market and we reach similar conclusions, although by exploiting a pretty different framework.

The bias introduced by lenders in pricing correlation leads us to a second analysis. We shift focus to borrowers and examine whether consumers have an advantage in shopping around, considering the differential pricing of correlation. To investigate this, we employ a stylized example and select a reference mortgage as the basis for calculating the varying impact of correlation on total interest paid by different banks. The reference mortgage is a 30-Year Fixed-Rate Mortgage with an original balance of \$300,000 and an origination rate of 5.5%. By computing all possible combinations of attributes, including those of the top lenders listed in Table 5, we determine correlation ρ for each entry and calculate the marginal effect of the lender/correlation interaction on the Excess Interest rate. This value is then added to the reference original rate, ensuring that the only source of variability is the interaction between correlation ρ and each individual bank. Although correlation is inherently dependent on mortgage features at this stage, our interest lies in quantifying how banks, all else being equal, price it differently. To accomplish this, we aggregate the portfolio and determine the maximum difference in total interest paid among banks for each unique mortgage combination.

Figure 8 reports the distribution in total interests paid, highlighting a concentration below \$ 12,000 (median value) and a mean value of \$ 13,700. Considering the total interests paid by the borrower (c. \$ 310,000), this variation can help borrowers save around 4.41% of the entire interest amount, with a standard deviation that stretches to a saving percentage equal to 7.27%. The right tail of the distribution shows that excess interest payments stretch to a maximum of \$ 40,000 (19.32% of total interests). This implies that the different pricing of correlation by banks translates into significant difference in total interests paid, further exacerbated by mortgage characteristics. For example, Figure A5 shows that geographic patterns exist, especially where the lending is concentrated in Midwest and Rocky mountains. In these areas borrowers are given an incentive in shopping around, as total interests charged can differ substantially. The same finding does not apply to the differential impact by *Original Loan-to-Value* and *Credit Score*. In fact, Figure A6 and Figure A7 highlight that financial institutions price mortgage correlation ρ in a similar way.

The last case analysed relates to *Debt-to-Income*. Figure A8 highlights that banks price mortgage default correlation ρ for increasing *Debt-to-Income* at origination quite differently, especially for ratio greater than 30 %. The result is interesting, especially because the distribution always stretches towards tail values. Hence, while banks are quite conservative and consistent in pricing mortgage default correlation ρ for increasing *Original Loan-to-Value*, the same approach seems not to be followed for *Debt-to-Income*.

6 Conclusion

This paper investigated whether residential mortgages correlation is effectively a flat-value (as recommended by BCBS (2005)) or if instead it variates depending on specific mortgage characteristics. Through the use of a comprehensive sample obtained from Freddie Mac, which spans the Great Financial Crisis, we provide evidence that correlation variability is significantly influenced by mortgage attributes. Relying on average correlations may obscure the actual level of interconnection in mortgage contagion, which can manifest at the intersection of specific features beyond the well-known risk patterns. Notably, our study reveals that correlation variability is strongly influenced by loan balance and borrower Debt-to-Income, which have not been previously explored. Risk managers should be vigilant about concentration risks to mitigate the potential increase in losses arising from correlated segments that may emerge unexpectedly during severe economic conditions.

Following the demonstration of non-flat mortgage default correlations, our framework is deployed to understand how lending institutions price correlation. We thus demonstrate that not all banks price contagion risk in the same way. In particular, we ascertain that the Global Systemically Important Banks within our sample are the only ones that price correlation negatively. This finding is particularly important, as it points out that current regulation could generate the incentive to increase portfolio correlation (and risk) in order to make more efficient use of capital. Such negative premium might also be the result of the intense market competition that pushes interest down and disconnects lenders from any portfolio concentration consideration, as compliance is ensured in any case by regulatory correlation value.

Finally, after isolating the effect of correlation by banks on the excess interest rate, we measure on a stylised example how correlation risk is effectively priced by lenders, introducing an effective benefit for borrowers in shopping-around.

References

Adams, Z., Füss, R. and Gluck, T. (2017), 'Are correlations constant? empirical and theoretical results on popular correlation models in finance', *Journal of Banking & Finance* **84**, 9–24.
URL: *https://www.sciencedirect.com/science/article/pii/S0378426617301590*

Adelino, M., Schoar, A. and Severino, F. (2016), 'Loan originations and defaults in the mortgage crisis: The role of the middle class', *The Review of financial studies* **29**(7), 1635–1670.

Ai, C. and Norton, E. C. (2003), 'Interaction terms in logit and probit models', *Economics Letters* **80**(1), 123–129.
URL: *https://www.sciencedirect.com/science/article/pii/S0165176503000326*

Arentsen, E., Mauer, D. C., Rosenlund, B., Zhang, H. H. and Zhao, F. (2015), 'Subprime mortgage defaults and credit default swaps', *The Journal of Finance (New York)* **70**(2), 689–731.

Banking Strategist (2022), 'Mortgage finance sector'. (accessed October 5, 2022).
URL: *https://www.bankingstrategist.com/mortgage-finance-sector*

BCBS (2005), 'An explanatory note on the Basel II IRB risk weight functions', *Basel Committee on Banking Supervision, Bank for International Settlements* (2).

BCBS (2021), 'Calculation of rwa for credit risk.CRE31 IRB approach: Risk weight functions', *Basel Committee on Banking Supervision, Bank for International Settlements* (2).

Blumke, O. (2018), 'On the cyclicality of default rates of banks: A comparative study of the asset correlation and diversification effects', *Journal of Empirical Finance* **47**, 65–77.
URL: *https://www.sciencedirect.com/science/article/pii/S0927539818300252*

Botha, M. and van Vuuren, G. (2010), 'Implied asset correlation in retail loans portfolios', *Journal of Risk Management in Financial Institutions* **3**(2), 156–173.

Buraschi, A., Porchia, P. and Trojani, F. (n.d.), 'Correlation risk and optimal portfolio choice', *The Journal of Finance* **65**(1), 393–420.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2009.01533.x*

Byrne, D., Kelly, R. and O'Toole, C. (2022), 'How does monetary policy pass-through affect mortgage default? evidence from the irish mortgage market', *Journal of Money, Credit and Banking* **54**(7), 2081–2101.

Calem, P. and Follain, J. (2003), 'The asset correlation parameter in basel ii for mortgages on single-family residences', *Board of Governors of the Federal Reserve System* .

Campbell, J. Y. and Cocco, J. F. (2015), 'A model of mortgage default', *The Journal of finance (New York)* **70**(4), 1495–1554.

Chamizo, A., Fonollosa, A. and Novales, A. (2019), 'Forward-looking asset correlations in the estimation of economic capital', *Journal of International Financial Markets, Institutions and Money* **61**, 264–288.
URL: *https://www.sciencedirect.com/science/article/pii/S1042443118304888*

Chernih, A., Henrard, L. and Vanduffel, S. (2006), 'Asset correlations: A literature review and analysis of the impact of dependent loss given defaults', *Katholieke University Leuven* **48**(17), 1–15.

Chernih, A., Henrard, L. and Vanduffel, S. (2010), 'Reconciling credit correlations', *Journal of risk model validation* **4**(2), 47–64.

Consumer Finance Protection Bureau (CFPB (2022), 'Mortgage data (hmda)'. (accessed October 8, 2022).
URL: *https://www.consumerfinance.gov/data-research/hmda/*

Cowan, A. M. and Cowan, C. D. (2004), 'Default correlation: An empirical investigation of a subprime lender', *Journal of Banking and Finance* **28**(4), 753–771. Retail Credit Risk Management and Measurement.
URL: *https://www.sciencedirect.com/science/article/pii/S0378426603001985*

Crook, J. and Bellotti, A. (2009), 'Asset correlations for credit card defaults', *Applied Financial Economics* **22**.

David, N. P. (2015), 'Factors affecting planned unit development implementation', *Planning, Practice & Research* **30**(4), 393–409.

Davidson, A., Sanders, A., Lan-Ling, W. and Ching, A. (2004), *Securitization: Structuring and investment analysis.*, John Wiley and Sons.

Demyanyk, Y., Koijen, R. S., Van Hemert, O. et al. (2011), 'Determinants and consequences of mortgage default', *Federal Reserve Bank of Cleveland Working Paper* (10-19R).

Driessen, J., Maenhout, P. J. and Vilkov, G. (2009), 'The price of correlation risk: Evidence from equity options', *The Journal of Finance* **64**(3), 1377–1406.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2009.01467.x*

Duellmann, K., Küll, J. and Kunisch, M. (2010), 'Estimating asset correlations from stock prices or default rates—which method is superior?', *Journal of Economic Dynamics and Control* **34**(11), 2341–2357. Special Issue: 2008 Annual Risk Management Conference held in Singapore during June 30 - July 2, 2008.
URL: *https://www.sciencedirect.com/science/article/pii/S0165188910001338*

Duellmann, K. and Scheule, H. (2003), 'Determinants of the asset correlations of german corporations and implications for regulatory capital', *Deutsches Bundesbank* .

Egami, M. and Kevkhishvili, R. (2017), 'An analysis of simultaneous company defaults using a shot noise process', *Journal of Banking & Finance* **80**, 135–161.

Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D. and Hunt, R. (2010), 'What "triggers" mortgage default?', *The American economic review* **100**(2), 490–494.

Federal Home Loan Mortgage Corporation (FHLMC (2022), 'Freddie mac single-family loan-level dataset'.
URL: *https://www.freddiemac.com/research/datasets/sf-loanlevel-dataset*

Financial Stability Board (2022), '2022 list of global systemically important banks (g-sibs)'.
URL: *https://www.fsb.org/2022/11/2022-list-of-global-systemically-important-banks-g-sibs/*

Floros, I. and White, J. T. (2016), 'Qualified residential mortgages and default risk', *Journal of Banking & Finance* **70**, 86–104.

for International Settlements (BIS), B. (2013), 'Capital requirements regulation (crr)', *Regulation (EU) No 575/2013 of the European Parliament and of the Council of 26 June 2013 on prudential requirements for credit institutions and investment firms and amending Regulation (EU) No 648/2012* .

Frey, R. and McNeil, A. (2003), 'Dependent defaults in models of portfolio credit risk', *Journal of Risk* **6**.

Frye, J. (2008), 'Correlation and asset correlation in the structural portfolio model', *The Journal of Credit Risk* **4**.

Furfine, C. (2020), 'The impact of risk retention regulation on the underwriting of securitized mortgages', *Journal of Financial Services Research* **58**(2-3), 91–114.

Fuster, A., Lucca, D. O. and Vickery, J. I. (2022), 'Mortgage-backed securities'.

Geidosch, M. (2014), 'Asset correlation in residential mortgage-backed security reference portfolios', *Journal of Credit Risk* **10**, 71–95.

Ghent, A. C. and Kudlyak, M. (2011), 'Recourse and residential mortgage default: Evidence from us states', *The Review of Financial Studies* **24**(9), 3139–3186.

Goodstein, R., Hanouna, P., Ramirez, C. D. and Stahel, C. W. (2017), 'Contagion effects in strategic mortgage defaults', *Journal of Financial Intermediation* **30**, 50–60.

Gordy, M. B. (2000), 'A comparative anatomy of credit risk models', *Journal of Banking and Finance* **24**(1), 119–149.
URL: *https://www.sciencedirect.com/science/article/pii/S0378426699000540*

Griffin, J. M., Nickerson, J. and Tang, D. Y. (2013), 'Rating shopping or catering? an examination of the response to competitive pressure for cdo credit ratings', *The Review of Financial Studies* **26**(9), 2270–2310.

Guiso, L., Sapienza, P. and Zingales, L. (2013), 'The determinants of attitudes toward strategic default on mortgages', *The Journal of Finance (New York)* **68**(4), 1473–1515.

Gupta, A. and Hansman, C. (2022), 'Selection, leverage, and default in the mortgage market', *The Review of Financial Studies* **35**(2), 720–770.

Huang, X., Zhou, H. and Zhu, H. (2009), 'A framework for assessing the systemic risk of major financial institutions', *Journal of Banking & Finance* **33**, 2036–2049.

Hull, J. (2015), *Risk management and financial institutions.Fourth Edition.*, Wiley finance series.

Hurst, E., Keys, B. J., Seru, A. and Vavra, J. (2016), 'Regional redistribution through the us mortgage market', *The American Economic Review* **106**(10), 2982–3028.

Jacobson, T., Lindé, J. and Roszbach, K. (2006), 'Internal ratings systems, implied credit risk and the consistency of banks' risk classification policies', *Journal of Banking & Finance* **30**(7), 1899–1926. Special Section: Banking and Finance in an Integrating Europe.
URL: *https://www.sciencedirect.com/science/article/pii/S0378426605001524*

Jones, T. and Sirmans, G. (2015), 'The underlying determinants of residential mortgage default', *Journal of Real Estate Literature* **23**, 169–205.

Krahnen, J.-P. and Wilde, C. (2022), 'Skin-in-the-game in abs transactions: A critical review of policy options', *Journal of Financial Stability* p. 100998.

Lee, Y., Rösch, D. and Scheule, H. (2021), 'Systematic credit risk in securitised mortgage portfolios', *Journal of Banking & Finance* **122**, 105996.

Linn, A. and Lyons, R. C. (2020), 'Three triggers? negative equity, income shocks and institutions as determinants of mortgage default', *The journal of real estate finance and economics* **61**(4), 549–575.

Longin, F. and Solnik, B. (2001), 'Extreme correlation of international equity markets', *The Journal of Finance* **56**(2), 649–676.
URL: *http://www.jstor.org/stable/222577*

Lopez, J. A. (2004), 'The empirical relationship between average asset correlation, firm probability of default, and asset size', *Journal of Financial Intermediation* **13**(2), 265–283. Bank Capital Adequacy Regulation under the New Basel Accord.
URL: *https://www.sciencedirect.com/science/article/pii/S1042957303000457*

Mandelker, D. R. (2018), 'Making puds work for you'.

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

Neumann, T. (2018), 'Mortgages: estimating default correlation and forecasting default risk', *Bank of England Working Paper* (708).
URL: *https://ssrn.com/abstract=3135271 or http://dx.doi.org/10.2139/ssrn.3135271*

Nickerson, J. and Griffin, J. M. (2017), 'Debt correlations in the wake of the financial crisis: What are appropriate default correlations for structured products?', *Journal of Financial Economics* **125**(3), 454–474.
URL: *https://www.sciencedirect.com/science/article/pii/S0304405X17301289*

of Governors of the Federal Reserve System Data, B. (2023), 'Assets and liabilities of commercial banks in the united states - h.8.'.
URL: *https://www.federalreserve.gov/releases/h8/current/*

Packham, N. and Woebbeking, C. (2019), 'A factor-model approach for correlation scenarios and correlation stress testing', *Journal of Banking & Finance* **101**, 92–103.
URL: *https://www.sciencedirect.com/science/article/pii/S0378426619300202*

Paulusch, J. and Schlütter, S. (2022), 'Sensitivity-implied tail-correlation matrices', *Journal of Banking & Finance* **134**, 106333.
URL: *https://www.sciencedirect.com/science/article/pii/S0378426621002843*

Pu, X. and Zhao, X. (2012), 'Correlation in credit risk changes', *Journal of Banking & Finance* **36**(4), 1093–1106.
URL: *https://www.sciencedirect.com/science/article/pii/S037842661100313X*

Quercia, R. G. and Stegman, M. A. (1992), 'Residential mortgage default: A review of the literature', *Journal of Housing Research* **3**(2), 341–379.
URL: *http://www.jstor.org/stable/24841960*

Rösch, D. and Scheule, H. (2004), 'Forecasting retail portfolio credit risk', *The Journal of Risk Finance* **5**, 16–32.

Springer, T. M. and Waller, N. G. (1993), 'Termination of Distressed Residential Mortgages: An Empirical Analysis', *The Journal of Real Estate Finance and Economics* **7**(1), 43–54.

Stoffberg, H. and Vuuren, G. (2015), 'Asset correlations in single factor credit risk models: an empirical investigation', *Applied Economics* **48**, 1–16.

Tarashev, N. (2010), 'Measuring portfolio credit risk correctly: Why parameter uncertainty matters', *Journal of Banking & Finance* **34**(9), 2065–2076.
URL: *https://www.sciencedirect.com/science/article/pii/S0378426610000269*

von Hoffman, A. (2012), 'History lessons for today's housing policy: the politics of low-income housing', *Housing Policy Debate* **22**(3), 321–376.
URL: *https://doi.org/10.1080/10511482.2012.680478*

Wang, Q., Yao, L. and Lai, P. (2009), 'Estimation of the area under roc curve with censored data', *Journal of Statistical Planning and Inference* **139**(3), 1033–1044.

Yang, L., Lahiri, K. and Pagan, A. (2023), 'Getting the roc into sync', *Journal of Business & Economic Statistics* pp. 1–13.

Yildirim, Y. (2008), 'Estimating default probabilities of cmbs loans with clustering and heavy censoring', *The Journal of Real Estate Finance and Economics* **37**, 93–111.

Zeng, G. and Zeng, E. (2019), 'On the three-way equivalence of auc in credit scoring with tied scores', *Communications in Statistics. Theory and Methods* **48**(7), 1635–1650.

Figure 1: Summary Statistics:States Distribution

The figure displays the distribution of mortgages by States across the entire sample. The sample covers all the Single-Family residential mortgages originated from 1999 to 2017 in Freddie Mac database.

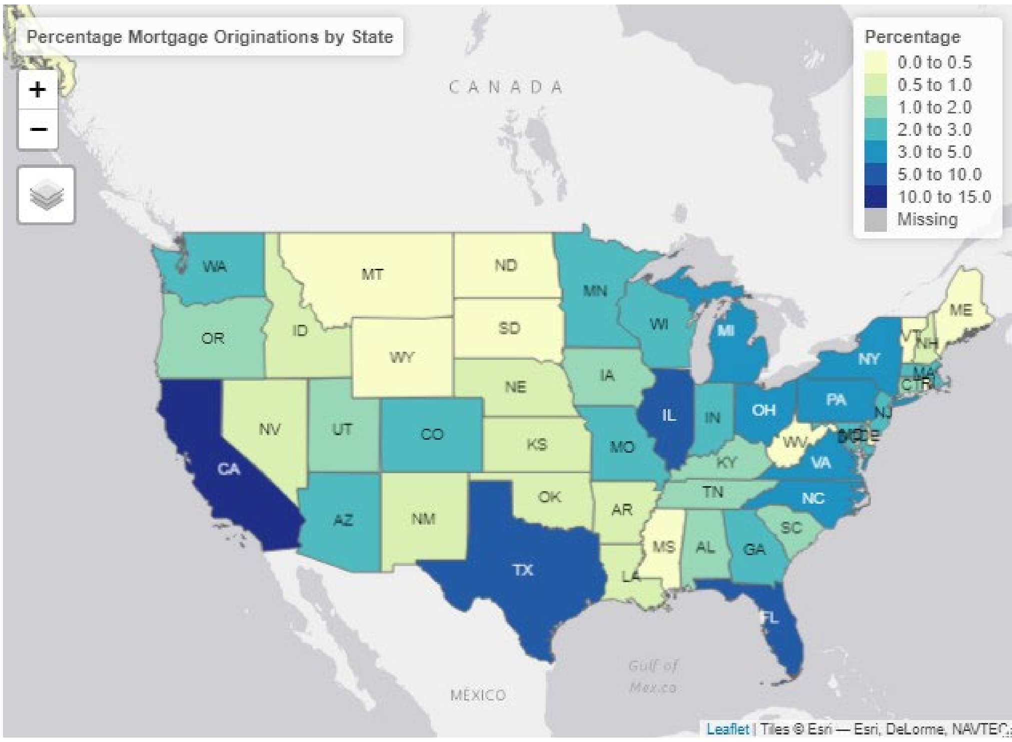
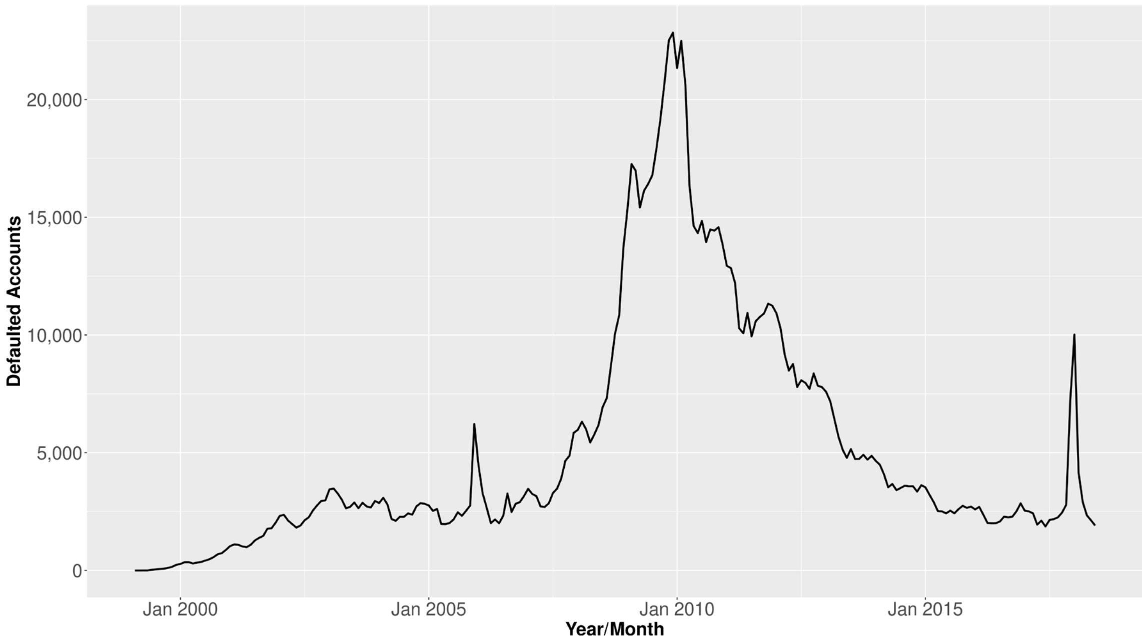
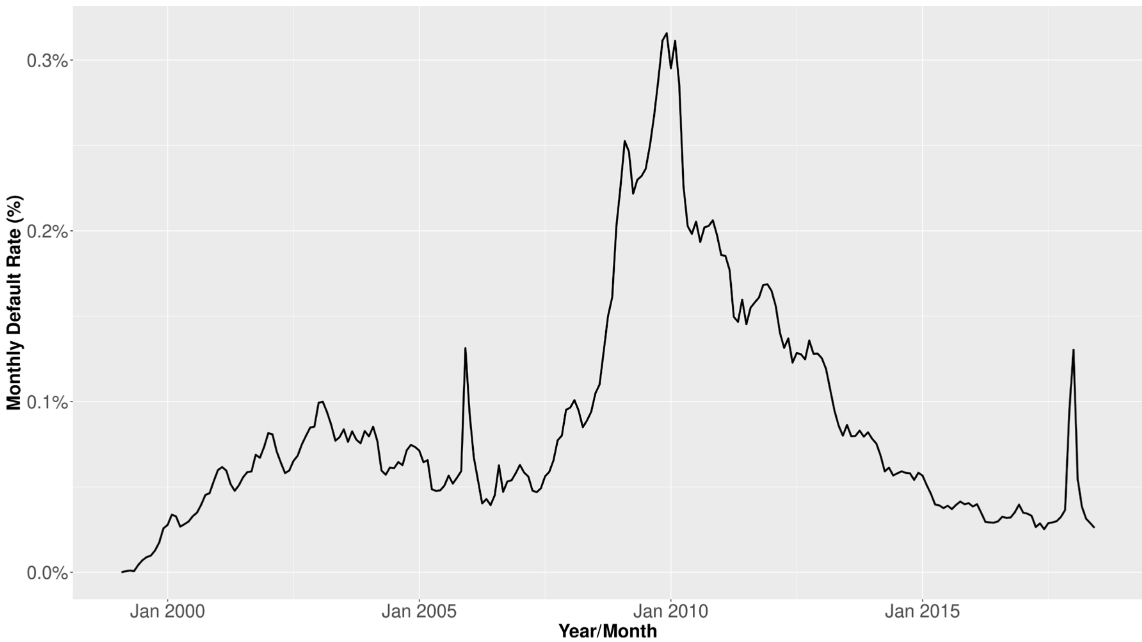


Figure 2: Defaulted Mortgages Over Time

Panel (a) displays the number of first default occurrences by month, from February 1999 to June 2018. Panel (b) displays the ratio between first default occurrences and outstanding mortgages by month, from February 1999 to June 2018.



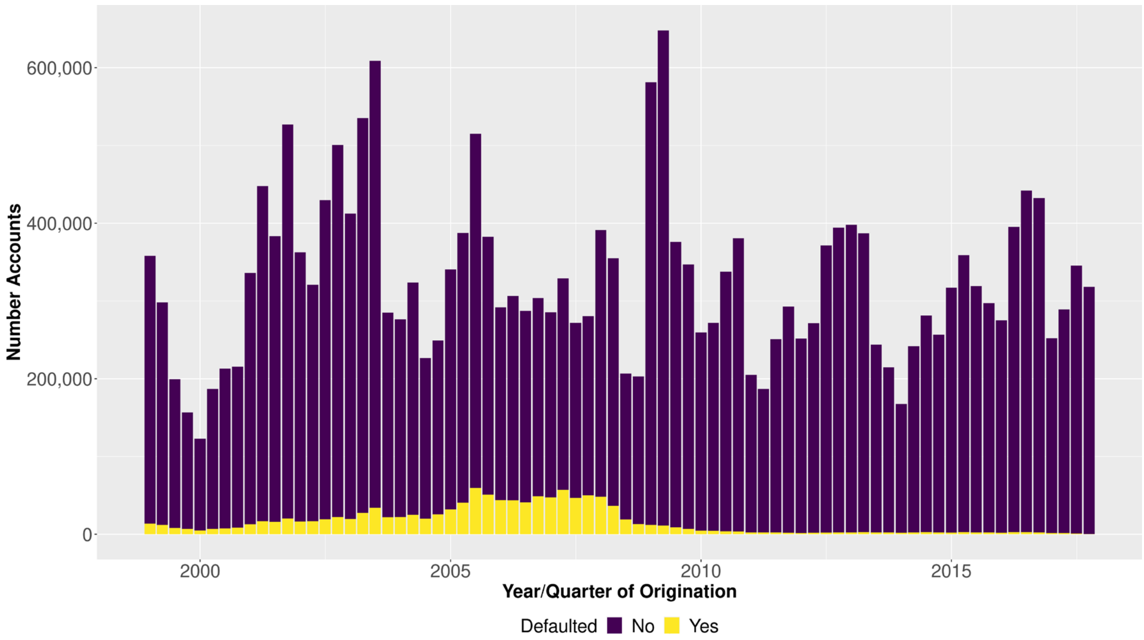
(a)



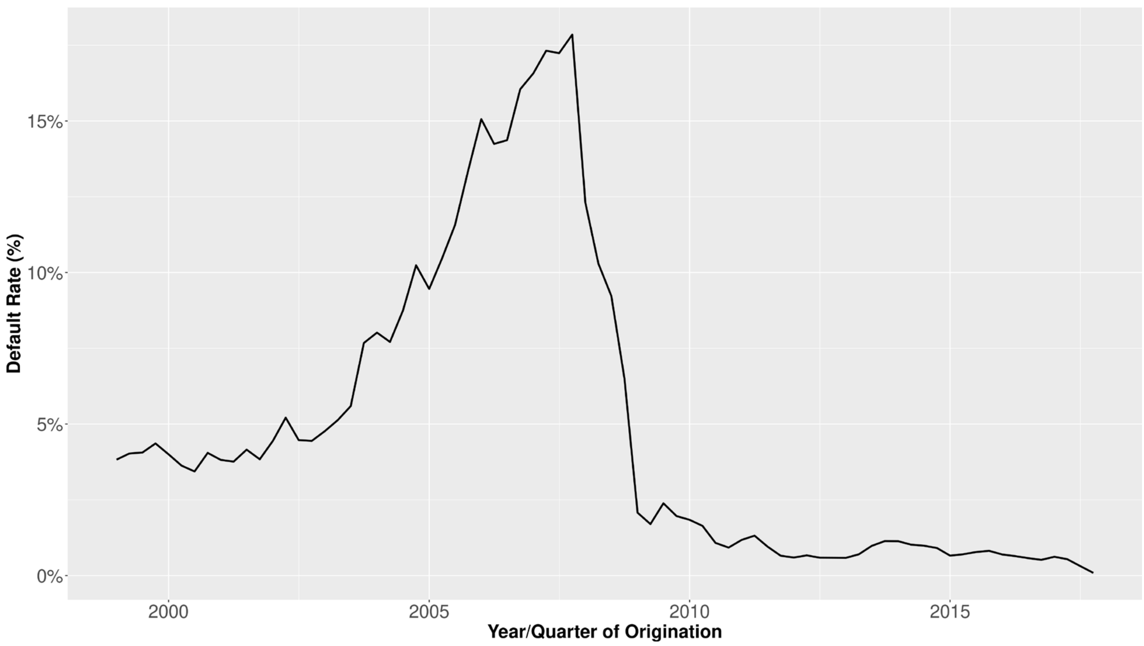
(b)

Figure 3: Defaulted Mortgages Over Vintage (Year/Quarter of Origination)

Panel (a) displays the number of mortgages by year and quarter of origination. The yellow portion of graph counts the number of mortgages belonging to that quarter of origination that have defaulted within the observation period. Panel (b) displays the ratio between first default occurrences and originated mortgages by quarter, from 1999q1 to 2017q4.



(a)



(b)

Figure 4: Ratio of Default Rate during Crisis over Default Rate during non-Crisis

The figure displays the ratio between average yearly default rate during Great Financial Crisis (GFC) and average yearly default rate before the GFC by State across the entire sample.

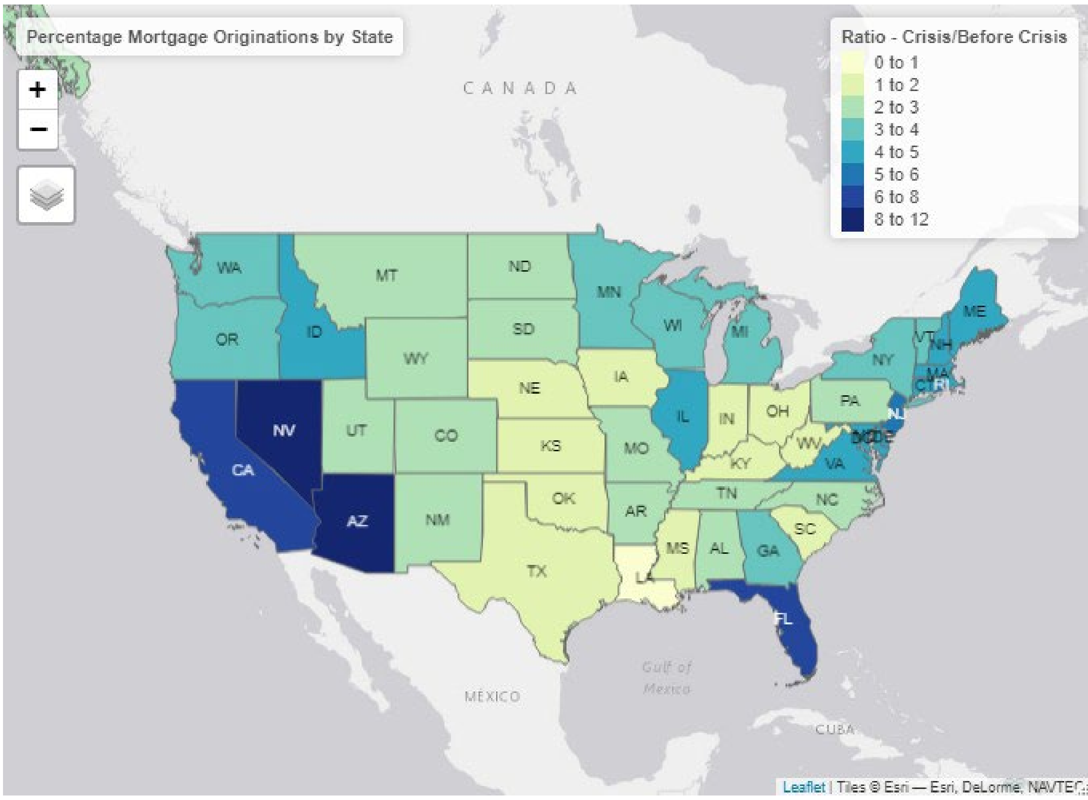
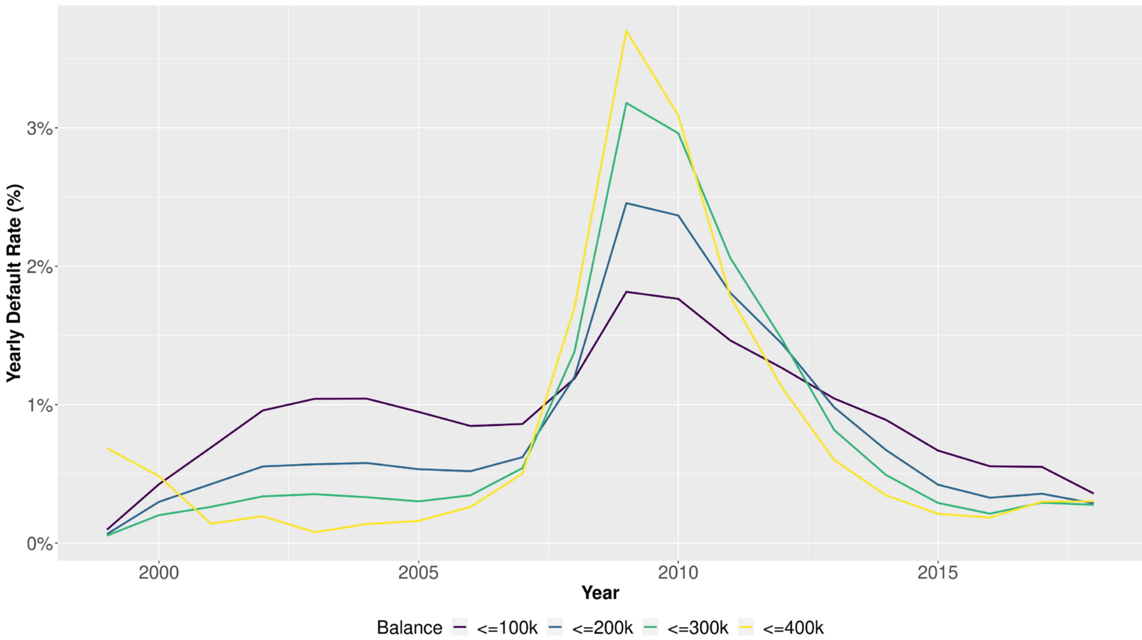
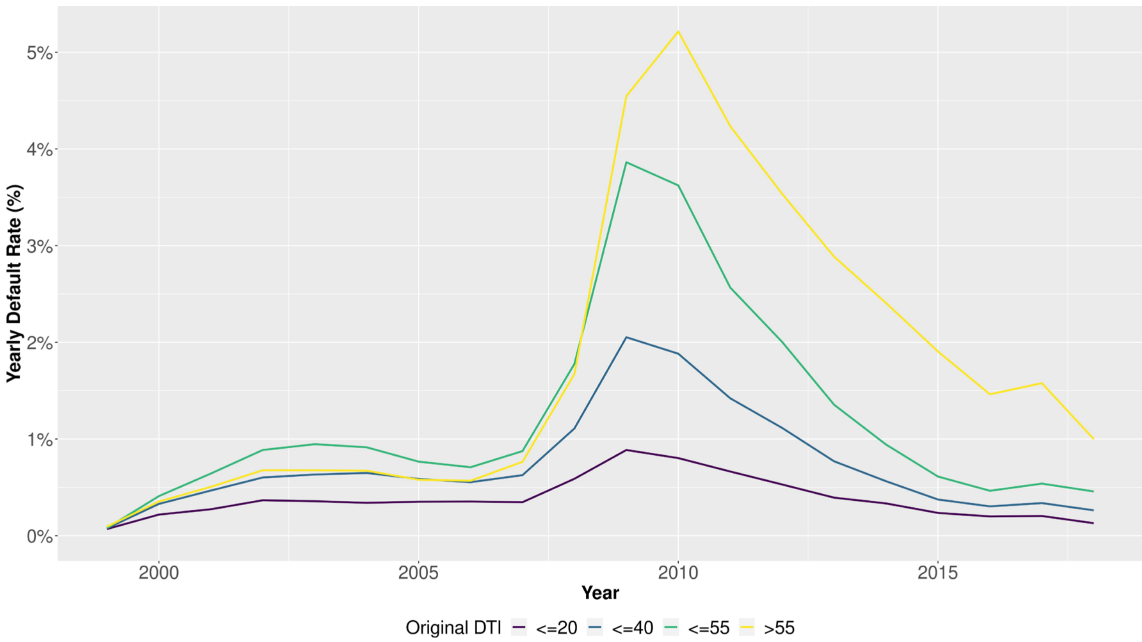


Figure 5: Default Rate by Balance and Original Debt-to-Income

The figure displays the yearly default rate by Balance (a) and Original Debt-to-Income (b) banding. The yearly default rate is the ratio between first default occurrences and outstanding mortgages by year, from 1999 until 2018.



(a)



(b)

Figure 6: Marginal Effects by Credit Score, Updated LTV, Original DTI and Balance

The figure displays marginal effects of Credit Score (a), Updated Loan-to-Value (b), Original Debt-to-Income (c) and Balance (ln) (d) on yearly default probability of both Long Run and Crisis, calculated from *Model 1* logistic regression specification. The dotted line reports the Average Marginal Effects, while the shaded area delimits the 99% confidence intervals

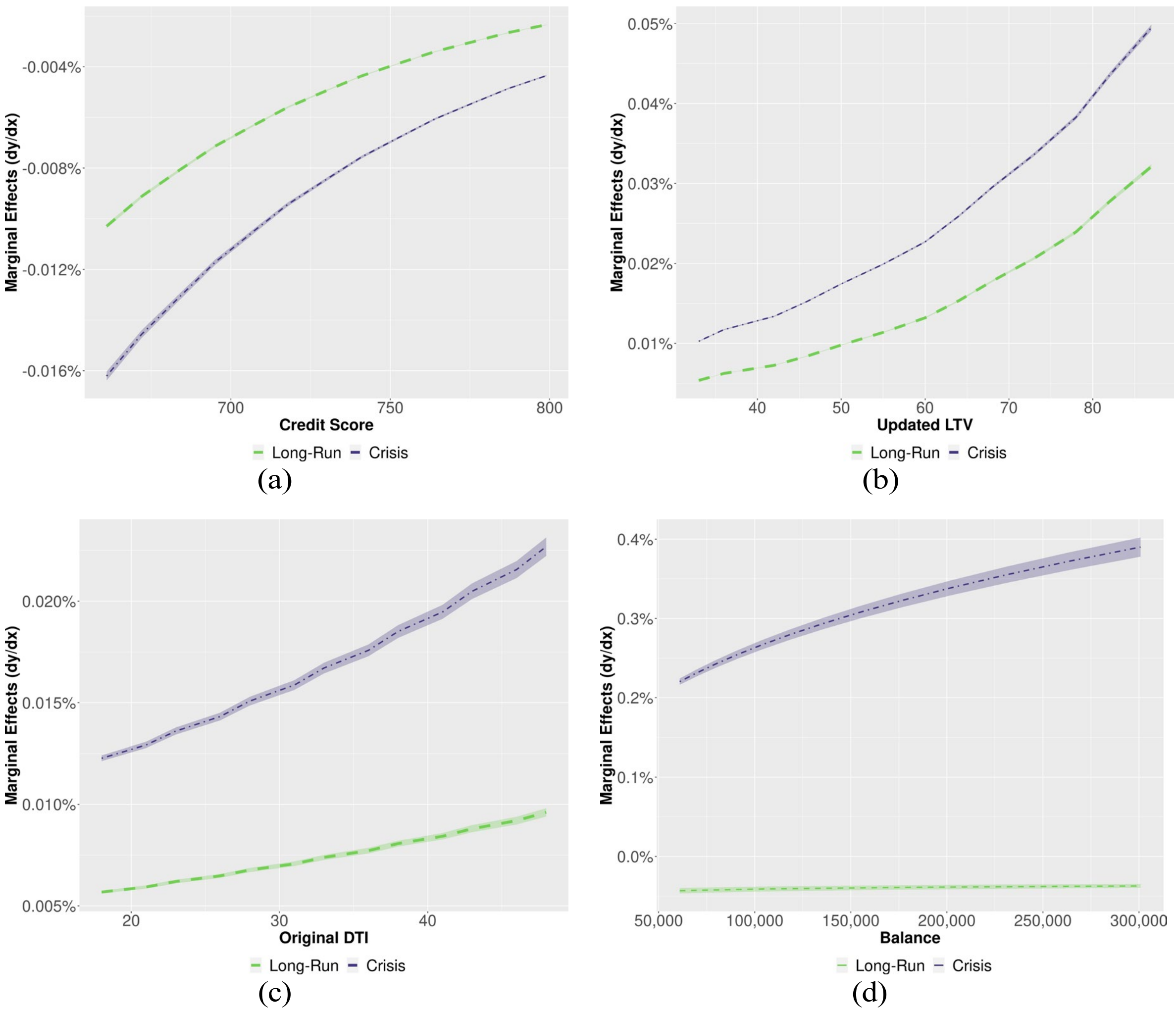
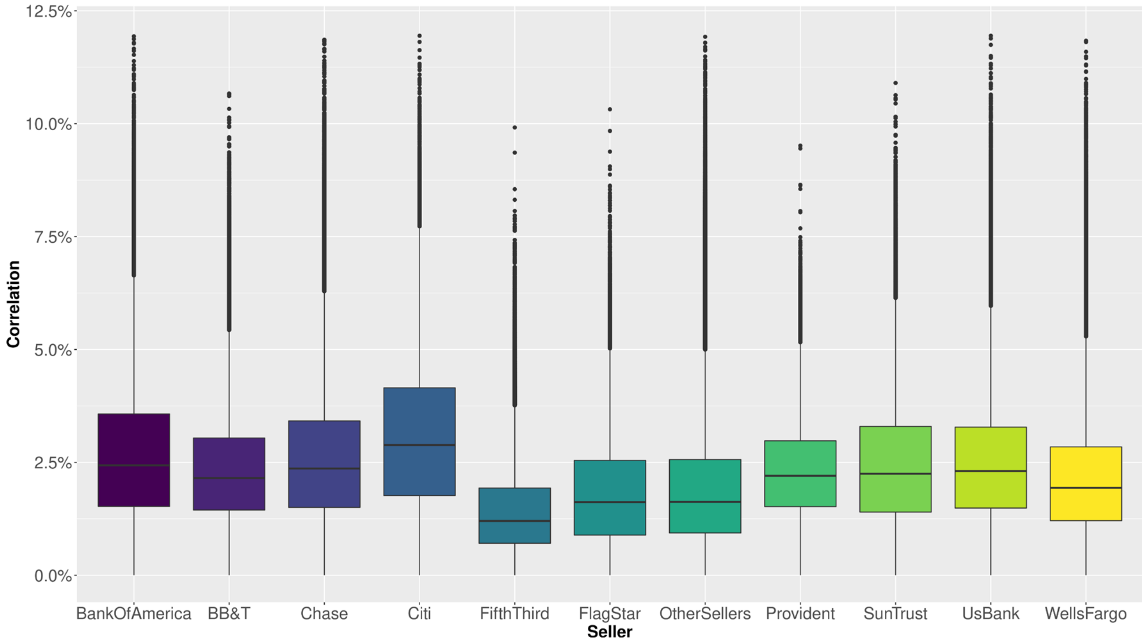
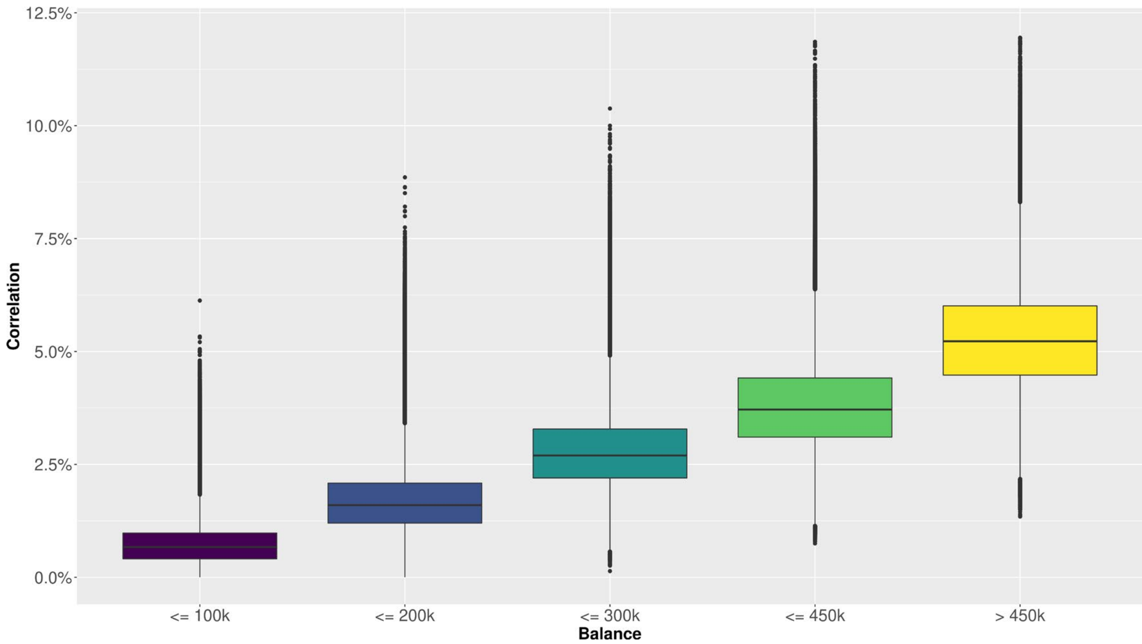


Figure 7: Correlation by Lenders and Balance

The graph displays correlation distribution by Lenders (a) and Balance (b). The correlations are calculated on the entire Freddie Mac cross-sectional sample, where each record is a mortgage at origination. The population is segmented on the x-axis by Lenders (a) and by increasing Balance buckets (b). Correlation distribution is plotted on the y-axis via box-plot. The box encloses the distribution from the 25th percentile (lower bound) to the 75th percentile (upper bound). The upper and lower whiskers represent correlation outside the middle 50% (i.e. the lower 25% and the upper 25% of correlations).



(a)



(b)

Figure 8: Excess Interests Paid

The graph shows the maximum difference amongst banks in pricing the effect of mortgage correlation. The reference loan is a 30-Year Fixed-Rate Mortgage with original balance of \$ 300,000 and 30-year Fixed-rate of 5.5%. The total interests paid for such mortgage are \$ 310,000. The isolated impact of correlation on Excess Interest rate is first calculated for each Bank; then the maximum difference amongst banks is produced and plotted. There is no other contributing factor to the difference in total interests paid. The distribution is concentrated below \$ 15,000 (4.8% of total interests), suggesting that banks differentiate in pricing correlation. The right tail of the distribution shows that the difference in pricing stretches beyond \$ 40,000 (12.8% of total interests). This proves that banks price in a different way specific mortgage segments based on correlation risk experienced by these clusters.

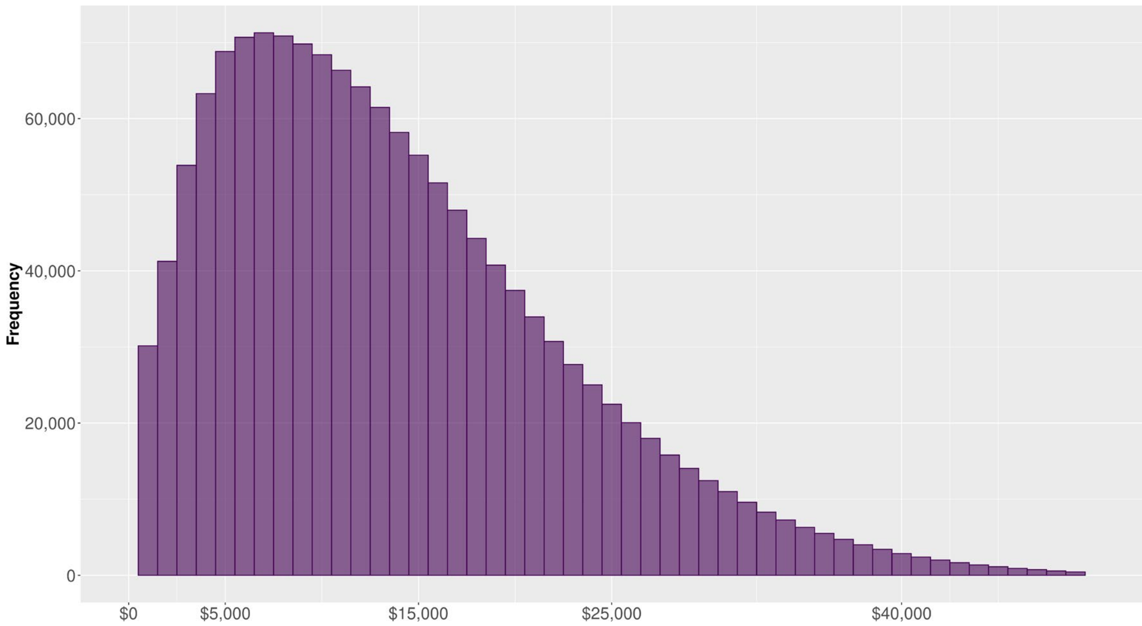


Table 1: Summary Statistics:Continuous Variables

The table reports number of accounts, 5^{th} quantile, mean, standard deviation and 95^{th} quantile of Credit Score, Loan-to-Value, Debt-to-Income, Interest rate and Balance by year of origination

		Credit Score				Loan-To-Value				Debt-to-Income				Interest Rate				Balance			
Year	N.Accounts	q5	Mean	Sd	q95	q5	Mean	Sd	q95	q5	Mean	Sd	q95	q5	Mean	Sd	q95	q5	Mean	Sd	q95
1999	1,095,011	621	711.8	52.0	785	45	76.7	15.2	95	15	32.8	11.0	51	6.5	7.3	0.6	8.3	50,000	125,942	54,599	230,000
2000	786,275	615	712.2	55.6	789	45	77.6	15.5	95	17	34.7	10.7	51	7.4	8.2	0.5	9.0	50,000	131,824	58,840	245,000
2001	1,755,390	617	714.1	58.7	791	45	75.4	14.7	95	15	33.2	11.0	50	6.4	7.0	0.4	7.8	58,000	147,801	64,408	273,000
2002	1,682,997	617	717.1	56.7	792	42	73.9	15.5	95	15	33.5	11.7	53	5.9	6.6	0.5	7.4	59,000	155,506	70,407	292,000
2003	1,927,050	632	724.9	51.6	794	40	72.1	15.7	95	12	32.3	12.3	53	5.1	5.8	0.4	6.5	62,000	161,475	74,414	302,000
2004	1,127,674	624	717.9	54.5	794	42	73.7	15.3	95	15	35.0	12.0	55	5.4	5.9	0.4	6.5	61,000	166,776	78,238	320,000
2005	1,690,993	626	724.4	58.3	802	34	69.6	17.4	95	15	35.3	12.3	56	5.3	5.8	0.4	6.5	57,000	171,054	86,077	343,000
2006	1,260,783	623	723.0	58.3	803	34	70.6	17.4	95	16	36.6	12.2	57	5.8	6.4	0.4	7.0	59,000	179,703	94,317	372,000
2007	1,220,654	621	722.9	58.4	804	35	72.0	17.7	95	16	36.8	12.5	58	5.8	6.4	0.4	7.1	59,000	183,435	98,207	388,000
2008	1,179,578	643	739.7	51.9	806	35	70.1	17.5	95	16	36.4	12.8	58	5.3	6.0	0.5	6.9	62,000	203,644	108,433	417,000
2009	1,974,690	685	762.4	39.6	809	32	65.2	17.0	88	14	31.6	11.6	52	4.4	4.9	0.4	5.6	68,000	214,357	117,130	417,000
2010	1,271,397	684	763.7	40.2	811	33	66.2	17.0	90	14	31.5	10.5	48	3.8	4.6	0.5	5.4	65,000	208,668	120,259	417,000
2011	955,418	685	764.2	39.8	811	33	67.3	17.2	90	15	31.7	10.1	47	3.4	4.3	0.6	5.3	65,000	217,863	125,849	417,000
2012	1,331,301	690	766.5	38.4	812	34	68.0	17.2	95	14	30.8	10.0	46	2.9	3.6	0.4	4.3	70,000	222,291	122,763	417,000
2013	1,300,286	681	759.5	41.0	810	36	70.6	17.2	95	15	32.0	9.8	46	2.8	3.7	0.6	4.8	68,000	218,019	118,454	417,000
2014	975,881	670	751.5	44.1	808	41	75.1	16.2	95	17	33.8	9.2	46	3.4	4.3	0.5	4.9	68,000	219,279	118,550	417,000
2015	1,316,566	670	752.3	44.0	809	40	73.6	16.5	95	17	33.8	9.4	47	3.1	4.0	0.5	4.6	75,000	229,030	119,353	417,000
2016	1,558,394	668	751.1	44.6	809	39	73.0	16.6	95	17	34.1	9.4	48	2.9	3.8	0.5	4.5	80,000	241,231	119,284	431,000
2017	1,217,105	662	747.0	45.9	808	40	74.1	16.5	95	18	35.0	9.4	48	3.4	4.2	0.4	4.9	75,000	235,267	120,715	424,000

Table 2: Yearly Default Rate by Credit Score, Debt-to-Income and Excess Interest Rate.

The table reports the yearly default rate (expressed in percentage) by year of observation. Yearly default rate, number of observations and number of defaults at portfolio level are reported in the first three columns. The yearly default rate is then segmented by Credit Score, Debt-to-Income and Excess Interest Rate by different buckets.

				Credit Score					Debt-to-Income				Excess IR					
Year	N.Accounts	Defaults	All	0-579	580-669	670-739	740-799	800-850	0-20	21-40	41-55	56-65	6-1%	6 -0.5%	6 0%	6 0.5%	6 1%	>1%
1999	939,197	718	0.08	1.04	0.17	0.05	0.02	0.02	0.07	0.08	0.08	0.11	0.03	0.03	0.05	0.09	0.21	0.36
2000	1,706,123	5,729	0.34	2.12	0.89	0.22	0.07	0.07	0.22	0.33	0.41	0.49	0.13	0.1	0.22	0.38	0.84	1.47
2001	3,292,855	16,049	0.49	2.67	1.35	0.33	0.09	0.10	0.27	0.47	0.64	0.69	0.15	0.18	0.31	0.52	1.31	3.2
2002	4,382,304	28,120	0.64	3.79	1.81	0.44	0.11	0.09	0.37	0.6	0.89	0.66	0.32	0.29	0.36	0.65	2.04	5.82
2003	5,311,191	35,279	0.66	4.68	2.03	0.47	0.11	0.08	0.36	0.63	0.95	0.66	0.38	0.26	0.39	0.75	1.73	5.66
2004	4,631,059	30,804	0.67	4.32	1.91	0.53	0.12	0.09	0.34	0.65	0.91	0.66	0.45	0.26	0.41	0.77	1.51	4.64
2005	5,424,031	32,325	0.60	3.48	1.73	0.52	0.14	0.07	0.35	0.59	0.77	0.60	0.27	0.27	0.42	0.65	1.44	4
2006	5,927,775	33,570	0.57	3.01	1.61	0.52	0.16	0.09	0.35	0.55	0.71	0.59	0.26	0.27	0.41	0.63	1.32	3.47
2007	6,641,321	44,175	0.67	3.61	1.93	0.64	0.16	0.08	0.35	0.63	0.87	0.81	0.26	0.26	0.44	0.8	1.58	3.23
2008	7,359,285	93,166	1.27	5.76	3.63	1.35	0.37	0.17	0.59	1.11	1.78	1.87	0.3	0.44	0.82	1.59	3.02	5.12
2009	8,645,786	217,599	2.52	10.34	7.74	3.22	0.86	0.30	0.89	2.05	3.86	5.11	0.79	1.12	1.79	3.1	5.29	8.23
2010	8,370,112	195,879	2.34	8.38	7.04	3.24	0.98	0.34	0.8	1.88	3.62	5.87	1.12	1.14	1.87	2.74	4.34	6.41
2011	7,803,125	134,082	1.72	6.36	5.14	2.47	0.81	0.31	0.66	1.42	2.57	4.73	0.92	0.82	1.46	1.95	2.66	4.7
2012	7,776,197	103,176	1.33	5.13	4.31	2.03	0.64	0.26	0.53	1.11	2	3.85	0.76	0.58	1.22	1.45	1.98	3.38
2013	7,245,881	65,902	0.91	4.39	3.45	1.4	0.42	0.17	0.39	0.77	1.35	3.16	0.43	0.34	0.87	0.97	1.4	2.5
2014	6,844,299	45,048	0.66	3.96	2.72	1.01	0.3	0.13	0.33	0.56	0.94	2.63	0.23	0.25	0.65	0.68	1.06	2
2015	7,515,347	32,781	0.44	3.1	1.86	0.68	0.21	0.10	0.24	0.37	0.61	2.07	0.15	0.17	0.41	0.46	0.75	1.45
2016	8,128,933	27,950	0.34	2.86	1.5	0.54	0.16	0.08	0.2	0.3	0.46	1.57	0.12	0.13	0.28	0.38	0.66	1.2
2017	8,415,102	32,457	0.39	3.57	1.65	0.61	0.18	0.09	0.2	0.34	0.54	1.69	0.14	0.15	0.29	0.43	0.79	1.23
2018	7,705,902	23,463	0.30	2.08	1.15	0.5	0.16	0.07	0.13	0.26	0.46	1.10	0.21	0.14	0.21	0.34	0.65	0.97

Table 3: Default Probability: Average Marginal Effects

The table shows average marginal effects of explanatory variables on default probability, split by Long Run and Crisis. Standard errors of marginal effects are calculated using the delta method and are reported in parentheses. Credit Score is borrower's Credit Score at origination. Debt-to-Income (DTI) is the sum of borrower's monthly debt payments, including housing expenses that incorporate mortgage payment, divided by total monthly income used for underwriting. Updated LTV is the ratio between outstanding $Balance_t$ and $PropertyPrice_t$, which is derived from State-level House Price Index at time t. Balance (ln) is the natural logarithm of mortgage outstanding balance. Excess Interest Rate is the difference between rate at origination and average interest rate of all mortgages generated in the same quarter. Ump_{12} is the 1-year growth rate of State-level Unemployment. The Crisis period covers the years of mortgage downturn (2009, 2010 and 2011). The sample includes mortgages originated from 1999 to 2017 and observed from 1999 to 2018. *** p<0.01; ** p<0.05; * p<0.1

Variables	Model1	Model1a	Model2	Model2a	Model3	Model3a	Model4
Credit Score	-0.0001071***	-0.0001089***	-0.0001103***	-0.0001074***	-0.0000938***	-0.0000943***	-0.000088***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Debt-to-Income	0.0002853***	0.000259***	0.000254***	0.0002293***	0.000183***	0.0001819***	0.000144***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Excess Int. Rate	0.0052216***	0.0054109***	0.0063047***	0.0066509***	0.0049472***	0.0050559***	0.004585***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Non-Recourse	-0.0010437***	-0.000944***	-0.0012086***	-0.0010887***	0.0008215***	0.0007274***	0.0000784*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.027)
Joint	-0.0049071***	-0.0048926***	-0.0057076***	-0.0056093***	-0.0051561***	-0.0051478***	-0.0048393***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Balance (ln)			0.0046437***	0.0040831***	0.0006959***	0.0007263***	-0.000777***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Updated LTV					0.0003088***	0.0002904***	0.000274***
					(0.000)	(0.000)	(0.000)
Ump12					0.0014191***	0.0012435***	0.00119***
					(0.000)	(0.000)	(0.000)
DummyCrisis(DC)		0.0154335***		0.0114012***		0.002683***	0.0029457***
		(0.000)		(0.000)		(0.000)	(0.000)
DC*CreditScore							-0.000106***
							(0.000)
DC*OriginalDTI							0.000227***
							(0.000)
DC*UpdatedLTV							0.000326***
							(0.000)
DC*lnBalance							0.003971***
							(0.000)
DC*ExcessIR							0.006182***
							(0.000)
DC*Non-Recourse							0.0016786***
							(0.000)
DC*First-Time Homebuyer							-0.0009764***
							(0.000)
DC*Joint							-0.0058166***
							(0.000)
LoanAge	No	No	Yes	Yes	Yes	Yes	Yes
BEA Territories FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Property FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	122,064,071	122,064,071	122,064,071	122,064,071	122,064,071	122,064,071	122,064,071
AUROC	78.780%	81.580%	83.600%	84.950%	87.760%	87.780%	87.970%
GINI	57.560%	63.160%	67.200%	69.900%	75.520%	75.560%	75.940%
Pseudo-R2	9.370%	12.410%	13.690%	15.670%	19.880%	20.010%	20.480%

Table 4: In-Sample Correlations

The table displays correlation distribution calculated in-sample across segments. These include Credit Score, Updated Loan-to-Value, Original Debt-to-Income, BEA Territories and Non-recourse. The correlations are calculated on the entire portfolio, where only the first observation is kept for each mortgage.

Variable	Segment	N.Observations	Mean	Median	SD	q75	q90	q99	Max
Credit Score	< 580	103,166	1.60%	1.33%	1.25%	2.30%	3.33%	5.34%	9.64%
	< 670	2,932,751	1.86%	1.61%	1.31%	2.61%	3.67%	5.81%	13.07%
	< 740	8,155,052	2.10%	1.86%	1.33%	2.87%	3.93%	6.05%	12.60%
	< 800	11,641,717	2.14%	1.93%	1.26%	2.88%	3.87%	5.82%	12.99%
	≥ 800	2,419,57	2.01%	1.80%	1.20%	2.72%	3.67%	5.50%	12.53%
Updated LTV	≤ 40	1,740,420	1.80%	1.54%	1.24%	2.52%	3.56%	5.45%	10.50%
	≤ 60	4,103,466	2.16%	1.93%	1.32%	2.95%	3.98%	5.96%	12.25%
	≤ 70	3,818,242	2.28%	2.06%	1.35%	3.08%	4.11%	6.20%	12.99%
	≤ 85	10,044,337	2.08%	1.86%	1.25%	2.80%	3.78%	5.79%	12.79%
	≤ 100	4,432,354	1.88%	1.67%	1.20%	2.60%	3.53%	5.36%	13.07%
	> 100	1,113,437	2.32%	2.05%	1.51%	3.22%	4.45%	6.61%	12.53%
Original DTI	≤ 15	1,370,102	1.39%	1.20%	0.96%	1.92%	2.71%	4.30%	10.54%
	≤ 30	8,606,096	1.76%	1.58%	1.10%	2.41%	3.27%	4.98%	11.31%
	≤ 45	11,746,676	2.24%	2.04%	1.32%	3.04%	4.05%	6.04%	12.60%
	≤ 55	2,904,321	2.54%	2.33%	1.42%	3.41%	4.48%	6.51%	13.07%
	> 55	625,061	2.66%	2.45%	1.48%	2.45%	4.68%	6.86%	12.53%
Balance	≤ 100k	5,191,020	0.73%	0.67%	0.46%	0.98%	1.33%	2.16%	6.13%
	≤ 200k	10,684,158	1.69%	1.60%	0.68%	2.08%	2.60%	3.71%	8.86%
	≤ 300k	5,704,745	2.80%	2.69%	0.83%	3.28%	3.89%	5.19%	10.38%
	≤ 450k	3,197,043	3.82%	3.71%	1.00%	4.41%	5.13%	6.67%	12.33%
	> 450k	475,290	5.28%	5.23%	1.18%	6.01%	6.79%	8.40%	13.07%
Region	FarWest	4,776,565	3.13%	3.00%	1.41%	4.03%	5.02%	6.90%	13.07%
	GreatLakes	4,549,138	1.46%	1.29%	0.92%	1.98%	2.74%	4.19%	10.10%
	Mideast	3,105,342	2.35%	2.21%	1.26%	3.13%	4.03%	5.86%	12.01%
	NewEngl.	1,388,126	2.60%	2.48%	1.18%	3.32%	4.16%	5.92%	12.09%
	Plains	2,077,061	1.61%	1.46%	0.95%	2.17%	2.92%	4.30%	9.08%
	RockyMount.	1,338,460	2.65%	2.53%	1.17%	3.37%	4.20%	5.82%	11.66%
	Southeast	5,603,178	1.59%	1.43%	0.98%	2.18%	2.95%	4.41%	9.95%
	Southwest	2,377,482"	1.68%	1.54%	1.00%	2.29%	3.06%	4.43%	9.81%
Recourse	Non-Recourse	8,848,243	2.65%	2.46%	1.39%	3.51%	4.55%	6.53%	13.07%
	Recourse	16,404,013	1.77%	1.57%	1.11%	2.43%	3.31%	4.99%	12.79%

Table 5: Determinants of Excess Mortgage Rates

The table reports the estimated coefficients for the linear regressions and their robust standard errors. The bottom of the table reports number of observations and Adjusted-R2. The regression is run on a cross-sectional sample that comprises all mortgages at origination in Freddie Mac database from January 2012 until December 2017. The dependent variable is the Excess (delta) IR from the average Interest Rate by quarter of origination. The independent variable ρ is the mortgage default correlation derived from the Logistic Regression on the Freddie Mac sample including all mortgages originated up to December 2011. This ensures independency of the estimated input correlation ρ into excess interest rate regression. Ump12 (HPI12) is the 1-year Unemployment rate (HPI) growth rate at State level. Loan-to-Value (LTV), Credit Score and Debt-to-Income (DTI) are continuous variables at origination. Non-Recourse and Joint are control variables in addition to the Fixed Effects (FE) listed at the bottom of the table. *** p<0.01; ** p<0.05; * p<0.1.

Variables	Model 1	Model 2	Model 3
Credit Score	-0.0019***	-0.0019***	-0.0019***
Original LTV	0.0063***	0.0062***	0.0062***
Original DTI	0.0024***	0.0026***	0.0026***
Joint	-0.0441***	-0.0439***	-0.0440***
Non Recourse	-0.0252***	-0.0261***	-0.0259***
ρ	1.0183***	0.4699***	-0.9464***
ρ*BB&T			1.9177***
ρ*Chase			-2.9899***
ρ*Citi			-0.6605***
ρ*FifthThird			2.2298***
ρ*Others			2.2724***
ρ*Provident			1.4351***
ρ*SunTrust			1.8623***
ρ*UsBank			3.454***
ρ*WellsFargo			0.4684***
Constant	1.3625***	1.3211***	1.3499***
Macro Controls	Yes	Yes	Yes
Bank FE	No	Yes	Yes
Region FE	Yes	Yes	Yes
Loan FE	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes
Property FE	Yes	Yes	Yes
Observations	7,680,619	7,680,619	7,680,619
Adjusted-R^2	22.54%	23.41%	23.53%

Table 6: Determinants of Excess Mortgage Rates by Lender

The table reports estimated coefficients and robust standard errors for linear regressions by lender. The bottom of the table reports number of observations and Adjusted-R^2. The regression is run on cross-sectional samples comprises all mortgages in Freddie Mac database, originated by each lender from January 2012 until December 2017. The dependent variable is the Excess (delta) IR from the average Interest Rate by quarter of origination. The independent variable ρ is the mortgage default correlation derived from the Logistic Regression on the Freddie Mac sample including all mortgages originated up to December 2011. This ensures independency of the estimated input correlation ρ into excess interest rate regression. Ump12 (HPI12) is the 1-year Unemployment rate (HPI) growth rate at State level. Loan-to-Value (LTV), Credit Score and Debt-to-Income (DTI) are continuous variables at origination. Non-Recourse, First-Time Homebuyer and Joint are control variables in addition to the Fixed Effects (FE) listed at the bottom of the table. *** p<0.01; ** p<0.05; * p<0.1.

Variables	BankofAmerica	BBAndT	JPMorganChase	Citi	FifthThird	OtherSellers	Provident	SunTrust	USBank	WellsFargo
ρ	-1.1381***	2.5277***	-6.0108***	-0.3661**	0.5623**	1.1007***	5.958***	1.1705***	3.8724***	-0.7462***
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mortgage Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Units FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	348,815	352,683	509,201	71,736	77,628	4,366,440	117,602	96,859	560,804	1,178,851
Adjusted-R^2	15.27%	19.46%	21.25%	21.02%	22.47%	24.60%	13.07%	20.83%	20.24%	25.54%

A Appendix

Figure A1: Yearly Default Rate by BEA Territories and Lenders

The figure displays yearly default rate by top Sellers.

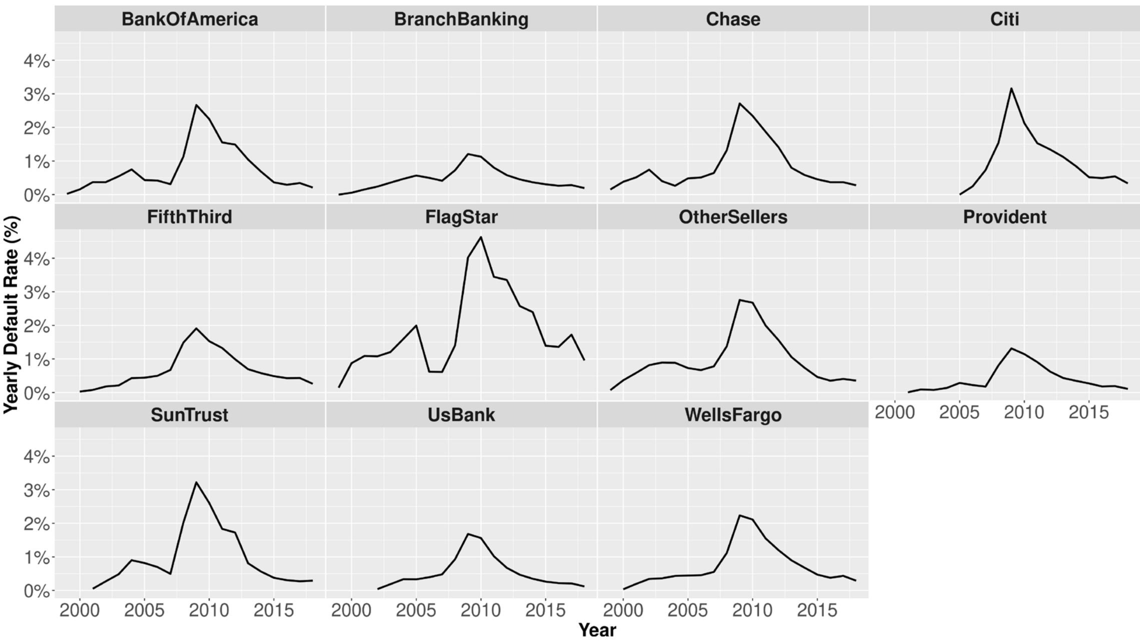
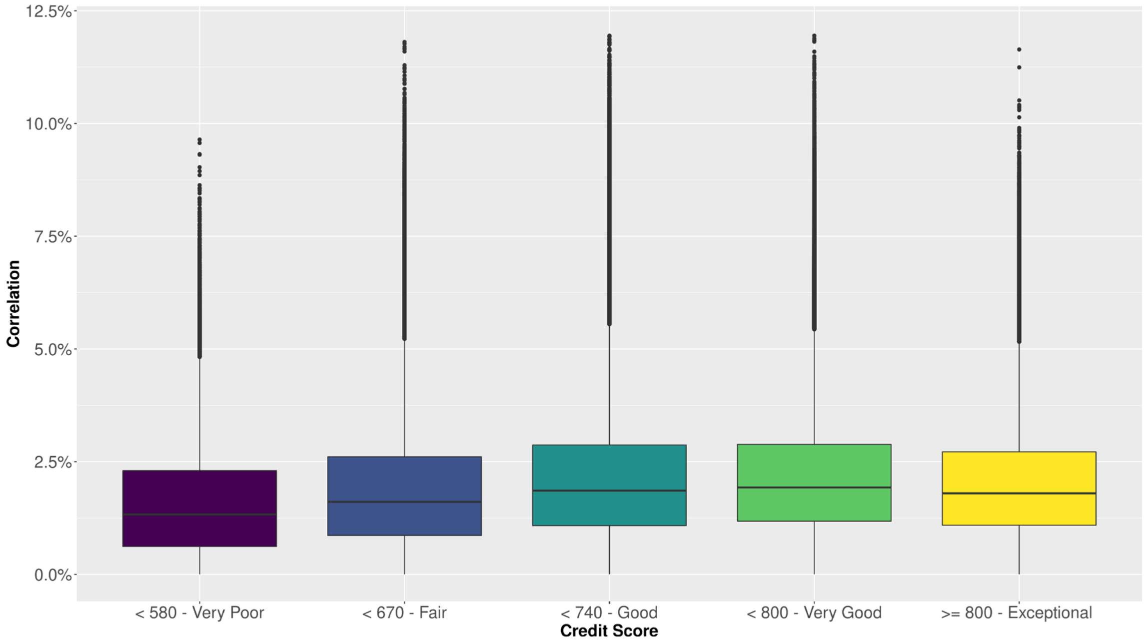
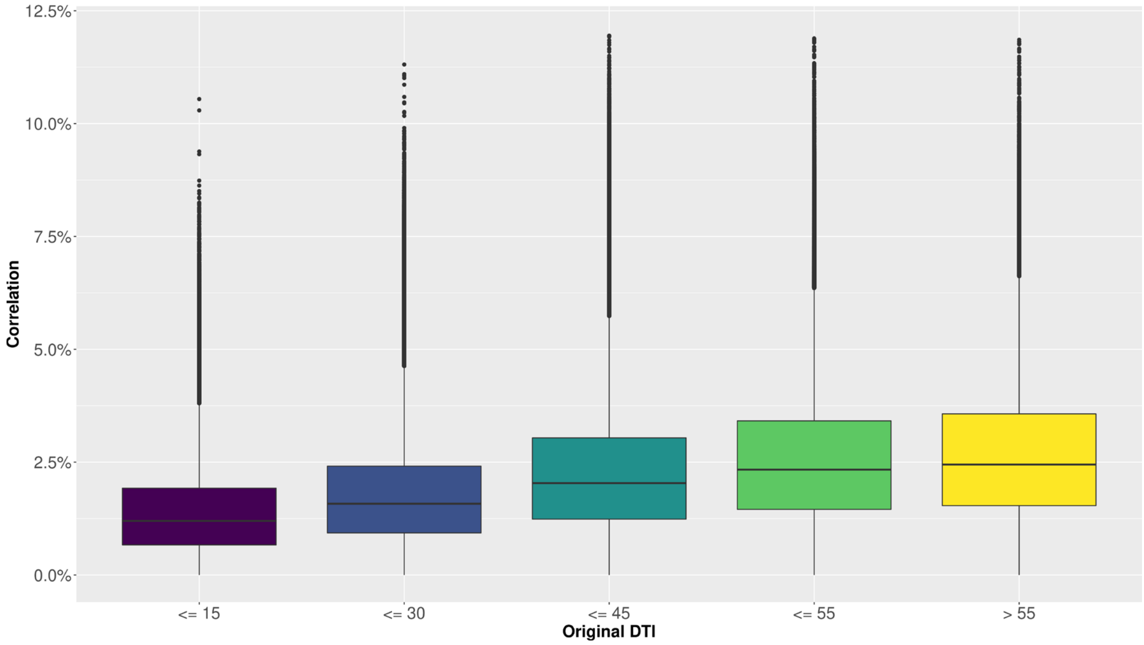


Figure A2: Correlation by Credit Score and Debt-to-Income

The graph displays correlation distribution by Credit Score (a) and Debt-to-Income (b). The correlations are calculated on the entire Freddie Mac cross-sectional sample, where each record is a mortgage at origination. The population is segmented on the x-axis by Lenders (a) and by increasing Balance buckets (b). Correlation distribution is plotted on the y-axis via box-plot. The box encloses the distribution from the 25th percentile (lower bound) to the 75th percentile (upper bound). The upper and lower whiskers represent correlation outside the middle 50% (i.e. the lower 25% and the upper 25% of correlations).



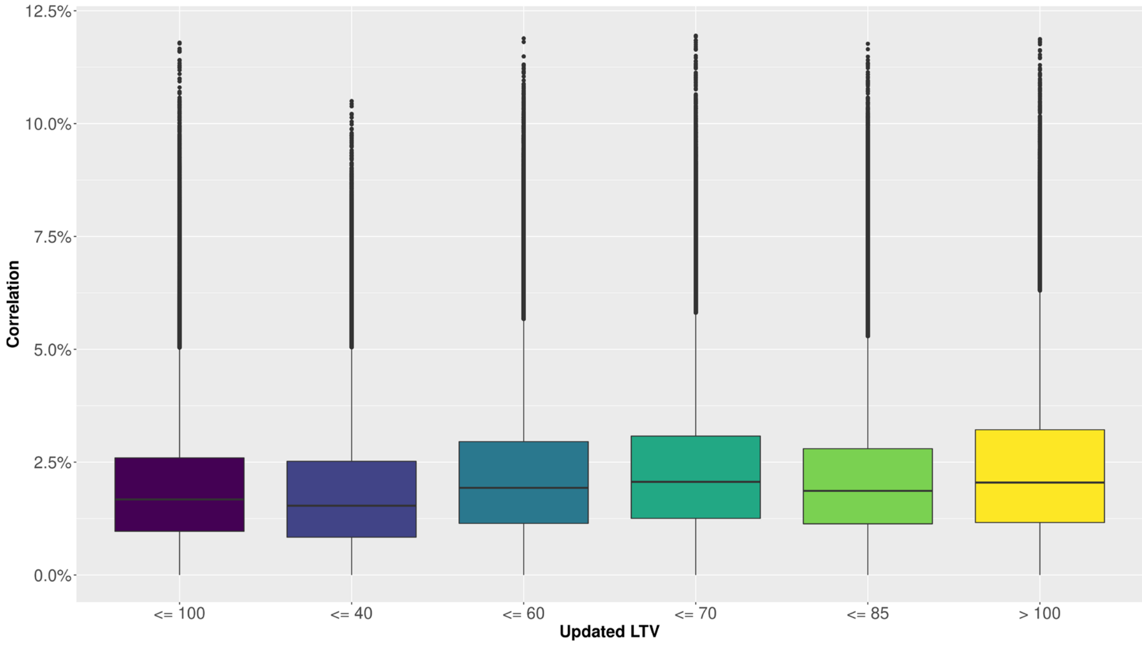
(a)



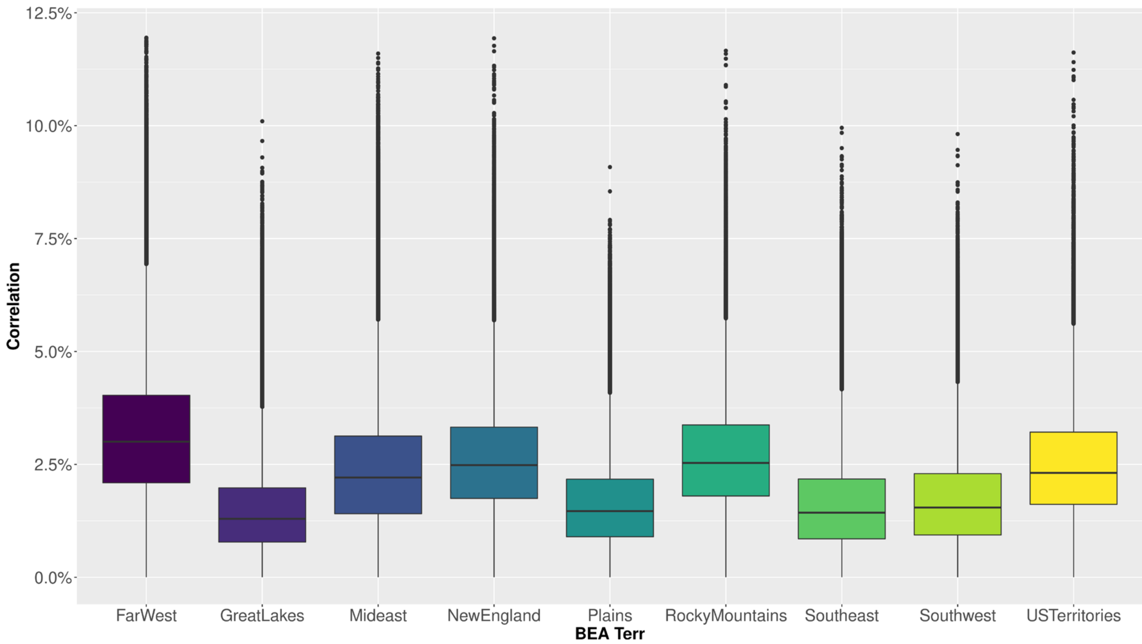
(b)

Figure A3: Correlation by Updated Loan-to-Value and Region

The graph displays correlation distribution by Updated Loan-to-Value (a) and Region (b) segments. The correlations are calculated on the entire Freddie Mac cross-sectional sample, where each record is a mortgage at origination. The population is segmented on the x-axis by Lenders (a) and by increasing Balance buckets (b). Correlation distribution is plotted on the y-axis via box-plot. The box encloses the distribution from the 25th percentile (lower bound) to the 75th percentile (upper bound). The upper and lower whiskers represent correlation outside the middle 50% (i.e. the lower 25% and the upper 25% of correlations).



(a)



(b)

Figure A4: Correlation by Non-Recourse

The graph displays correlation distribution by Recourse/Non-Recourse States. The correlations are calculated on the entire Freddie Mac cross-sectional sample, where each record is a mortgage at origination. The population is segmented on the x-axis by Recourse/Non-Recourse States. Correlation distribution is plotted on the y-axis using via box-plot. The box encloses the distribution from the 25th percentile (lower bound) to the 75th percentile (upper bound). The upper and lower whiskers represent correlation outside the middle 50% (i.e. the lower 25% and the upper 25% of correlations).

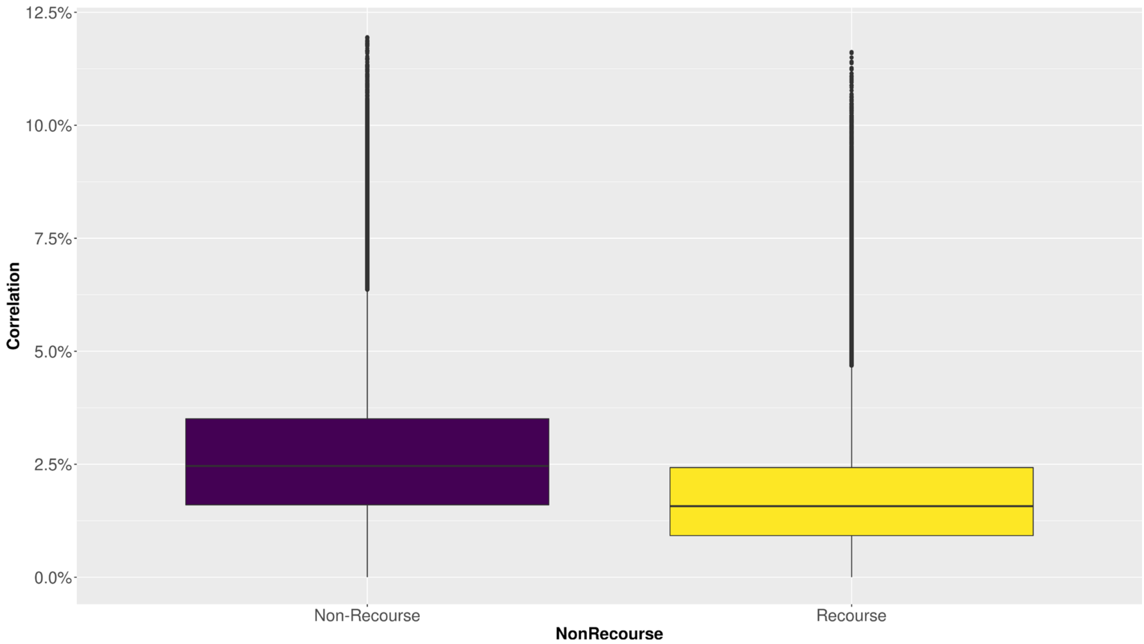


Figure A5: Excess Interests Paid by Region

The graph shows the maximum difference amongst banks in pricing the effect of mortgage correlation, by breaking down by Bureau of Economic Analysis (BEA) territories. The reference mortgage is a 30-Year Fixed-Rate Mortgage with original balance of $ 300,000 and 30-year Fixed-rate of 5.5 %. The isolated impact of correlation on Excess Interest rate is first calculated for each Bank; then the maximum difference amongst banks is produced and plotted. There is no other contributing factor to the difference in total interests paid. The graph highlights geographic region is not discriminatory factor when pricing mortgage correlation. This is in line with Hurst et al. (2016)

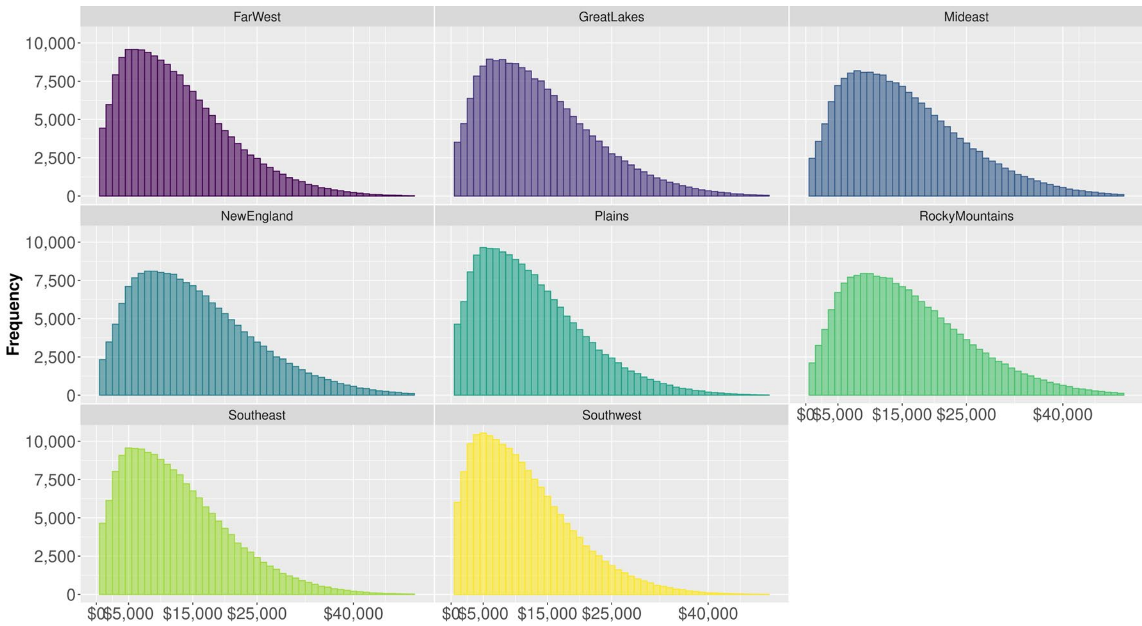


Figure A6: Excess Interests Paid by Original LTV

The graph shows the maximum difference amongst banks in pricing the effect of mortgage correlation, by breaking down by Loan-to-Value (LTV) at origination. The reference mortgage is a 30-Year Fixed-Rate Mortgage with original balance of $ 300,000 and 30-year Fixed-rate of 5.5 %. The isolated impact of correlation on Excess Interest rate is first calculated for each Bank; then the maximum difference amongst banks is produced and plotted. There is no other contributing factor to the difference in total interests paid. The graph highlights mortgage pricing discrepancy due correlation by increasing Original LTV. Financial institutions are aware that mortgages with higher LTVs experience contagion effect under adverse economic conditions and consequently price it, although not that differently. Borrowers might little benefit if they want to be priced differently based on their Original LTV.

Figure A7: Excess Interests Paid by Credit Score

The graph shows the maximum difference amongst banks in pricing the effect of mortgage correlation, by breaking down by Credit Score at origination. The reference mortgage is a 30-Year Fixed-Rate Mortgage with original balance of $ 300,000 and 30-year Fixed-rate of 5.5 %. The isolated impact of correlation on Excess Interest rate is first calculated for each Bank; then the maximum difference amongst banks is produced and plotted. There is no other contributing factor to the difference in total interests paid. The graph highlights that the lower the score, the higher the discrepancy amongst banks in pricing mortgage correlation. Borrowers might significantly benefit if they want to access lower interest charge based on their Credit Score.

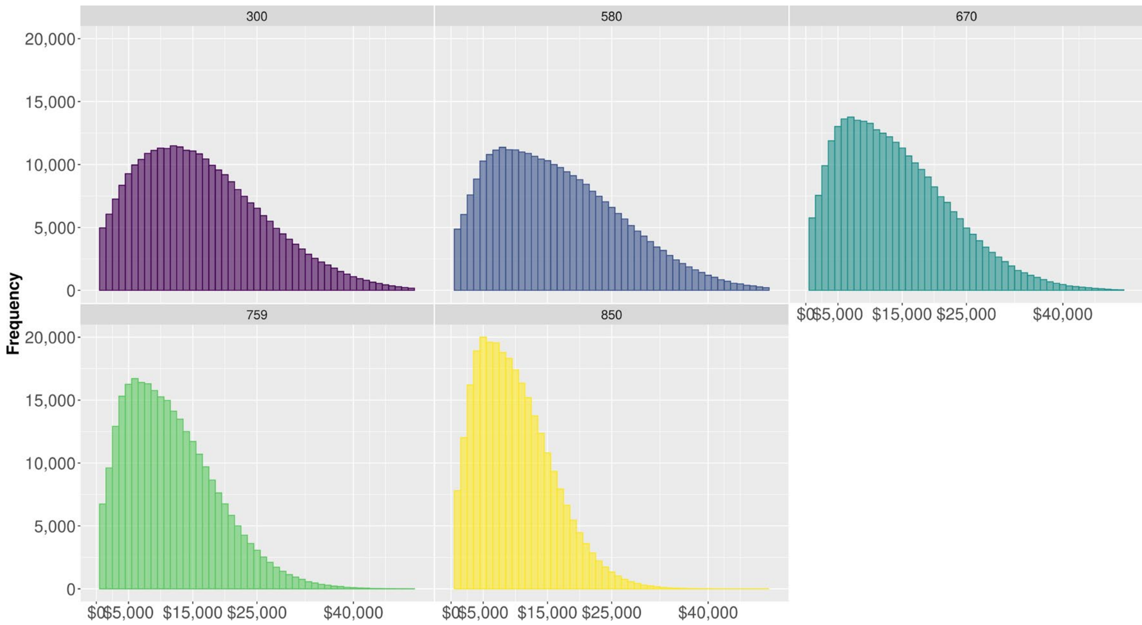


Figure A8: Excess Interests Paid by Debt-to-Income

The graph shows the maximum difference amongst banks in pricing the effect of mortgage correlation, by breaking down by Debt-to-Income (DTI) at origination. The reference mortgage is a 30-Year Fixed-Rate Mortgage with original balance of $ 300,000 and 30-year Fixed-rate of 5.5 %. The isolated impact of correlation on Excess Interest rate is first calculated for each Bank; then the maximum difference amongst banks is produced and plotted. There is no other contributing factor to the difference in total interests paid. The graph highlights that banks price mortgage correlation by increasing DTI at origination quite differently, especially when Debt-to-Income ratio is greater than 30 %. While banks are quite conservative and consistent in pricing mortgage default correlation for increasing LTVs, the same approach seems not to be followed for DTI. This implies that some financial institutions might underestimate mortgage correlation effect for borrowers having riskier profiles when it comes to their debt-to-income ratio.

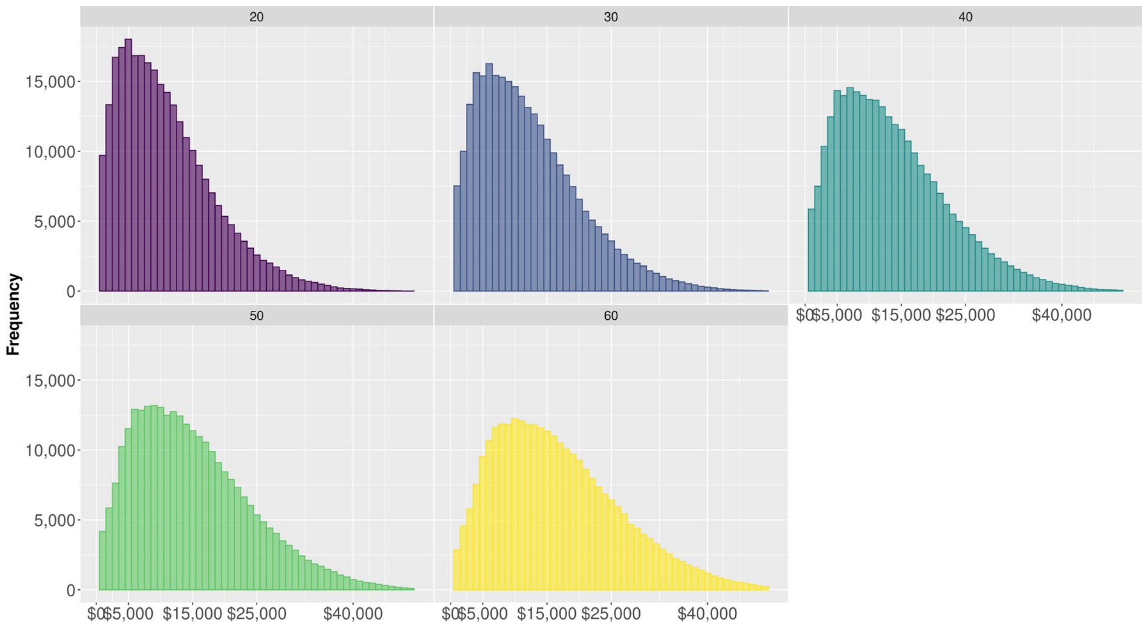


Table A1: Summary Statistics: Categorical Variables

The table reports percentage distribution by year of property and borrower types at origination: 1st Time Buyer; Occupancy: Investment (I), Primary (P), Second Home (S); Origination Channel: Broker (B), Correspondent (C), Retail (R), TPO Not Specified (T); Property Type: Condominium (CO), Co-op (CP), Manufactured Housing (MH), Planned Unit Development (PU), Single-Family (SF); Purpose: Cash-out Refinance (C), No Cash-out Refinance (N), Purchase (P); Number of Borrowers: Single (S), Joint (J)

		1stHomebuyer		Occupancy			Channel				Property					Purpose			N.Borrowers	
Year	N.Accounts	No	Yes	I	P	S	B	C	R	T	CO	CP	MH	PU	SF	C	N	P	1	2
1999	1,095,011	91.72	8.28	3.87	93.01	3.12	0.04	0.1	47.42	52.44	6.79	0.09	0.29	10.46	82.37	17.16	25.91	56.93	36.42	63.51
2000	786,275	82.16	17.84	4.98	90.92	4.10	0.05	0.03	49.02	50.91	8.31	0.13	0.45	12.99	78.11	11.24	13.41	75.34	39.54	60.41
2001	1,755,390	91.28	8.72	4.31	92.66	3.03	0.02	0.01	42.99	56.99	6.86	0.08	0.38	10.95	81.70	25.3	35.1	39.60	37.21	62.76
2002	1,682,997	92.02	7.98	4.46	92.19	3.35	0.04	0.03	42.97	56.96	6.8	0.11	0.56	10.33	82.20	26.5	37.04	36.46	38.03	61.95
2003	1,927,050	94.17	5.83	3.71	93.04	3.25	0.18	0.15	50.7	48.97	6.68	0.14	0.66	11.62	80.90	24.38	47.03	28.59	37.14	62.83
2004	1,127,674	90.89	9.11	4.34	91.08	4.59	0.09	0.3	44.62	54.99	7.24	0.42	1.07	13.88	77.37	25.39	28.1	46.51	41.74	58.23
2005	1,690,993	91.63	8.37	3.35	91.89	4.76	0.06	0.18	45.82	53.93	6.73	0.33	1.28	12.41	79.25	38.1	22.86	39.04	41.85	58.12
2006	1,260,783	89.35	10.65	4.68	90.08	5.25	0.03	0.05	40.09	59.83	8.12	0.37	1.63	14.44	75.44	36.83	16.33	46.84	44.55	55.41
2007	1,220,654	88.54	11.46	7.23	87.81	4.95	0.06	0.07	41.5	58.37	8.4	0.37	1.34	13.98	75.92	37.02	19.15	43.83	46.86	53.09
2008	1,179,578	89.85	10.15	7.82	86.92	5.26	10.19	16.95	45.73	27.12	7.99	0.39	0.58	16.48	74.55	33.45	28.27	38.29	46.52	53.43
2009	1,974,690	93.27	6.73	3.1	92.38	4.52	17.64	27.73	54.62	0.00	5.48	0.26	0.23	19.75	74.28	31.44	46.98	21.57	37.9	62.09
2010	1,271,397	91.13	8.87	4.98	90.54	4.48	11.46	38.79	49.74	0.00	5.24	0.21	0.24	19.64	74.67	31.1	41.29	27.62	38.19	61.81
2011	955,418	90.86	9.14	5.84	89.54	4.62	11.16	40.85	48	0.00	4.92	0.15	0.26	21.25	73.41	26.78	41.72	31.50	38.49	61.51
2012	1,331,301	92.33	7.67	5.25	90.76	3.99	10.14	37.34	52.52	0.00	4.52	0.11	0.27	22.47	72.62	23.21	50.56	26.23	37.92	62.08
2013	1,300,286	87.61	12.39	7.06	88.81	4.13	8.76	35.45	55.79	0.00	6.33	0.23	0.26	25.08	68.10	23.1	38.45	38.45	43.04	56.96
2014	975,881	79.97	20.03	7.88	88.13	3.99	9.92	34.46	55.62	0.00	7.47	0.17	0.3	27.19	64.87	20.12	20.91	58.97	48.19	51.81
2015	1,316,566	83.6	16.40	7.57	88.77	3.66	10.67	30.9	58.43	0.00	7.82	0.19	0.28	27.2	64.51	22.86	29.17	47.97	48.22	51.78
2016	1,558,394	84.84	15.16	7.77	88.84	3.39	10.06	31.39	58.55	0.00	8.11	0.16	0.27	27.57	63.89	24.42	30.57	45.01	49.42	50.58
2017	1,217,105	81.5	18.50	9.95	86.01	4.04	9.73	33.53	56.74	0.00	7.97	0.14	0.42	27.62	63.85	25.36	17.39	57.24	50.58	49.42

Table A2: Top Lenders

The table reports percentage distribution by year of origination of mortgage Sellers

Year	NAccounts	BankOfAmerica	BB&T	JPMorganChase	Citi	FifthThird	FlagStar	Other	Provident	SunTrust	UsBank	WellsFargo
1999	1,095,011	3.70	0.20	1.90	0.00	0.00	3.20	90.90	0.00	0.00	0.00	0.00
2000	786,275	10.20	1.80	3.30	0.00	0.60	0.00	61.70	0.00	0.00	0.00	22.50
2001	1,755,390	5.40	1.80	2.40	0.00	1.00	0.00	57.00	1.40	4.00	0.00	27.10
2002	1,682,997	5.30	2.00	1.30	0.00	1.60	0.00	50.90	1.90	1.70	3.10	32.20
2003	1,927,050	0.90	1.30	7.20	0.00	1.60	0.00	48.60	2.10	0.00	3.30	35.00
2004	1,127,674	0.00	1.50	14.60	0.00	1.50	0.00	46.00	0.30	0.30	3.80	32.00
2005	1,690,993	3.10	0.30	9.00	0.30	1.00	0.00	57.10	0.50	0.00	4.40	24.40
2006	1,260,783	3.30	1.30	8.40	2.20	1.70	0.80	49.40	1.20	0.80	5.20	25.70
2007	1,220,654	9.10	2.00	7.20	3.30	2.30	0.80	49.30	2.00	1.70	5.50	16.80
2008	1,179,578	7.60	3.90	9.90	4.80	2.20	1.70	44.00	0.60	0.30	7.30	17.70
2009	1,974,690	7.40	4.90	7.30	3.90	2.50	0.60	37.20	4.20	1.70	8.10	22.20
2010	1,271,397	11.80	5.80	5.50	3.70	2.20	0.20	36.40	2.60	0.60	10.90	20.30
2011	955,418	4.90	6.60	6.60	2.40	2.50	0.00	35.70	5.00	0.30	11.60	24.50
2012	1,331,301	3.60	6.50	5.20	0.30	2.50	0.00	38.40	5.90	0.00	14.00	23.50
2013	1,300,286	5.80	7.00	12.40	1.50	1.90	0.00	45.30	1.30	2.40	8.50	13.80
2014	975,881	7.00	4.90	6.60	1.60	0.70	0.00	59.50	0.20	1.30	6.30	12.10
2015	1,316,566	5.60	4.00	3.60	1.50	0.40	0.00	65.40	0.30	0.70	6.30	12.40
2016	1,558,394	3.90	3.00	5.20	0.60	0.00	0.00	65.20	0.60	1.30	4.60	15.50
2017	1,217,105	1.80	2.30	8.20	0.20	0.80	0.00	66.70	0.50	1.90	4.00	13.50

Table A3: Yearly Default Rate by Loan-to-Value

The table reports yearly default rate (expressed in percentage) by year of observation. Yearly default rate, number of observations and number of defaults at portfolio level are reported in the first three columns. The yearly default rate is then segmented by Loan-to-Value at origination (Original LTV) and Updated Loan-to-Value.

				Original LTV						Updated LTV					
Year	Accounts	Defaults	All	,; 30%	,; 50%	,; 70%	,; 80%	,; 90%	> 90%	,; 30%	,; 50%	,; 70%	,; 80%	,; 90%	> 90%
1999	939,197	718	0.08	0.0415	0.028	0.0438	0.0662	0.1275	0.11	0.0428	0.0265	0.0433	0.0692	0.1215	0.1186
2000	1,706,123	5,729	0.34	0.1023	0.1166	0.1712	0.271	0.5195	0.59	0.0958	0.1141	0.1889	0.3101	0.5859	0.6528
2001	3,292,855	16,049	0.49	0.1805	0.1535	0.2447	0.3661	0.8163	0.99	0.1901	0.1692	0.3632	0.419	1.0033	0.7113
2002	4,382,304	28,120	0.64	0.1322	0.1581	0.2839	0.4807	1.2205	1.46	0.1523	0.1824	0.43	0.6688	1.5213	1.6083
2003	5,311,191	35,279	0.66	0.1421	0.1394	0.2877	0.5155	1.3344	1.72	0.1994	0.2168	0.5125	0.726	1.556	1.236
2004	4,631,059	30,804	0.67	0.1023	0.1446	0.2635	0.5148	1.4233	1.83	0.1356	0.2545	0.5508	0.8067	1.8769	1.3792
2005	5,424,031	32,325	0.60	0.1188	0.1397	0.2851	0.4989	1.3462	1.69	0.1351	0.2822	0.6497	0.7492	1.4613	0.6996
2006	5,927,775	33,570	0.57	0.1264	0.1789	0.3323	0.5027	1.2308	1.52	0.1703	0.3212	0.6297	0.6412	1.0125	0.9294
2007	6,641,321	44,175	0.67	0.1654	0.2243	0.4352	0.623	1.3135	1.61	0.1862	0.3311	0.6584	0.8127	0.8862	1.3502
2008	7,359,285	93,166	1.27	0.2332	0.3802	0.8423	1.2563	2.334	2.95	0.2345	0.4873	0.8905	1.2707	1.5065	2.1287
2009	8,645,786	217,599	2.52	0.3254	0.6484	1.8049	2.7316	4.7452	5.53	0.2199	0.454	1.0165	1.4708	2.7914	7.5737
2010	8,370,112	195,879	2.34	0.3499	0.6862	1.7586	2.5861	4.3412	5.00	0.2368	0.5082	1.0691	1.5581	3.1207	7.8126
2011	7,803,125	134,082	1.72	0.2973	0.5599	1.2999	1.9232	3.1203	3.56	0.2038	0.4295	0.8787	1.1549	1.9976	5.3259
2012	7,776,197	103,176	1.33	0.255	0.4648	1.0072	1.4891	2.3772	2.64	0.1824	0.3821	0.7171	1.0767	2.4538	6.2198
2013	7,245,881	65,902	0.91	0.218	0.3638	0.7275	1.0062	1.5767	1.55	0.1799	0.422	0.742	1.1266	2.0439	4.2464
2014	6,844,299	45,048	0.66	0.2135	0.3115	0.5549	0.7166	1.0729	0.93	0.189	0.42	0.6908	0.8729	1.2738	1.7796
2015	7,515,347	32,781	0.44	0.1717	0.2389	0.3844	0.461	0.6452	0.60	0.1803	0.3639	0.5262	0.4662	0.5984	0.5479
2016	8,128,933	27,950	0.34	0.1521	0.2002	0.2963	0.359	0.4845	0.48	0.1707	0.3316	0.421	0.3328	0.3828	0.3267
2017	8,415,102	32,457	0.39	0.1594	0.2231	0.3269	0.3916	0.5248	0.60	0.2068	0.3843	0.4411	0.3961	0.5228	0.389
2018	7,705,902	23,463	0.30	0.1299	0.1791	0.2514	0.3077	0.3949	0.50	0.1508	0.2612	0.3303	0.4135	0.5362	0.9217

Table A4: Yearly Default Rate by Type of Property and Borrower

The table reports yearly default rate (expressed in percentage) by year of observation. Yearly default rate, number of observations and number of defaults ate portfolio level are reported in the first three columns. The yearly default rate is then segmented by 1st Time Homebuyer; Occupancy: Investment (I), Primary (P), Second Home (S); Origination Channel: Broker (B), Correspondent (C), Retail (R), TPO Not Specified (T); Property Type: Condominium (CO), Co-op (CP), Manufactured Housing (MH), Planned Unit Development (PU), Single-Family (SF); Purpose: Cash-out Refinance (C), No Cash-out Refinance (N), Purchase (P); Number of Borrowers: Single (S), Joint (J)

		1^{st} Time Buyer		Occupancy			Channel				Property					Purpose		
Year	All	No	Yes	I	P	S	B	C	R	T	CO	CP	MH	PU	SF	C	N	P
1999	0.08	0.0748	0.09	0.1268	0.0747	0.07	0	0.1548	0.0678	0.08	0.0656	0.1383	0.1124	0.0499	0.08	0.0699	0.102	0.0662
2000	0.34	0.3349	0.34	0.2996	0.3422	0.21	0.2016	0.6006	0.2557	0.41	0.2405	0.1156	1.0604	0.207	0.36	0.4077	0.4533	0.2819
2001	0.49	0.4772	0.56	0.5565	0.4916	0.28	1.4388	0.8706	0.3893	0.57	0.3217	0.0994	2.0832	0.3237	0.52	0.4678	0.5139	0.4814
2002	0.64	0.6311	0.73	0.9713	0.6371	0.33	2.0496	0.9943	0.4524	0.79	0.4293	0.3266	2.3181	0.4218	0.68	0.5649	0.6633	0.6656
2003	0.66	0.6476	0.84	1.0695	0.6562	0.35	1.1952	1.5873	0.4455	0.86	0.3598	0.1896	2.0569	0.408	0.71	0.6136	0.6197	0.7436
2004	0.67	0.6442	0.91	0.8106	0.6718	0.31	0.7273	2.5641	0.4763	0.84	0.3601	0.1305	2.1229	0.3707	0.72	0.6359	0.6355	0.7158
2005	0.60	0.5869	0.70	0.7328	0.6034	0.28	1.6908	1.9231	0.4369	0.74	0.3187	0.1544	1.5665	0.292	0.65	0.5401	0.6478	0.5906
2006	0.57	0.5565	0.67	0.6344	0.5654	0.52	1.4925	1.2195	0.4605	0.66	0.57	0.1756	1.15	0.3274	0.60	0.5337	0.6242	0.5468
2007	0.67	0.656	0.76	0.6003	0.6814	0.40	1.0274	0.6803	0.5316	0.78	0.5147	0.2567	1.2593	0.4447	0.71	0.7255	0.6843	0.6047
2008	1.27	1.25	1.42	1.3962	1.27	1.05	0.2547	0.0945	0.971	1.61	1.2845	0.4069	2.0553	1.127	1.28	1.4852	1.1131	1.1983
2009	2.52	2.4963	2.73	3.0695	2.5183	1.93	0.9771	0.5657	1.8519	3.95	2.7131	0.7879	3.8923	2.3078	2.53	3.2673	1.8847	2.4245
2010	2.34	2.3153	2.60	2.6351	2.3554	1.74	1.0891	0.5602	1.7989	4.20	2.701	0.7313	4.0692	2.0779	2.35	3.0194	1.6717	2.374
2011	1.72	1.7008	1.90	2.0052	1.7188	1.38	0.7759	0.5164	1.3706	3.45	2.1338	0.7136	3.1225	1.4511	1.72	2.1903	1.2011	1.8185
2012	1.33	1.3054	1.55	1.4061	1.3357	1.08	0.5999	0.3763	1.0962	3.15	1.7693	0.838	2.7906	0.9972	1.35	1.7434	0.8827	1.4563
2013	0.91	0.8993	1.01	0.8923	0.9221	0.70	0.47	0.3032	0.7576	2.63	1.085	0.8984	2.1207	0.5523	0.97	1.2452	0.6261	0.9423
2014	0.66	0.6587	0.65	0.5978	0.6725	0.47	0.368	0.2736	0.561	2.31	0.6925	0.6947	1.7062	0.3342	0.73	0.9314	0.4895	0.6179
2015	0.44	0.4364	0.43	0.3906	0.4452	0.32	0.2607	0.2306	0.3698	1.80	0.441	0.5357	1.2892	0.2176	0.49	0.6162	0.3311	0.4069
2016	0.34	0.3386	0.38	0.305	0.351	0.26	0.2353	0.2203	0.2956	1.55	0.3162	0.3887	1.1024	0.1943	0.39	0.4597	0.2613	0.339
2017	0.39	0.3716	0.47	0.3095	0.3986	0.25	0.3466	0.3065	0.3343	1.57	0.3263	0.375	0.9339	0.3414	0.40	0.4847	0.2862	0.4004
2018	0.30	0.2875	0.41	0.2334	0.3143	0.22	0.3474	0.2405	0.2907	0.85	0.2963	0.2382	0.4864	0.3419	0.29	0.3832	0.1988	0.3363

Table A5: HMDA Representativeness

The table shows mortgage applications breakdown across the United States sourced by HMDA (Home Mortgage Disclosure Act) from 2007 to 2017. The Home Mortgage Disclosure Act (HMDA) requires financial institutions to maintain, report, and publicly disclose loan-level information about mortgages, hence providing a reliable source for mortgage market dynamics. The table shows the breakdown of mortgage applications and originations, with a particular focus on conventional loans issued by Fannie Mae (FNMA) and Freddie Mac (FHLMC). Freddie Mac mortgages cover 25.6 % of total conventional originated mortgages and 17.7 % of total mortgage originations. The sample used for the analysis in this paper relates to conventional originated mortgages.

Data	Percentage	Volumes
Total Mortgage Applications		187,462,446
Total Mortgages Originated		90,171,323
	% Total Applications	48.1 %
Conventional Originated		62,317,732
	% Total Originated	69.1 %
FHLMC and FNMA Originated		41,550,067
	% Total Originated	46.1 %
FHLMC and FNMA Conventional Originated		40,849,709
	% Total Conventional Originated	65.6 %
FHLMC Conventional Originated		15,976,438
	% Total Conventional Originated	25.6 %
	% Total Originated	17.7 %