Correlation and Residential Mortgage Defaults

Chiara Maria Ventura, Simone Varotto

January 14, 2023

Abstract

We employ a sample of 25 million US mortgages originated between 1999 and 2017 to compute pairwise mortgage correlations implied from mortgage defaults. We find that the flat correlation of 15% adopted by bank regulators does not adequately capture its wide variability in US mortgage portfolios. Such variability is mainly driven by the borrower's credit score and geographical location. Other borrower's and loan's characteristics also play an important role. Moreover, we find evidence that borrowers could make average savings of $4,708 with a standard deviation of $4,400 on a standard mortgage by "shopping around", as lenders may charge economically different rates to the same customer to reflect (1) differences in their mortgage portfolio composition and (2) differences in the extent to which they account for correlation risk.

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

1 Introduction

The US mortgage market has played a critical role in major financial crises of 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 are characterised by a highly correlated behaviour of borrowers which leads to a substantial increase in mortgage defaults. In this study, we analyse the factors behind the increase in pairwise correlations in mortgage portfolios by exploiting a substantial loan-level database that covers the Great Recession. Our work contributes to the 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 computed mortgage correlations at an aggregate or portfolio level from MBS prices (Geidosch (2014)) or available time series (Nickerson and Griffin (2017), Botha and van Vuuren (2010), Stoffberg and Vuuren (2015))), or from niche 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 wider sample representative of US mortgage market. Our estimation strategy employs the intuition that correlation is what drives the difference between long run default probabilities (PD-LR) and default probabilities in a crisis (PD-Crisis). With a logit model we estimate both probabilities by using the Great Recession as a benchmark crisis period. Then, by exploiting the correlation-driven relationship between PD-LR and PD-Crisis based on a popular model adopted by bank regulators (BCBS (2005),Blumke (2018)), we determine the implied average pairwise correlation of any given borrower and the other borrowers in the lender's mortgage portfolio. The advantage of this approach is that it allows us to determine how borrower and loan characteristics can influence mortgage portfolio correlations in a crisis. We find that mortgage correlations are influenced primarily by the borrower's geographical location, credit score and loan-to-value.

Third, we investigate if banks price correlation risk in the interest rates applied to mortgage borrowers. New borrowers that are more (less) correlated with the existing borrowers in a bank's portfolio should be charged a higher (lower) interest by the bank to compensate the lender for the higher (lower) risk of joint default in its mortgage portfolio in a crisis. We find that while some borrowers charge a positive premium for correlation risk (Wells Fargo, US Bank, Sun Trust) others apply a negative premium (JP Morgan Chase, Citi and Provident). We conjecture that a negative premium may be the result of (1) greater market competition that pushes interests down and decouple them from portfolio concentration considerations, (2) aggressive expansion strategy by the lender to increase market share in a given market segment which would lead to the same outcome as in point (1), (3) portfolio correlation risk not being considered because mortgages would be securitised and skin-in-the-game provisions (Fuster et al. (2022) and Krahnen and Wilde (2022)) fail to generate the incentive for some banks to align mortgage prices to correlation risk. Such correlation-price connection may also not be justified as Freddie Mac (Federal Home Loan Mortgage Corporation) and other agencies combine in the same securitization mortgages from different banks. This potentially increases the diversification of the underlying pool of loans relative to the diversification in the portfolio of the originators.

A corollary of the above is that the sensitivity to correlation risk varies significantly among banks. We find that the implication for borrowers is that they could make average savings of \$4,708 with a standard deviation of \$4,400 on a standard mortgage by "shopping around". This is because lenders may charge economically different rates to the same customer not only to reflect differences in the extent to which they account for correlation risk, but also because of differences in their mortgage portfolio composition and the marginal contribution made by a new borrower to the bank's portfolio concentration.

While the literature on correlation in the context of corporate exposures is extensive (Adams et al. (2017), Driessen et al. (2009), Longin and Solnik (2001), Chernih et al. (2006), Gordy (2000), Blumke (2018)), few studies have investigated the correlation in portfolios of retail exposures. The lack for market prices for retail exposures implies that mortgage correlations at the loan level have to be calculated with default/loss data. Cowan and Cowan (2004) are the first to adopt this approach. We extend their analysis by considering a wider sample that includes the Great Recession and by adopting a different methodology and a more extensive database. Our data includes 25 million mortgages issued from 1999 to 2017 across the Unites States. The data that is made publicly available by

Freddie Mac.

Our paper is organised as follows. In Section 2 we review the relevant literature. Section 3 includes a description of the data. In Section 4 we explain our methodology while Section 5 discusses our result. Section 6 concludes the paper.

2 Literature Review

In this Section we review the literature related to mortgage correlation as well as relevant research in the corporate loan market.

The hit of Great Financial Crisis raised questions on the validity of correlation values and on the methodological assumptions set by eitherBCBS (2005) or alternative risk assessment frameworks. Literature and studies on this topic grew bigger, with a particular focus on corporate asset classes or securities (Nickerson and Griffin (2017), Chamizo et al. (2019), Chernih et al. (2010) and 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. The authors also argue that a commonly assumed lesson from the financial crisis is that default correlations were not well understood, and despite this available period of massive default, not much academic work has been done to understand it for structured products. Similarly, Chamizo et al. (2019) start from pointing 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. (2010), who compare asset correlations calculated on monthly asset value with both Basel II and previous literature. They generally find that their results are in line with previous literature, while a clear difference arises in comparison with Basel II and 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 within different asset classes, as the Great Financial Crisis highlighted a flaw in risk assessment frameworks to correctly measure contagion effect in financial markets. Nonetheless, mortgage correlation studies are quite limited in the current literature despite the relevance of this asset class in banking books and securitised markets.

Most of quoted literature is focused on corporate portfolios, while instead little investigation has been carried out on mortgages. In general, a misconception on residential mortgages correlation can arise i.e. that it would not variate much and that the value set by BCBS (2005) can be considered for any capital calculation (also internal capital allocation). Assuming that the conservativism is well-proven (despite being questioned by Hull (2015)), there is not much evidence in the existing literature on the accuracy of a flat correlation value for residential mortgages. This is also caused by the difficulty of measuring mortgage correlation, as it is not straightforward to derive its asset value. The study of Duellmann et al. (2010), for example, examines if it is better to estimate asset correlation from stock prices or default rates, reaching the conclusion that, whenever time series of market prices are available, it is advisable to use stock prices instead of default rates, as these latter yield a underestimation and are generally characterised by scarce data. However, only one option is possible for mortgage exposures (i.e. rely on default data).

Following the Great Financial Crisis, augmented performance of mortgage portfolio has been collected, becoming the starting point of increased analyses. Interesting research was carried out by Gupta and Hansman (2022), Goodstein et al. (2017) and Mian and Sufi (2009). Gupta and Hansman (2022) analyses the mortgage market with a focus on the connection between leverage and default. In particular, the authors investigate the defaulting behaviour of highly leveraged borrowers when house prices fall and separate moral hazard (i.e. leverage increases the probability of default) from adverse selection (i.e. risky borrowers prefer high-leverage mortgages). While we cannot separate these two triggers, we similarly highlight the effect of Updated LTV on default contagion and, as opposite, we also corroborate the relevance of other factors such as FICO, purpose etc. which are instead not verified by Gupta and Hansman (2022). Such difference might be driven by the sample used, as only non-agency options ARMs are deployed for Gupta and Hansman (2022) analysis, while instead the data employed in our research is more representative of US mortgage market.

Goodstein et al. (2017) and Mian and Sufi (2009) investigate mortgage credit and its risk at ZIP code level.Goodstein et al. (2017) analyse the contagion effect among strategic defaulters as a consequence of increasing delinquency in the same ZIP code area. Again, strategic defaulters are linked to negative equity as in Gupta and Hansman (2022), even though the authors make a different step by diversifying strategic defaulters by other borrowers. Similarly, we investigate the mortgage contagion implied by default experience, even though we do not solely focus on strategic/not strategic behaviour because we adopt the lender perspective, blind at this end, and we control for a wider set of covariates to estimate correlation. On the other hand, Mian and Sufi (2009) link the surge in GFC mortgage defaults to disproportionate lending to subprime ZIP codes. Therefore, the crisis can be also explained with credit expansion to risky borrowers. Similarly, in our research we 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. As opposed to Mian and Sufi (2009), who aggregate default rates by Zip codes, we do not carry out any aggregation process, while instead we preserve the unique combination of mortgage characteristics at borrower level thanks to the large coverage of our sample. Second, the authors do not effectively quantify the difference/discrepancy conditional on other drivers, which instead we consider when calculating correlation.

Additional studies have examined the dependency of correlation from firms' characteristics. For example, Lopez (2004) investigates the empirical relationship between average asset correlation, firm's probability of default and asset size. While still within the corporate world, the conclusions reached by the authors are interesting and relevant for our research, as they adopt a similar framework to ours. The authors find that average asset correlation is a decreasing function of probability of default. We reach a similar conclusion in the mortgage market. Second, the empirical results indicate that average asset correlation is increasing in asset size. That is, as firms increase the book value of their assets, the correlation with economic environment increases. This result is intuitive in the sense that larger firms can generally be viewed as portfolios of smaller firms, and such portfolios would be relatively more sensitive to common risks than to idiosyncratic risks. While our case is somehow different, we demonstrate that sub-portfolios of mortgages can behave quite differently in relation to the systemic risk factor. A comparable conclusion is reached by Duellmann and Scheule (2003), who explore asset correlation and its dependency on firm size and probability of default, finding a significant relation with both.

The work of Tarashev (2010) is centred on parameters uncertainty, and he find that higher levels of PD and asset-return correlations are generally associated with greater noise in VaR. Such analysis is developed by using Bayesian inference, reaching the conclusion that the impact of parameters uncertainty is strong for a wide range of portfolio characteristics. This has a strong parallelism with our analysis, where we empirically demonstrate that beyond prior assumptions, correlation is effectively a characteristic-dependant parameter.

As mentioned earlier, correlation value set in the regulatory capital is equal to 15% for residential mortgages (this value is derived by Calem and Follain (2003)) and to 4% for credit cards as in BCBS (2005). Part of existing literature on retail asset classes focuses on testing the accuracy of this value and often reaches the conclusion that it tends to be fairly conservative. See for example, Botha and van Vuuren (2010) who study charge-off information loss data derived from the 100 largest US banks, and Crook and Bellotti (2009), who analyse UK credit cards. This is also in line with results on US credit cards obtained by Rösch and Scheule (2004). Geidosch (2014) investigates asset correlation of residential mortgages using RMBS data and including toxic RMBS deals. The author leverages different estimation methodologies (SFGC, methods of moments, maximum likelihood estimation, parametric approach), again reaching to the conclusion that inferred correlation is surprisingly low if compared with Basel parameter, despite having included extremely low-quality deals. On the other hand, Neumann (2018) uses UK and US loss data to infer residential mortgage correlation via using multiple estimators and conclude that Basel 15% parameter is at the appropriate level. As opposed to Geidosch (2014) and Neumann (2018), our methodology relies on popular copula models to extract correlations from default data (as in Lee et al. (2021)), even though part of literature has found these models faulty (Egami and Kevkhishvili (2017)). Nonetheless, being aware of the limitations of copula models to compute correlation, we use it to compute a correlation indicator to show its heterogeneity and sensitivity to portfolio composition.

3 Data

On March 2013, Freddie Mac (Federal Home Loan Mortgage Corporation (FHLMC (2022)) released loan-level credit performance data on a majority of the fully amortizing fixed rate, single-family mortgages it had purchased since 1999. The aim of such disclosure is to allow investors and researchers to build more accurate estimates and models, guaranteeing increasing transparency in the sector. This is the data that we leverage for our analyses; it is composed by single family residential mortgages issued by a multiplicity of entities (more than 100) and then acquired by Freddie Mac for securitisation purposes. The mortgages in the sample are originated from the first quarter of 1999 up to the end of 2017; all the facilities are followed up from origination to the latest reporting date, which is the second quarter of 2018. The portfolio peculiarity is its being "live"; the book is constantly fed with new acquisitions and the performance of existing loans is quarterly updated as well.

Throughout the entire period covered in our research, the size of the sample builds up to c. 25 millions units (unique Mortgage IDs). Freddie Mac acquisitions have kept a constant growth even after the financial crisis, as it is observed in the non-decreasing pattern of originated accounts after 2009. Moreover, the granularity of mortgage originators makes the sample extremely capillary in terms of US geographical composition, as shown in Figure 1. In line with demographic distribution, states like California (> 3million), Florida, Texas and Illinois (>1 million) have a higher share of mortgages within the sample. However, this does not prevent less-populated states to have a significant number of facilities. For example, despite having the lower share of mortgages after excluding US territories like Puerto Rico, Guam and Virgin Islands, the state of Wyoming counts c. 44k originated loans.

For each purchased mortgage, information on both origination and performance is collected, tracked and stored in Origination and Performance files respectively. The origination file includes borrower, property and mortgage related characteristic at origination (e.g. credit score, first time home-buyer flag, debt to income ratio, occupancy status, loan-to-value and interest rate). Table 1 reports key distribution statistics of Credit Score, Loan-to-Value, Debt-to-income, Interest Rate and Balance at origination. It seems, indeed, that the year of origination is determinant in the issuance of mortgages and captures market breaks. For example, Table 1 shows the average quantiles evolution of Credit score and Debt-to-Income ratio by year of origination. The GFC structural break is highlighted by the average increase/decrease of Credit Score and Debt-to-Income respectively, which mirrors the change due to stricter underwriting standards. Following the drop, the aftermath of Global Financial Crisis is characterised by a stable distribution. Likewise, average Loan-to-Value ratio and Interest Rate disclose relevant information on origination patterns. The GFC yields a decrease in the average LTV right after 2009, even if there is a recent reverse trend mainly caused by the implementation of supporting schemes for homebuyers. On the other hand, Interest Rate at origination follows market rate trends with minor fluctuations, as shown by the small standard deviation. We'll focus on this pattern when analysing Excess Interest rate.

While continuous variables are influenced by economic cycle at time of origination, some other characteristics remain quite homogeneous throughout the years, as shown by the Occupancy Status distribution in Table 2. The majority of borrowers purchase a primary residence and a smaller share buys investment or second homes. The pattern of Loan Purpose, instead, shows an interesting increase in refinance mortgages right after the GFC, most likely explained by decreasing interest rates. Channel variables is characterised by a hard break in TPO non-specified mortgages right after the financial crisis, due to increased transparency and stricter reporting criteria. All the other mortgage characteristics are evenly distributed by year of origination, with the sole exception of Planned Unit Development property type, whose share in the mortgage market is steadily increasing.
As each loan's performance is monthly monitored since origination, current delinquency status, interest rate, remaining months to maturity and unpaid balance are consistently updated throughout the entire lifetime of the loan. Availability of performance variables helps us to track the evolution of each mortgage's credit history and collateral information. For example, knowing Property State we track the changes in House Prices at state level to update property value and derive Updated LTV from the outstanding unpaid balance. Likewise, we calculate the age of the loan and follow its lifecycle from origination to the latest available observation.

Amongst performance variables, repayment information is key to derive the default flag. We focus on two indicators that track the repayment performance of each facility. The first is the "zero balance code" flag, which signals the reason for a loan's balance being reduced to zero (charge-off, REO acquisition, repurchase prior to property disposition and third party sale). The second is "current loan delinquency status", which corresponds to the number of days the borrower has been delinquent. Both variables are used to identify risky customers and default is triggered at the first occurrence of either 90-days delinquency or zero-balance code indicator being populated. This is in line with the recently updated regulatory definition of default. We consider first default occurrence as a termination event, hence we remove any observation after first default observation.

Figure 2 and Figure 3 display two different aspects of default occurrence within the data. Figures 2 shows that the actual surge in defaults is slightly delayed from the start of the 2008 crisis. We therefore consider as the real mortgage crisis the years spanning from 2009 until 2011. Figures 3 displays number of mortgagees by year of origination, highlighting the relevant share of mortgages originated just before the crisis having a richer delinquency experience. This is a combined effect of the hit of financial downturn and mortgage lifecycle, as default rate is typically higher in the first 5 years since origination. Both these elements are controlled for when running the logistic regression, to avoid that any bias is transmitted to other drivers. Overall, based on the default definition above outlined, 4.68 % facilities defaulted within the time frame considered in our data.
Table 5 to Table 7 show yearly default rate across key characteristics, and we can already observe that not all the segments are affected by the GFC in the same way. For example, Debt-to-Income and Excess Interest Rate rank order by implied riskiness the yearly default rate before and during the crisis. Credit score has a similar pattern for the most populated segments (i.e. where Credit Score is greater than 550), while the subprime segment default rate is less influenced by the business cycle, most likely explained by lower concentrations. Original LTV and Updated LTV default rates (Table 6) are aligned with economic intuition, and it is worth noticing the spike in Updated LTV default rate for underwater mortgages. This is the main reasons that supports the choice of incorporating Updated LTV rather than LTV at origination when estimating the logistic regression, as we deemed more accurate to grasp interaction of key variables with economic environment. Finally, Table 7 splits default rate by main categorical variables. First Time Homebuyer flag, Number of borrowers, Occupancy yearly default rates are well separated between categories, implying that riskier buckets have consistently higher default rates throughout the entire series; sensitivity to economic downturn is obviously different and this will be object of study for correlation. Loan purpose categories are instead characterised by very similar default rates before GFC crisis, which instead separate in economic downturn and remain differentiated even after, with Cash-Out refinance becoming the riskier. Property types have a less defined behaviour, especially for Manufactured Housing and Co-op; we decide to keep these variables in the regressions and test their significance wihtout merging into a single category. Finally, we observe that Channel default rate patterns have an erratic behaviour, mainly due to the structural break in its buckets population. We decide to remove this variable from logistic regression estimates, but to keep it for the Excess Interest rate regression because in this latter only loans originated after 2011 are considered, where the categories are already consolidated.

Figure 4 displays the ratio between average yearly default rate before and after the crisis by State. Only few States experienced a default rate that less than doubled during the crisis, while mortgage default rate was at least three times than in non-crisis period for the majority of US Sates. California, Nevada, Florida, Arizona mortgage default rate was six-times than observed in non-crisis. Among these, only California and Arizona are mortgage non-recourse states.

Having explored the distribution of the data source considered, an important aspect of our research is the claim of representativeness of US mortgage market. This shifts from Cowan and Cowan (2004), whose analyses are based on a single subprime lender. Data representativeness is ensured by comparing the sample used for our analyses with HMDA database (Consumer Finance Protection Bureau (CFPB (2022)). HMDA database offers the right instrument to assess the representativeness of our sample, being the most comprehensive source of publicly available information on 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. Even if HMDA data does not have a full coverage of US mortgage market, it is the widest publicly available loan-level source and helps us to gain an accurate view of our data coverage.

Table 4 displays the number of applications and originated loans over time. Out of c. 187 million mortgage applications from 2007 to 2017, 48.1% of the mortgages have been effectively taken-up. The predominant portion is Conventional loans (69.1%), the most common loan type in the US mortgage market. Conventional mortgages are not directly insured by the US Government (differently from FHA-insured, FSA/RHS-guaranteed and VA-guaranteed), while instead they are either kept on Banks' balance sheet or taken-up by GSEs business (i.e. Freddie Mac and Fannie Mae) that predominantly operate into this category. Fannie Mae and Freddie Mac buy (mostly) conventional loans not insured by the government (46.1%), and further filter their business by setting guidelines (conformity rules) that depository and non-depository lenders have to follow in order to sell the loans. 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. The conformity rules set by Freddie Mac and Fannie Mae, although not exactly overlapping, definitively contribute to shape the acceptance/rejection mechanism within the broader mortgage market.
Even if there is not a specific market split between the two entities, it is common knowledge that Freddie Mac used to focus its business towards smaller banks and thrifts, while Fannie Mae mainly purchased mortgages from larger commercial banks and bigger institutions. However, the post financial crisis mortgage market saw a large number of mergers and acquisitions; this split in business is therefore less applicable and in any case does not really influence the scope of our research. While Fannie Mae mortgage volumes are higher than Freddie Mac, Table 4 shows that Freddie Mac share on conventional loans is (overall) around 25%, which is still a relevant portion of US mortgage market analysed by our study.

4 Empirical Methodology

4.1 Correlation

Given the granular nature of the data, we do not have methodological restrictions to extrapolate default correlation from our sample. We initially relied on pooled approach in line with Botha and van Vuuren (2010), where the sample is segmented by different characteristics and default rate series are then calculated for each pool. Once the default rate series are constructed (either unidimensional or multidimensional), second step for the pooled methodology is the extrapolation of mortgage default correlation. On this scope, we apply the framework of Botha and van Vuuren (2010) who examine the extraction of retail asset correlations, assess robustness of the methods and compare implied correlations with BCBS (2005) specifications. The paper introduces two different distributions (Vasicek and beta) to extract empirical correlation from gross loss data time series at portfolio level. Pooled approach has been initially very helpful to skim the main drivers and effectively understand the soundness in variability of correlation; however, such methodology is affected by volatile results when segmented default rate time series are calculated on low volumes, bringing to unstable results in correlation, especially for multi-dimensional analysis.

We have therefore decided to privilege an alternative approach, where loan-level available performance is used ensuring a multidimensional view (hence, a non-flat correlation) without de-stabilising the estimations. Loan-level estimates are based on a panel-logit discrete hazard model, where long-run and a downturn PDs are calculated for each loan and then used to compute default correlation. Aim of the selected approach is to exploit the sample in its fullness and not to bias the estimation of correlation ρ_i via data pooling, as the large volumes we are dealing with ensure confidence in the results presented.

Before any regression is performed, the panel loan-level dataset is built so that one-year PDs can be estimated, in line with BCBS (2005) requirements. Each loan's yearly performance is tracked through time and the target variable default flag (0/1) is assigned in each year depending on the delinquency status of the loan at the end of the year considered. Default flag is triggered following the definition of default introduced earlier. Explanatory variables for each loan include characteristics at origination (e.g. Credit Score, Loan-to-Value, State), which we cannot track after the loan has been securitised, and time-varying characteristics (e.g. Loan Age, Updated LTV and macroeconomic variables). Given the large amount of available data and the richness of mortgage drivers, we maximise the usage of explanatory variables as long as significance and alignment with economic intuition are granted. It is worth highlighting that none of the continuous variables are segmented and the raw values are used, leaving no room to potential bias caused by additional segmentation. Once the panel loan-level dataset is built, multi-period logit model is estimated as per Equation 3:

$$L(\alpha;\beta_1,\cdots,\beta_{N_b};\gamma;\delta_1,\cdots,\lambda_{N_l}) = \tag{1}$$

$$= \prod_{i=1,t=t_{t0}}^{N,T} \pi^{y_{i,t}}(1-\pi)^{1-y_{i,t}} = \tag{2}$$

$$= \prod_{i=1,t=t_{t0}}^{N,T} \left(\frac{exp^{(\alpha+\sum_{b=1}^{N_b}\beta_b x_{b|it}+\sum_{m=1}^{N_m}\mu_m x_{m|t}+\gamma d_{crisis|t}+d_{crisis,t}\sum_{l=1}^{N_l}\lambda_l x_{l|it})}}{1+exp^{(\alpha+\sum_{b=1}^{N_b}\beta_b x_{b|it}+\sum_{m=1}^{N_m}\mu_m x_{m|t}+\gamma d_{crisis,t}+d_{crisis|t}\sum_{l=1}^{N_l}\lambda_l x_{l|it})}}\right) \tag{3}$$

where π is the estimated probability of default and $y_{i,t}$ takes values 1 or 0 whether the borrower i in $(1, \cdots, N)$ defaults or not in year t $(t0, \cdots, T)$. On the right hand side we find the explanatory variables and their related coefficients. Beside the intercept α, the coefficients β_i (i in 1, $\cdots$, N_b) capture mortgage characteristics. Dummy $d_{crisis|t}$ captures the effect of the great Financial Crisis, and is made interact with the same set of mortgage variables, whose coefficients are λ_i (i in 1, $\cdots$, N_l). In doing so, we are able to separate the downturn predicted PDs from long-run values. The coefficients β_b capture the effect of macroeconomic trends. The dummy $d_{crisis|t}$ is activated for the years running from 2009 to 2011 included, as we have observed in the data that the effect of the financial crisis on the mortgage market was not immediate.

Panel-logit discrete hazard model is run over different trials, which have been tested across a heterogeneity of explanatory variables and selected samples. For instance, the first models have been tested on the entire sample, where missing information has been kept as a separate category. This approach has been later discarded and the development sample has been cleaned from records with missing information (especially on missing debt-to-income, credit score and LTV), having ensured that the estimated values had little fluctuation. This methodological choice is supported by the higher (almost total) incidence of missing information (especially on Debt-to-Income) in the years before the crisis, when loosening standards in mortgage applications were applied. As such practices are no longer allowed, missing records are removed from the sample to avoid any bias in the estimations that could invalidate future applicability, as Excess Interest Rate regression. This is also motivates the exclusion of Channel, whose distribution reflects the break in reporting criteria implemented after the crisis. Macroeconomic variables have been tested both interacted and not-interacted with the dummy crisis, but to avoid any double-counting effect the preferred approach is not to interact them because they already embed economic downturn. Additional trials have involved the use of Interest Rate at origination, which has been eventually excluded for its non-stationary pattern, as shown in Table 1. All the models are validated based on a set of criteria which involve rank-ordering (measured by GINI and AUROC coefficients), pseudo-R^2, robust standard errors, soundness of coefficients' sign and stability of the estimates on reduced sample. Following the champion model selection, the next methodological steps requires calculating correlations.

The analysis on correlations is not performed on the development sample while instead on a synthetic dataset created by Cartesian products across all possible mortgage variables, an approach which is also implemented by Packham and Woebbeking (2019). PD^{DT} and a PD^{LR} are predicted for each combination following the model estimated in Equatio 3. PD^{LR} is calculated by switching off the dummy crisis, while PD^{DT} is calculated by activating the dummy crisis.

Creating the synthetic portfolio is an important step of our analysis for two reasons. First, it would be misleading to draw conclusions from the mere regression coefficients, as pointed out by Ai and Norton (2003). By building a training dataset, instead, we can observe first-hand the variation of correlation based on the underlying mortgage variables. Second, we are able to stretch our analysis towards patterns not delimited by the development sample. Keeping fixed as few dimensions as possible, we let the coefficient ρ_i variate across the variables left free to move. We are therefore able to find specific patterns and to identify those mortgage segments where correlation is higher/lower.

Therefore, for each combination of mortgage characteristics i in $(1, \cdots, N)$, we are able to feed Equation 4 with PD_{ik}^{DT} and PD_i^{LR} and then reverse the BCBS (2021) formula to calculate the correlation ρ_k, as per Equation 5.

$$PD_i^{DT} = N\left(\frac{G(PD_i^{LR})}{\sqrt{1-\rho}} + \sqrt{\frac{\rho}{1-\rho}}G(0.999)\right) \tag{4}$$

where N(x) denotes the cumulative distribution function for a standard normal random variable, G(z) denotes the inverse cumulative distribution function for a standard normal random variable (i.e. the value of x such that N(x) = z), PD_i^{DT} is the downturn PD for mortgage i calculated by activating the dummy crisis, while PD_i^{BL} is the long-run PD for mortgage combination i, estimated in normal economic conditions. Equation 4 has to be resolved in $\sqrt{\rho_i}$ to obtain correlation values dependent on mortgage characteristics.

$$\sqrt{\rho_i} = \left(\frac{-2G(PD_i^{LR})G(0.999) \pm \sqrt{Q}}{2(G(0.999)^2 + G(PD_i^{DT})^2)}\right) \tag{5}$$

where

$$Q = 4G(PD_i^{LR})^2G(0.999)^2 - 4(G(0.999)^2 + G(PD_i^{DT})^2)(G(PD_i^{LR})^2 - G(PD_i^{DT})^2)$$

Being Equation 5 quadratic, there are two solutions for $\sqrt{\rho_i}$. We keep the solution with minus and then raise to the power of two, as correlations would otherwise be always greater than 90%

when using the plus operator. This procedure is applied to each mortgage in the synthetic portfolio, enabling the creation of mortgage-dependant correlation distribution.

4.2 Excess Interest Rate

The second methodological step mandates to define Excess Interest Rate at mortgage level. Our interest is to demonstrate whether mortgage providers started to account for varying correlation ρ_i when pricing new issued mortgages in the aftermath of the Great Financial Crisis. To achieve our scope, it is essential to ensure independence between the sample used to estimate correlation from the one used for Excess Interest Rate model; for this reason, we calculate correlation ρ_i based on the reduced sample model, which includes only mortgages originated up to 2011 (included). On the other hand, Excess Interest rate model is estimated on a sample that only includes mortgages originated after 2011.

Being constructed as the difference between mortgage Interest Rate at Origination and average Interest Rate at origination of all loans of 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 in Equation 6 and Equation 7

$$\delta_i = OriginalIR_i - \frac{\sum_{j=1}^{N_J^{T_Q}} OriginalIR_j}{N_J^{T_Q}} \tag{6}$$

$$\delta_i = OriginalIR_i - \frac{\sum_{j=1}^{N_J^{T_Y}} OriginalIR_j}{N_J^{T_Y}} \tag{7}$$

where Equation 6 calculates δ_i as the difference with quarterly average Interest Rate at origination and Equation 7 calculates δ_i as the difference with yearly average Interest Rate at origination. By adopting such approach, we want to isolate the pure effect of the premium charged by lenders, without any noise coming from interest rate trends. Equation 6 is selected as the champion approach, as it brings higher precision. As with the preparation of development sample for the discrete hazard model, δ_i has been tested both on full and cleaned samples. Being the difference negligible, we prefer the clean approach in consistency with discrete hazard model assumptions, reinforced by lower volumes of missing values following 2008 crisis.

Having built our target variable, Excess Interest Rate δ_i is then linearly regressed against the explanatory drivers, also including correlation ρ_i. Equation 8 outlines model specification:

$$\delta_i = \alpha + \sum_{b=1}^{N_b} \beta_b x_{b|i} + \sum_{m=1}^{N_m} \mu_m x_{m|T_i} + \omega * \rho_i + \sum_{f=1}^{N_B} \phi_f x_{f|i} + \sum_{p=1}^{N_B} \psi_p x_{p|i} * \rho_i \tag{8}$$

where δ_i is the estimated Excess Interest Rate at origination for mortgage i in $(1, \cdots, N)$. On the right hand side we find the intercept α, $beta_i$ coefficients (i in $1, \cdots, N_b$) capturing mortgage characteristics at origination, ensuring that we do not bias the effect of correlation; mu_m (m in $1, \cdots, N_m$) capture the state of the economy at time of origination. The rest of the equation involves bank-specific and correlation coefficients; phi_f (f in $1, \cdots, N_B$) relate to bank fixed-effects, ω measures the pure effect of correlation, while psi_p (p in $1, \cdots, N_B$) captures the interaction between correlation and banks.

Excess interest rate model is estimated as cross-sectional analysis, as our target variable is only measured at origination. This requires a set of assumptions to correctly incorporate correlation ρ, which instead has been estimated using a panel dataset to encompass mortgage dynamics (Equation 3). Therefore, the time-varying information needs to be flattened out for three variables: Updated LTV, LoanAge and macro drivers. First, Updated LTV is substituted with the LTV at origination, as the most natural choice to be made. This approach results is likely to result in lower correlations, however it is preferred not to make any further correction (i.e. stressing Original LTV). The second assumption is related to LoanAge, whose optimal value is selected based on observed peak in both PD^{DT} and PD^{LR}. Third, macroeconomic variables are inputted by calculating long-run and downturn averages by BEA territories. All the models are validated based on a set of criteria which involve goodness-of-fit via R^2, robust standard errors, soundness of coefficients' sign and stability of the estimates on different interactions.

5 Results

5.1 Correlation

The main result is that correlation in residential mortgage portfolios is a non-flat value, and that the magnitude of variation is dependant on specific mortgage characteristics. Furthermore, financial institutions have priced correlated segments in the aftermath of great Financial Crisis, even not with the same weight. We now illustrate how our results support both these claims.

Before presenting any conclusion on correlation patterns, the starting point is the correct assessment of PD^{DT} and PD^{LR}, as derived from the multi-period logit model reported in Table 8. Selection criteria have been already outlined in the methodology section, bringing to the final choice between two challenger models that only differ for selected macroeconomic drivers. Model 2 includes yearly growth rate of HPI ($HPI12$) and Unemployment (Ump), while Model 3 only incorporates yearly growth rate of Unemployment ($Ump12$). The latter approach is eventually chosen, because the macroeconomic coefficient remains stable even on the reduced sample model. Furthermore, we do not risk anyway to omit house price movements because these are already included in Updated Loan-to-Value.

Mortgage characteristics coefficients little change between Model 2 and Model 3, and they remain stable and significant even on reduced sample models. In particular, we make sure that signs are aligned with economic intuition for the non-interacted terms; for example, increasing Credit Score reduces default probability, as like as reducing Updated LTV and Debt-to-income do. However, as logistic regression establishes a non-linear relationship with the target variable through interacted coefficients, it is hard to interpret the sign of these latter, as pointed out by Ai and Norton (2003). We therefore pause any consideration related to the signs of mortgage variables when interacted with the dummy crisis, but we make sure that PD^{DT} is always greater than PD^{LR}. These tests, for example, led us to remove any non-stationary variables like Interest rate at origination and channel, that were otherwise producing counter-intuitive results, and to ensure that the final model correctly rank orders PD^{DT} and PD^{LR}, despite the non-intuitive sign of interacted coefficients.
Beyond soundness and significance of the estimates, rank ordering and predictive power are ensured

by the high levels of AUROC (87.9%) and Gini (75.8%) that have notably improved after our initial estimations, where only variables at originations were considered. LoanAge and Updated LTV, in fact, better capture mortgage dynamics and help the model to increase rank ordering. Robust standard errors are also deployed.

Model 3 is applied to the synthetic portfolio created as the cross-combinations of all mortgage features in the development sample. We avoid using the development data for the study over correlation patterns in order to assign an equal weight to each combination and avoid concentrations that shape Freddie Mac sample. In doing so, it is obtained a full distribution of correlations ρ_i that variate for each combination of mortgage characteristics i in the synthetic portfolio. Figure 5 to 8 provide a first insight of correlation distribution, which is then quantified in Table 9 and Table 11.

First and foremost, Figure 5 shows that even in residential mortgage asset class, correlation is a decreasing function of PD_i^{LR}. This finding is important for two reasons. First, it is sufficient to unfold the non-flat nature of correlation for retail asset class. Second, it is aligned with BCBS (2005) correlation calculations for corporate exposures, which is a decreasing function of increasing implied PD.

Effects of variability are presented from Figure 6 to Figure 8, which visually break the relation between correlation ρ_i and PD_i^{LR} by most relevant mortgage features. Figure 6 unfolds one of the most expected findings, which is that correlation almost duplicates with increasing Updated LTV. This is generally experienced across all BEA territories, and is linked to falling house prices caused by 2008 economic turmoil. While it is natural to expect that contagion effect is magnified for underwater mortgages, some regions experience higher levels of correlation, such as MidEast, New England and Rocky Mountains, which are not necessary the regions that experienced higher default rates during downturn.

Figure 7 splits the relationship between default correlation ρ_i and PD_i^{LR} by Credit Score and Number of Borrowers. In line with expectation, single-borrower mortgages are riskier (higher PD_i^{LR}), but on the other hand this segment exhibits a lower default correlations ρ_i compared with Joint applicants. In addition, Single/Joint segments show a different sensitivity to correlation when interacted with Credit Score. In fact, ρ_i tends to increase with decreasing credit score for Joint mortgages, while Single borrowers' correlation is characterised by a hump shape with lower contagion for sub-prime mortgages. This is most probably linked to the deteriorating PD_i^{LR}, which makes mortgages cluster less when subject to adversely changing economic conditions.

Another striking pattern in exhibited in Figure **??**. While Recourse and Non-Recourse mortgages do not differ substantially in PD_i^{LR} distribution, the difference in default correlation ρ_i is remarkable. Non-recourse states experience higher levels of correlations regardless of PD Long-run values. This is a consequence of non-recourse policies, as borrowers are more reckless in defaulting because lenders cannot attempt to take possession of other assets to make up for the loss. Such behaviour is magnified by BEA territories of the mortgage. Finally, we observe in Figure 7 that high Debt-to-Income borrowers are more prone in default clustering during a downturn, while we can observe from Figure 8 that when considering correlation by Sellers, Bank of America, Branch banking, Chase, Citi, Sun trust and US bank generally manifest higher contagion effects compared with smaller regional banks, with the only exception of Wells Fargo.

To further complement the analysis, we summarise residential mortgage default correlation by cross-tabulations that compute average, maximum and standard deviation of ρ_i. Table 9 demonstrates that on average correlations are pretty much aligned with Cowan and Cowan (2004) findings. However, Table 11 also highlights that even within the same category/combinations, the variability brought by other characteristics can results in correlation values more than 10 times the average value, and even higher than the conservative 15% flat value recommended by BCBS (2005). While it is true that banks portfolio are generally well diversified, risk managers have to be careful because if the portfolio is concentrated in cross-segments, correlation risk can be significantly high and the entire segment is exposed to a greater contagion effect.
Tables 9, 10 and 11 also offer an augmented insight, as they include additional dimensions that have not been reported graphically (e.g. Purpose, First Time Homebuyer and Property type), and that offer additional insight if read jointly. We see, for example, that average correlation increases with credit score, while maximum correlation has an opposite trend. This is due to higher levels of volatility in the high risk buckets. On the other hand, Debt-to-Income and Updated LTV are consistent both in average and standard deviations, with the risky buckets having higher correlation both on average and on maximum. Mortgage default correlations (all else equal) can also triplicate from low-risk segments to high risk ones. Both Debt-to-Income and Updated LTV are amongst the variables that show a stronger variability, if compared with other continuous variables like Credit score.

Other interesting patterns highlight higher correlations for refinance mortgages compared with purchase, as well as Not-First Time homebuyers with first time ones. This is a quite interesting result, as it would be expected that first time homebuyer would cluster more than experienced buyers, possibly due to lower savings and higher risk of contagion. However, this finding might reveal that under distressed economic conditions, borrowers already holding a property are more reckless in defaulting. A similar finding has been graphically measured for Non-recourse states, and here also confirmed quantitatively, as we can see that Non-recourse states consistently have higher correlation values than recourse ones. Again, this is a segment of lending where financial institutions should be careful about. Last, Property type is analysed. We can see that PUDs, Condos and Single family properties experience a correlation which is four times higher than Co-op properties, by taking both average and maximum values. This effect is curbed when combined with high Debt-to-Income.

5.2 Excess Interest Rate

Having ascertained that correlation is effectively a non-flat value, and that it can be high for specific segments, we now assess if financial institutions take into account correlation risk when pricing through-the-door mortgages. To achieve such objective, Excess Interest Rate is linearly regressed by a series of factors, including segment-varying correlation ρ_i. Differently from panel-logit discrete hazard model, the frequency of observations is now quarterly, and Excess Interest rate is obviously only measured at origination, as the sample is composed of fixed-rate mortgages. We are not anyway interested in interest rate resets, as our objective is to quantify if lenders price correlation risk differently at the time of mortgage application.

Regression results are reported in Table 12, where Bank-specific effects have been progressively included in the estimations. Model 1 does not include any Bank-effect, Model 2 incorporates Bank fixed effects only, and Model 3 finally accounts for interaction between correlation and Bank fixed effects. All the 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.

In general, we observe that lending institutions price correlation risk, and tend to assign a premium proportional to correlation ρ_i and variable on mortgage attributes. As we can see from Table 12 results, correlation ρ_i coefficient is positive for all the regressions tested. We also ensure that non-flat correlation is not biasing the results by making sure that the usual factors driving mortgage premium are correctly estimated. Credit score and Joint applications lower Excess Interest Rate, while Original LTV, DTI and first-time homebuyers increase it. Even the sign of macroeconomic drivers, that we use to account for quarter of origination, correctly measures that mortgage premium increases when economy is expanding.

Whilst correlation ρ_i has a positive impact on Excess Interest Rate, we observe from Model 3 that banks generally do not assign it the same weight. Provident, Citi and Chase, in fact tend to price less correlation risk, compared to the other institutions like FifthThird, which is amongst the banks that price it the most.

To quantify the impact, we rely on a stylized example and take a reference mortgage to calculate the differential impact of priced correlation by banks on total interests paid. The reference mortgage is a 30-Year Fixed-Rate Mortgage with original balance of \$ 300,000 and rate at origination equal to 5.5 %. Similarly to the synthetic portfolio used to analyse non-flat correlation, we compute for our reference mortgage all possible combinations of attributes, including lenders in Table 12. Correlation is then calculated for each entry and the marginal effect on Excess Interest rate of the interacted correlation with banks effect is computed. The Excess Interest Rate is then added to the reference original rate, so that the only source of variability is produced by correlation ρ_i and its interaction of with each Bank. At this stage, correlation is obviously dependant on mortgage features, but what we are interested about is to understand how banks, all else equal, differently price correlation risk. To achieve such result we collapse the portfolio and calculate the maximum difference in total interests paid amongst banks within each unique combination. Results are then plotted to check whether the difference in total interests is material or not, and whether difference in pricing is stressed for specific mortgage segments.

Figure 9 shows that difference in total interests paid is concentrated below \$ 7,000, suggesting that overall banks do not excessively differentiate in pricing default correlation ρ_i. However, the right tail of the distribution shows that the distribution outstretches to \$ 40,000. This implies that banks tend to price in a different way specific segments in the mortgage market based on default correlation risk of specific clusters. For example, Figure 10 clearly shows that geographic patterns exist in pricing mortgage default correlation ρ, especially where the lending is concentrated in Midwest, New England, Rocky mountains. In these area characterised by higher correlation, financial institutions differently price default correlation ρ_i, giving an incentive to borrowers in shopping around, as total interests charged can differ substantially. The same finding does not apply to the differential impact by Original LTV and Credit Score. In fact, Figure 11 highlights that financial institutions price mortgage default correlation ρ by increasing Original LTV in somehow a similar way. Lenders are therefore aware that mortgages with higher LTVs tend to cluster and to experience a contagion effect under adverse economic conditions, hence pricing it quite similarly. On the other hand, Figure 12 shows borrowers with lower Credit Scores are priced similarly, even if they tend to cluster and to experience contagion effect under adverse economic conditions. Borrowers might not significantly benefit if they want to access lower interest charges based on their Credit Score or Original LTV.

The last three cases analysed relate to Debt-to-Income, Recourse/Non recourse and Property type. Figure 12 highlights that banks price mortgage default correlation ρ by increasing DTI at origination quite differently, especially when Debt-to-Income ratio is greater than 30 %. The result is interesting, especially as the support of the distribution always stretches towards tail values for increasing DTI. While banks are quite conservative and consistent in pricing mortgage default correlation ρ for increasing LTVs, the same approach seems not to be followed for DTI. This implies that some financial institutions might underestimate default correlation risk for borrowers having higher debt-to-income ratio.

To conclude our analysis, we focus on Non-recourse and Property type. Figure 13 highlights that correlation risk for non-recourse states is generally priced differently. This is most likely attributed to different business capillarity in non-recourse states, resulting on a disparity when pricing mortgage default correlation ρ_i. For what concerns property type, it seems that it is one of the key characteristic where mortgage issuers apply the most diverse behaviour in pricing default correlation ρ_i. For coop and manufactured housing, the Excess interest rate is more concentrated around zero, compared with the other property types. This means that the measurement of default correlation risk ρ_i is perceived quite similarly amongst financial institutions, while the same cannot be said for Condos, PUD and Single-family, where correlation is generally higher priced.

6 Conclusion

This paper has investigated if default correlation for residential mortgages is effectively a flat-value (as recommended by BCBS (2005)) or if instead it variates depending on specific mortgage characteristics. By deploying an extensive sample sourced by Freddie Mac that also includes the Great Financial Crisis, we have demonstrated that the magnitude in default correlation variability is highly dependant on mortgage attributes. Taking average correlations might shade the true degree of interconnection in mortgage defaults that can occur at the intersection of specific features, even outside the most known risk patterns. Therefore, risk managers should pay attention to the risks caused by concentration, to avoid the rise of unexpected losses caused by increased correlation.

Following the demonstration of non-flat mortgage default correlations, our framework is deployed to understand if lending institutions effectively price correlation risk. After having isolated the effect of correlation by different banks on the excess interest rate, we have shown on a stylised example that correlation risk is priced differently depending on mortgage segments. Correlation caused by the usual risk drivers such as Credit Score and Loan-to-value is generally similarly priced, as if lending institutions are aware of the risks coming from contagion effect in these segments. The same is not observed for other mortgage buckets, such as property or geographical area, where default correlation risk is priced much differently by banks.

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	NAccounts	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	711.79	51.96	785	45	76.68	15.18	95	15	32.78	11	51	6.5	7.31	0.55	8.25	50,000	$1.26 \cdot 10^5$	$5.46 \cdot 10^4$	230,000
2000	786,275	615	712.21	55.58	789	45	77.58	15.48	95	17	34.66	10.71	51	7.38	8.18	0.5	9	50,000	$1.32 \cdot 10^5$	$5.88 \cdot 10^4$	245,000
2001	1,755,390	617	714.05	58.74	791	45	75.4	14.71	95	15	33.24	11.04	50	6.38	7.01	0.42	7.75	58,000	$1.48 \cdot 10^5$	$6.44 \cdot 10^4$	273,000
2002	1,682,997	617	717.11	56.68	792	42	73.9	15.47	95	15	33.51	11.68	53	5.88	6.57	0.48	7.38	59,000	$1.56 \cdot 10^5$	$7.04 \cdot 10^4$	292,000
2003	1,927,050	632	724.9	51.56	794	40	72.1	15.74	95	12	32.26	12.28	53	5.13	5.78	0.41	6.5	62,000	$1.61 \cdot 10^5$	$7.44 \cdot 10^4$	302,000
2004	1,127,674	624	717.93	54.54	794	42	73.67	15.25	95	15	34.99	11.97	55	5.38	5.86	0.37	6.5	61,000	$1.67 \cdot 10^5$	$7.82 \cdot 10^4$	320,000
2005	1,690,993	626	724.35	58.25	802	34	69.63	17.35	95	15	35.34	12.26	56	5.25	5.8	0.39	6.5	57,000	$1.71 \cdot 10^5$	$8.61 \cdot 10^4$	343,000
2006	1,260,783	623	722.95	58.26	803	34	70.64	17.35	95	16	36.62	12.2	57	5.75	6.41	0.39	7	59,000	$1.8 \cdot 10^5$	$9.43 \cdot 10^4$	372,000
2007	1,220,654	621	722.89	58.4	804	35	71.97	17.67	95	16	36.8	12.48	58	5.75	6.37	0.43	7.13	59,000	$1.83 \cdot 10^5$	$9.82 \cdot 10^4$	388,000
2008	1,179,578	643	739.68	51.86	806	35	70.08	17.52	95	16	36.35	12.79	58	5.25	6.03	0.54	6.88	62,000	$2.04 \cdot 10^5$	$1.08 \cdot 10^5$	417,000
2009	1,974,690	685	762.44	39.58	809	32	65.24	16.98	88	14	31.63	11.61	52	4.38	4.94	0.38	5.63	68,000	$2.14 \cdot 10^5$	$1.17 \cdot 10^5$	417,000
2010	1,271,397	684	763.69	40.18	811	33	66.24	16.96	90	14	31.48	10.49	48	3.75	4.59	0.49	5.38	65,000	$2.09 \cdot 10^5$	$1.2 \cdot 10^5$	417,000
2011	955,418	685	764.17	39.84	811	33	67.33	17.16	90	15	31.65	10.13	47	3.38	4.29	0.57	5.25	65,000	$2.18 \cdot 10^5$	$1.26 \cdot 10^5$	417,000
2012	1,331,301	690	766.52	38.35	812	34	68.04	17.23	95	14	30.84	10.02	46	2.88	3.58	0.44	4.25	70,000	$2.22 \cdot 10^5$	$1.23 \cdot 10^5$	417,000
2013	1,300,286	681	759.45	41	810	36	70.58	17.19	95	15	31.99	9.76	46	2.75	3.74	0.62	4.75	68,000	$2.18 \cdot 10^5$	$1.18 \cdot 10^5$	417,000
2014	975,881	670	751.5	44.05	808	41	75.08	16.22	95	17	33.75	9.15	46	3.38	4.27	0.46	4.88	68,000	$2.19 \cdot 10^5$	$1.19 \cdot 10^5$	417,000
2015	1,316,566	670	752.3	44.04	809	40	73.63	16.48	95	17	33.79	9.39	47	3.13	3.95	0.45	4.63	75,000	$2.29 \cdot 10^5$	$1.19 \cdot 10^5$	417,000
2016	1,558,394	668	751.08	44.62	809	39	73.01	16.62	95	17	34.07	9.39	48	2.88	3.76	0.47	4.5	80,000	$2.41 \cdot 10^5$	$1.19 \cdot 10^5$	431,000
2017	1,217,105	662	747.04	45.92	808	40	74.11	16.46	95	18	35.02	9.41	48	3.38	4.17	0.44	4.88	75,000	$2.35 \cdot 10^5$	$1.21 \cdot 10^5$	424,000

Table 2: Summary Statistics:Categorical Variables

The table reports percentage distribution by year of origination of 1^{st} Time Homebuyer; Occupancy: Investment (I), Primary (P), Second Home (S); 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}Homebuyer		Occupancy			Channel				Property					Purpose			N.Borrowers	
Year	NAccounts	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	4.1	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	81.7	25.3	35.1	39.6	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	82.2	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	80.9	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	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	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	4.92	0.15	0.26	21.25	73.41	26.78	41.72	31.5	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	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	6.33	0.23	0.26	25.08	68.1	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	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	16.4	7.57	88.77	3.66	10.67	30.9	58.43	0	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	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	18.5	9.95	86.01	4.04	9.73	33.53	56.74	0	7.97	0.14	0.42	27.62	63.85	25.36	17.39	57.24	50.58	49.42

Table 3: Summary Statistics:Categorical Variables
The table reports percentage distribution by year of origination of mortgage Sellers

Year	NAccounts	BankOfAmerica	BranchBanking	Chase	Citi	FifthThird	FlagStar	Other	Provident	SunTrust	UsBank	WellsFargo
1999	1,095,011	0.037	0.002	0.019	0.000	0.000	0.032	0.909	0.000	0.000	0.000	0.000
2000	786,275	0.102	0.018	0.033	0.000	0.006	0.000	0.617	0.000	0.000	0.000	0.225
2001	1,755,390	0.054	0.018	0.024	0.000	0.010	0.000	0.570	0.014	0.040	0.000	0.271
2002	1,682,997	0.053	0.020	0.013	0.000	0.016	0.000	0.509	0.019	0.017	0.031	0.322
2003	1,927,050	0.009	0.013	0.072	0.000	0.016	0.000	0.486	0.021	0.000	0.033	0.350
2004	1,127,674	0.000	0.015	0.146	0.000	0.015	0.000	0.460	0.003	0.003	0.038	0.320
2005	1,690,993	0.031	0.003	0.090	0.003	0.010	0.000	0.571	0.005	0.000	0.044	0.244
2006	1,260,783	0.033	0.013	0.084	0.022	0.017	0.008	0.494	0.012	0.008	0.052	0.257
2007	1,220,654	0.091	0.020	0.072	0.033	0.023	0.008	0.493	0.020	0.017	0.055	0.168
2008	1,179,578	0.076	0.039	0.099	0.048	0.022	0.017	0.440	0.006	0.003	0.073	0.177
2009	1,974,690	0.074	0.049	0.073	0.039	0.025	0.006	0.372	0.042	0.017	0.081	0.222
2010	1,271,397	0.118	0.058	0.055	0.037	0.022	0.002	0.364	0.026	0.006	0.109	0.203
2011	955,418	0.049	0.066	0.066	0.024	0.025	0.000	0.357	0.050	0.003	0.116	0.245
2012	1,331,301	0.036	0.065	0.052	0.003	0.025	0.000	0.384	0.059	0.000	0.140	0.235
2013	1,300,286	0.058	0.070	0.124	0.015	0.019	0.000	0.453	0.013	0.024	0.085	0.138
2014	975,881	0.070	0.049	0.066	0.016	0.007	0.000	0.595	0.002	0.013	0.063	0.121
2015	1,316,566	0.056	0.040	0.036	0.015	0.004	0.000	0.654	0.003	0.007	0.063	0.124
2016	1,558,394	0.039	0.030	0.052	0.006	0.000	0.000	0.652	0.006	0.013	0.046	0.155
2017	1,217,105	0.018	0.023	0.082	0.002	0.008	0.000	0.667	0.005	0.019	0.040	0.135

Figure 1: Summary Statistics:States Distribution
The figure displays the distribution of mortgages by States across the entire sample. A higher concentration $(> \%) Z$ is observed in California, Florida, Texas and Illinois. A lower concentration is observed in Montana, North Dakota, South Dakota , Wyoming, Mississippi, West Virginia, Vermont and Delaware.

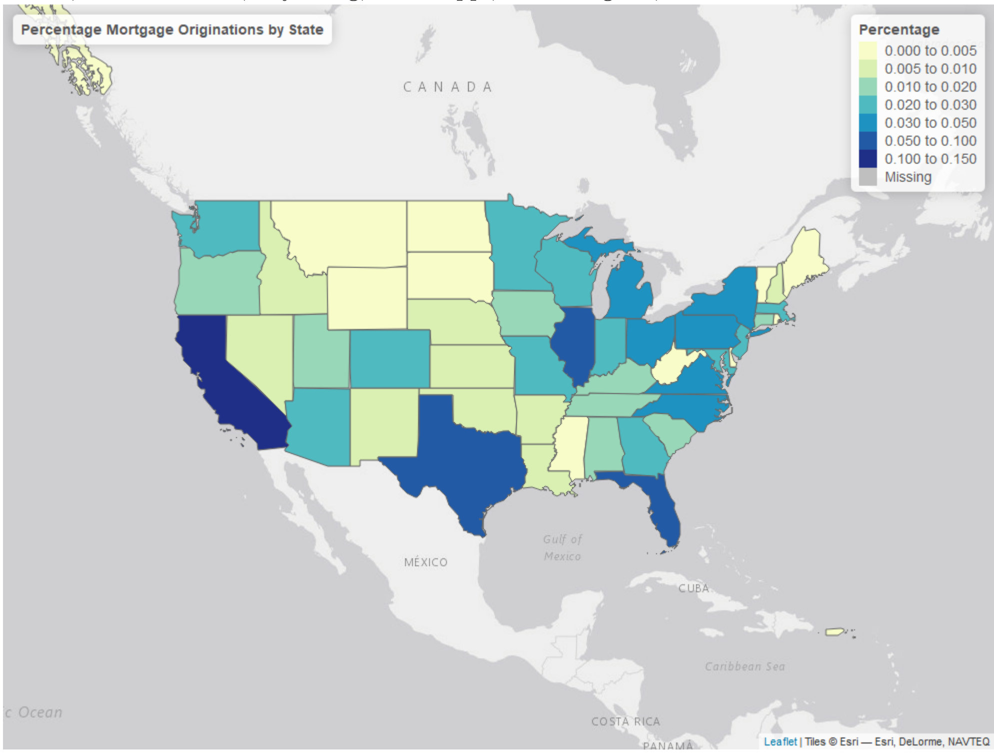

Table 4: HMDA Representativeness
The table shows the breakdown of mortgage applications across the United States sourced by HMDA (Home Mortgage Disclosure Act) from 2007 to 2017. The Home Mortgage Disclosure Act (HMDA) requires financial institutions to maintain, report, and publicly disclose loan-level 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). In particular, Freddie Mac mortgages cover 25.6 % of total conventional originated loans and 17.7 % of total mortgage originations. The sample used for the analysis in this paper sits under this latter bucket, hence the analysis performed is representative for more than a quarter of conventional originated mortgages, which is a relevant share of US mortgage market.

Data	Percentage	Volumes
Total Mortgage Applications		187,462,446
Total 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 %

Table 5: Yearly Default Rate:Continuous Variables

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	Accounts	Defaults	All	0-450	450-550	550-650	650-750	750-850	0-20	20-40	40-60	60-65	-Inf–1	-1–0.5	-0.5–0	0-0.5	0.5-1	1-Inf
1999	939,197	718	0.0764	0.0000	1.8634	0.2302	0.0554	0.0194	0.0690	0.0764	0.0823	0.0794	0.0341	0.0287	0.0496	0.0886	0.2118	0.3632
2000	1,706,123	5,729	0.3358	0.0000	2.6050	1.1218	0.2551	0.0619	0.2184	0.3288	0.4170	0.4067	0.1250	0.1040	0.2209	0.3801	0.8409	1.4684
2001	3,292,855	16,049	0.4874	0.1902	3.5544	1.7026	0.3783	0.0775	0.2734	0.4671	0.6475	0.6191	0.1543	0.1794	0.3067	0.5220	1.3097	3.2005
2002	4,382,304	28,120	0.6417	0.6728	5.2867	2.2939	0.5004	0.0983	0.3665	0.6018	0.8751	0.4971	0.3198	0.2861	0.3640	0.6506	2.0356	5.8150
2003	5,311,191	35,279	0.6642	0.8640	5.9220	2.6124	0.5472	0.0954	0.3563	0.6326	0.9231	0.6683	0.3829	0.2634	0.3892	0.7450	1.7266	5.6552
2004	4,631,059	30,804	0.6652	1.2413	5.3324	2.3810	0.6039	0.1028	0.3398	0.6484	0.8947	0.6019	0.4459	0.2615	0.4125	0.7688	1.5055	4.6401
2005	5,424,031	32,325	0.596	0.5119	4.6402	2.1289	0.5817	0.112	0.3513	0.5866	0.7528	0.5531	0.2665	0.2710	0.4167	0.6538	1.4363	3.9954
2006	5,927,775	33,570	0.5663	0.6971	3.8933	1.9626	0.5725	0.1302	0.3530	0.5534	0.6966	0.5908	0.2614	0.2735	0.4055	0.6290	1.3211	3.4684
2007	6,641,321	44,175	0.6652	1.3213	4.4629	2.3102	0.7019	0.1294	0.3465	0.6259	0.8668	0.8268	0.2631	0.2631	0.4420	0.8017	1.5816	3.2285
2008	7,359,285	93,166	1.266	2.0737	6.9529	4.2531	1.4468	0.2929	0.5885	1.1087	1.7820	1.9306	0.3030	0.4400	0.8158	1.5874	3.0241	5.1204
2009	8,645,786	217,599	2.5168	3.6265	11.5346	8.8030	3.3564	0.6527	0.8860	2.0525	3.9609	5.4569	0.7857	1.1211	1.7916	3.1030	5.2888	8.2339
2010	8,370,112	195,879	2.3402	3.3732	8.4092	7.8030	3.3277	0.7575	0.8022	1.8821	3.7859	6.2401	1.1223	1.1389	1.8701	2.7368	4.3386	6.4086
2011	7,803,125	134,082	1.7183	3.2623	6.5781	5.6793	2.5215	0.6299	0.6636	1.4210	2.7077	4.9128	0.9180	0.8215	1.4633	1.9537	2.6630	4.6967
2012	7,776,197	103,176	1.3268	2.8960	5.5253	4.7541	2.0511	0.4971	0.5295	1.1144	2.1056	3.9549	0.7636	0.5842	1.2195	1.4462	1.9763	3.3827
2013	7,245,881	65,902	0.9095	2.3198	4.8210	3.8816	1.4320	0.3196	0.3933	0.7691	1.4297	3.3085	0.4298	0.3418	0.8702	0.9723	1.3976	2.4984
2014	6,844,299	45,048	0.6582	1.4541	4.4424	3.1806	1.0387	0.2338	0.3337	0.5633	0.9919	2.8576	0.2288	0.2523	0.6452	0.6820	1.0607	1.9982
2015	7,515,347	32,781	0.4362	1.9920	3.8349	2.2195	0.6923	0.1646	0.2355	0.3740	0.6460	1.9991	0.1535	0.1705	0.4065	0.4597	0.7539	1.4455
2016	8,128,933	27,950	0.3438	2.2544	2.2944	1.8247	0.5467	0.1306	0.1992	0.3033	0.4829	1.5847	0.1229	0.1336	0.2833	0.3792	0.6589	1.2011
2017	8,415,102	32,457	0.3857	1.7578	3.4142	1.9454	0.6223	0.1417	0.2038	0.3373	0.5534	1.6404	0.1433	0.1545	0.2871	0.4334	0.7945	1.2333
2018	7,705,902	23,463	0.3045	0.9174	2.1877	1.3282	0.4958	0.1241	0.1290	0.2623	0.4644	1.1636	0.2130	0.1430	0.2097	0.3432	0.6544	0.9665

Table 6: Yearly Default Rate:Continuous Variables

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 Lona-to-Value at origination (Original LTV) and Updated Loan-to-Value. Original LTV is a static fields, hence all the accounts belonging to a bucket do not migrate to other buckets. Updated LTV is a dynamic field, hence accounts that belong to a bucket in time t can migrate to a different Updated LTV bucket in $t+1$, depending on the ratio between amortised balance and updated property price. This latter is subject to variation in House Price Index at state level.

				Original LTV						Updated LTV					
Year	Accounts	Defaults	All	0-30	30-50	50-70	70-80	80-90	90-Inf	0-30	30-50	50-70	70-801	80-90	90-Inf
1999	939,197	718	0.0764	0.0415	0.0280	0.0438	0.0662	0.1275	0.1147	0.0428	0.0265	0.0433	0.0692	0.1215	0.1186
2000	1,706,123	5,729	0.3358	0.1023	0.1166	0.1712	0.2710	0.5195	0.5888	0.0958	0.1141	0.1889	0.3101	0.5859	0.6528
2001	3,292,855	16,049	0.4874	0.1805	0.1535	0.2447	0.3661	0.8163	0.9888	0.1901	0.1692	0.3632	0.4190	1.0033	0.7113
2002	4,382,304	28,120	0.6417	0.1322	0.1581	0.2839	0.4807	1.2205	1.4593	0.1523	0.1824	0.4300	0.6688	1.5213	1.6083
2003	5,311,191	35,279	0.6642	0.1421	0.1394	0.2877	0.5155	1.3344	1.7234	0.1994	0.2168	0.5125	0.7260	1.5560	1.2360
2004	4,631,059	30,804	0.6652	0.1023	0.1446	0.2635	0.5148	1.4233	1.8296	0.1356	0.2545	0.5508	0.8067	1.8769	1.3792
2005	5,424,031	32,325	0.596	0.1188	0.1397	0.2851	0.4989	1.3462	1.6875	0.1351	0.2822	0.6497	0.7492	1.4613	0.6996
2006	5,927,775	33,570	0.5663	0.1264	0.1789	0.3323	0.5027	1.2308	1.5171	0.1703	0.3212	0.6297	0.6412	1.0125	0.9294
2007	6,641,321	44,175	0.6652	0.1654	0.2243	0.4352	0.6230	1.3135	1.6118	0.1862	0.3311	0.6584	0.8127	0.8862	1.3502
2008	7,359,285	93,166	1.266	0.2332	0.3802	0.8423	1.2563	2.3340	2.9459	0.2345	0.4873	0.8905	1.2707	1.5065	2.1287
2009	8,645,786	217,599	2.5168	0.3254	0.6484	1.8049	2.7316	4.7452	5.5336	0.2199	0.4540	1.0165	1.4708	2.7914	7.5737
2010	8,370,112	195,879	2.3402	0.3499	0.6862	1.7586	2.5861	4.3412	5.0046	0.2368	0.5082	1.0691	1.5581	3.1207	7.8126
2011	7,803,125	134,082	1.7183	0.2973	0.5599	1.2999	1.9232	3.1203	3.5565	0.2038	0.4295	0.8787	1.1549	1.9976	5.3259
2012	7,776,197	103,176	1.3268	0.2550	0.4648	1.0072	1.4891	2.3772	2.6408	0.1824	0.3821	0.7171	1.0767	2.4538	6.2198
2013	7,245,881	65,902	0.9095	0.2180	0.3638	0.7275	1.0062	1.5767	1.5539	0.1799	0.4220	0.7420	1.1266	2.0439	4.2464
2014	6,844,299	45,048	0.6582	0.2135	0.3115	0.5549	0.7166	1.0729	0.934	0.1890	0.4200	0.6908	0.8729	1.2738	1.7796
2015	7,515,347	32,781	0.4362	0.1717	0.2389	0.3844	0.4610	0.6452	0.5992	0.1803	0.3639	0.5262	0.4662	0.5984	0.5479
2016	8,128,933	27,950	0.3438	0.1521	0.2002	0.2963	0.3590	0.4845	0.4829	0.1707	0.3316	0.4210	0.3328	0.3828	0.3267
2017	8,415,102	32,457	0.3857	0.1594	0.2231	0.3269	0.3916	0.5248	0.5964	0.2068	0.3843	0.4411	0.3961	0.5228	0.3890
2018	7,705,902	23,463	0.3045	0.1299	0.1791	0.2514	0.3077	0.3949	0.5025	0.1508	0.2612	0.3303	0.4135	0.5362	0.9217

Table 7: Yearly Default Rate:Categorical Variables

The table reports the 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 1^{st} Time Homebuyer; Occupancy: Investment (I), Primary (P), Second Home (S); 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}Homebuyer		Occupancy			Property					Purpose			N.Borrowers	
Year	Accounts	Defaults	All	No	Yes	I	P	S	CO	CP	MH	PU	SF	C	N	P	1	2
1999	939,197	718	0.0764	0.0748	0.0874	0.1268	0.0747	0.0653	0.0656	0.1383	0.1124	0.0499	0.0806	0.0699	0.102	0.0662	0.1235	0.0497
2000	1,706,123	5,729	0.3358	0.3349	0.3407	0.2996	0.3422	0.2098	0.2405	0.1156	1.0604	0.207	0.3599	0.4077	0.4533	0.2819	0.5439	0.2094
2001	3,292,855	16,049	0.4874	0.4772	0.558	0.5565	0.4916	0.2784	0.3217	0.0994	2.0832	0.3237	0.5183	0.4678	0.5139	0.4814	0.772	0.3161
2002	4,382,304	28,120	0.6417	0.6311	0.7327	0.9713	0.6371	0.3264	0.4293	0.3266	2.3181	0.4218	0.6802	0.5649	0.6633	0.6656	1.0292	0.403
2003	5,311,191	35,279	0.6642	0.6476	0.8448	1.0695	0.6562	0.3521	0.3598	0.1896	2.0569	0.408	0.7147	0.6136	0.6197	0.7436	1.0361	0.4346
2004	4,631,059	30,804	0.6652	0.6442	0.91	0.8106	0.6718	0.3121	0.3601	0.1305	2.1229	0.3707	0.7193	0.6359	0.6355	0.7158	0.9934	0.4464
2005	5,424,031	32,325	0.596	0.5869	0.7008	0.7328	0.6034	0.2808	0.3187	0.1544	1.5665	0.292	0.6525	0.5401	0.6478	0.5906	0.8622	0.4151
2006	5,927,775	33,570	0.5663	0.5565	0.6732	0.6344	0.5654	0.5223	0.57	0.1756	1.15	0.3274	0.5954	0.5337	0.6242	0.5468	0.8117	0.3951
2007	6,641,321	44,175	0.6652	0.656	0.7572	0.6003	0.6814	0.4039	0.5147	0.2567	1.2593	0.4447	0.7058	0.7255	0.6843	0.6047	0.9573	0.4549
2008	7,359,285	93,166	1.266	1.25	1.4221	1.3962	1.27	1.046	1.2845	0.4069	2.0553	1.127	1.2798	1.4852	1.1131	1.1983	1.7829	0.8832
2009	8,645,786	217,599	2.5168	2.4963	2.7349	3.0695	2.5183	1.9276	2.7131	0.7879	3.8923	2.3078	2.5299	3.2673	1.8847	2.4245	3.4287	1.8696
2010	8,370,112	195,879	2.3402	2.3153	2.6027	2.6351	2.3554	1.7376	2.701	0.7313	4.0692	2.0779	2.3473	3.0194	1.6717	2.374	3.1016	1.7964
2011	7,803,125	134,082	1.7183	1.7008	1.9004	2.0052	1.7188	1.3825	2.1338	0.7136	3.1225	1.4511	1.7244	2.1903	1.2011	1.8185	2.2647	1.3228
2012	7,776,197	103,176	1.3268	1.3054	1.5544	1.4061	1.3357	1.075	1.7693	0.838	2.7906	0.9972	1.3471	1.7434	0.8827	1.4563	1.7534	1.0208
2013	7,245,881	65,902	0.9095	0.8993	1.01	0.8923	0.9221	0.703	1.085	0.8984	2.1207	0.5523	0.9674	1.2452	0.6261	0.9423	1.2246	0.6734
2014	6,844,299	45,048	0.6582	0.6587	0.6523	0.5978	0.6725	0.4692	0.6925	0.6947	1.7062	0.3342	0.7297	0.9314	0.4895	0.6179	0.8915	0.476
2015	7,515,347	32,781	0.4362	0.4364	0.4337	0.3906	0.4452	0.3248	0.441	0.5357	1.2892	0.2176	0.491	0.6162	0.3311	0.4069	0.605	0.3011
2016	8,128,933	27,950	0.3438	0.3386	0.3791	0.305	0.351	0.2561	0.3162	0.3887	1.1024	0.1943	0.3867	0.4597	0.2613	0.339	0.4839	0.2271
2017	8,415,102	32,457	0.3857	0.3716	0.4747	0.3095	0.3986	0.2491	0.3263	0.375	0.9339	0.3414	0.4014	0.4847	0.2862	0.4004	0.5362	0.2555
2018	7,705,902	23,463	0.3045	0.2875	0.4097	0.2334	0.3143	0.2248	0.2963	0.2382	0.4864	0.3419	0.2916	0.3832	0.1988	0.3363	0.4135	0.2101

Figure 2: Realised Defaults by Month of Observation
The figure displays the number of first default occurrence by year and month.

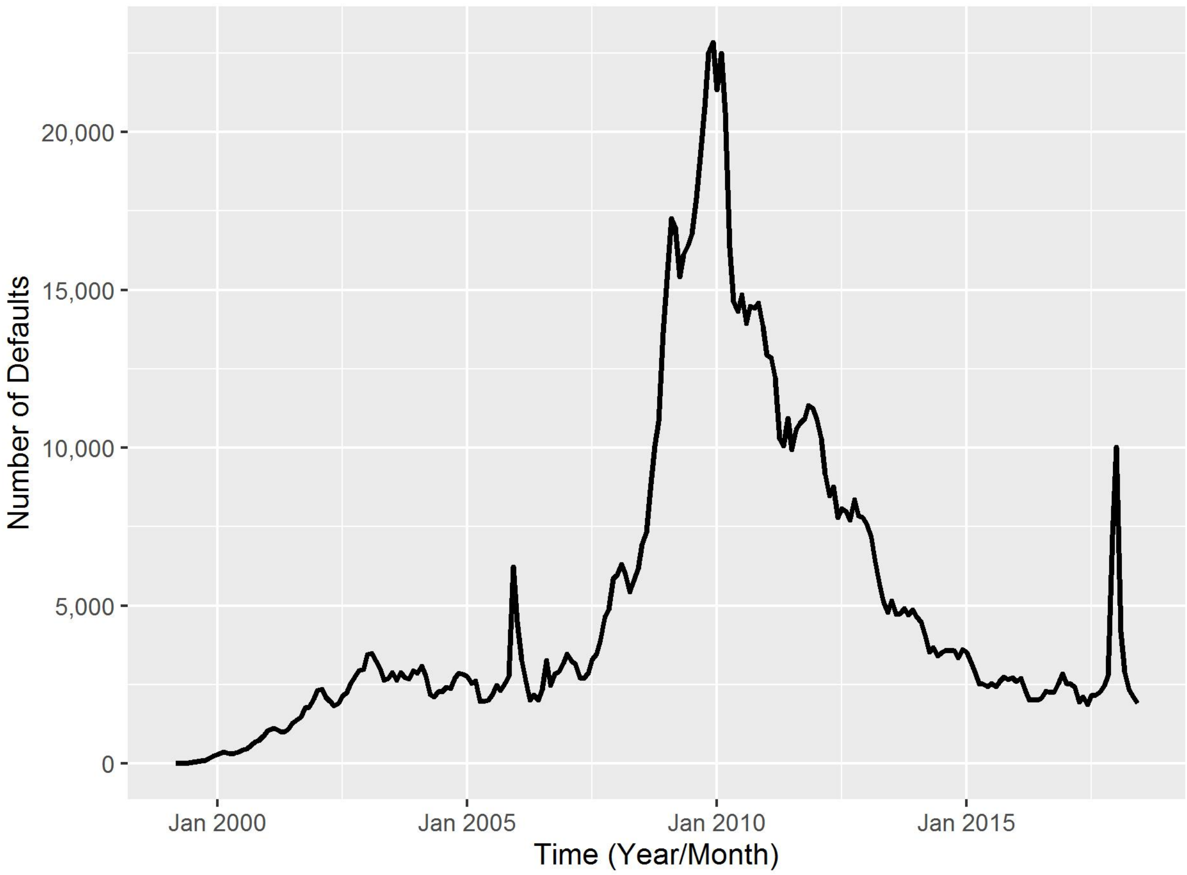

Figure 3: Share of Defaults by Year of Origination
The figure displays the number of mortgages by year of origination. The blue portion of the stacked barlpot counts the number of mortgages belonging to that year of origination that have defaulted within the observation period.

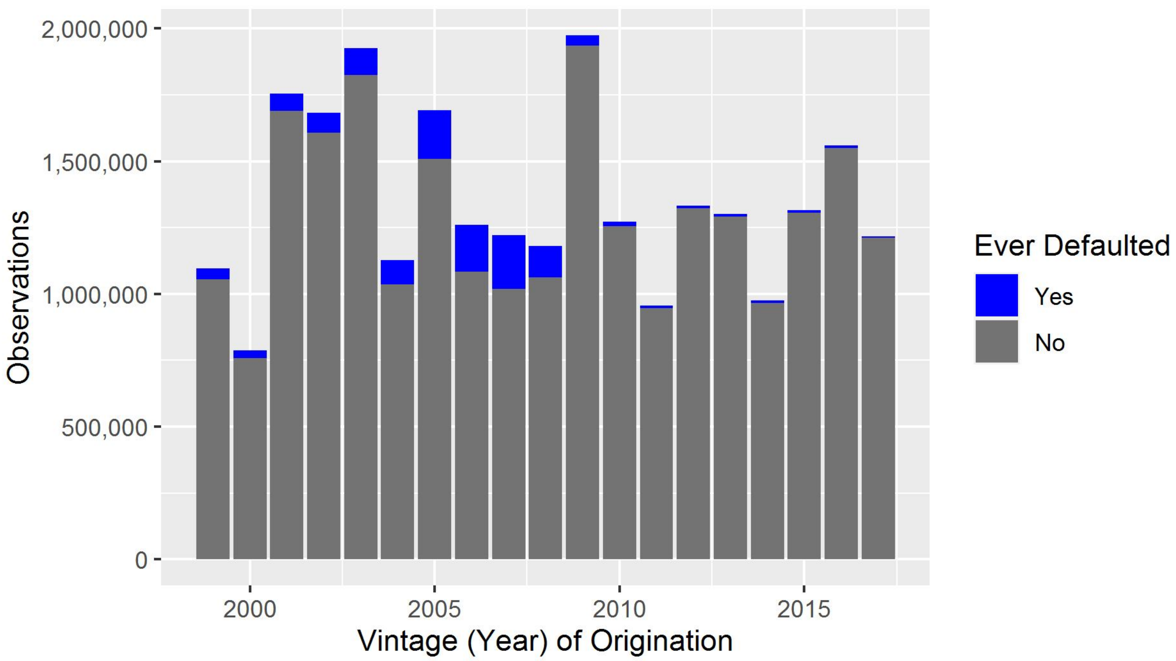

Figure 4: States Default Rate Crisis/Pre-Crisis Ratio
The figure displays the ratio between average yearly default rate before the Great Financial Crisis (GFC) and average yearly default rate during the GFC by States across the entire sample. Only few States experienced a default rate that less than doubled during the crisis, while mortgage default rate was at least three times than in non-crisis period for the majority of US Sates. California, Nevada, Florida, Arizona mortgage default rate was six-times than observed in non-crisis. Among these, only California and Arizona are mortgage non-recourse states.

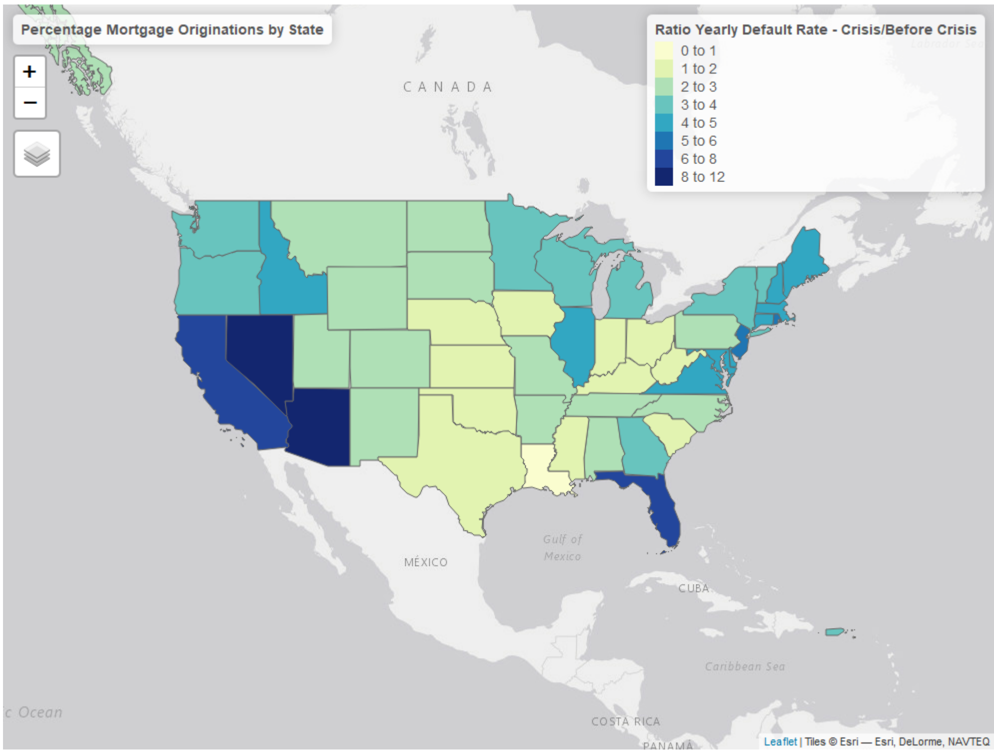

Table 8: Default Probability: Logistic Regression

The table reports the estimated coefficients for the logistic regressions and their robust standard errors. The bottom of the table reports number of observations, AUROC, GINI and Pseudo-R^2. The dependent variable takes a value of 0 if the mortgage is active and a value of 1 if it is defaulted. The frequency of observation is one year. Ump12 (HPI12) is the 1-year Unemployment rate (HPI) growth rate at State level, while Ump is the Unemployment rate. Excess IR is the difference between Interest rate at origination and the average quarterly IR. Credit Score and Debt-to-Income (DTI) are measured at origination. Updated LTV is the ratio between $Balance_t$ and $PropertyPrice_t$, which is derived based on HPI_t. Non-Recourse, First-Time Homebuyer an Joint are control variables in addition to the Fixed Effects (FE) listed at the bottom of the table. Dummy Crisis (DC) is activated during the years of mortgage crisis (2009, 2010 and 2011). The Full sample includes mortgages originated from 1999 to 2017 and observed from 1999 to 2018. The Reduced sample includes mortgages originated from 1999 to 2011 and observed from 1999 to 2018. *** p<0.01; ** p<0.05; * p<0.1.

	Full Sample		Reduced Sample	
Variables	$Model2$	$Model3$	$Model2$	$Model3$
HPI12	-0.962***		-0.729***	
Ump	0.0283***		-0.00341***	
Ump12		0.135***		0.123***
Original Credit Score	-0.0112***	-0.0110***	-0.0110***	-0.0110***
Original Debt-to-Income	0.0173***	0.0173***	0.0159***	0.0161***
Updated LTV	0.0322***	0.0331***	0.0336***	0.0331***
Excess Interest Rate	0.578***	0.581***		
Non-Recourse	-0.00344	0.00131	0.00467	-0.00193
First-Time Homebuyer	-0.0784***	-0.0739***	-0.0503***	-0.0455***
Joint	-0.611***	-0.613***	-0.640***	-0.642***
DummyCrisis(DC)	-1.136***	-1.238***	-0.674***	-0.967***
DC*CreditScore	0.00124***	0.00139***	0.000452***	0.000659***
DC*OriginalDTI	0.00567***	0.00558***	0.00709***	0.00690***
DC*UpdatedLTV	-0.00133***	-0.000982***	-0.00106***	-1.89E-05
DC*ExcessIR	-0.0976***	-0.0837***		
DC*Non-Recourse	0.176***	0.196***	0.153***	0.179***
DC*First-Time Homebuyer	-0.0644***	-0.0546***	-0.0782***	-0.0664***
DC*Joint	0.156***	0.149***	0.158***	0.152***
Constant	1.227***	1.302***	1.665***	1.651***
BEA Territories FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
LoanAge	Yes	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes	Yes
Occupancy FE	Yes	Yes	Yes	Yes
N. Units FE	Yes	Yes	Yes	Yes
Property Type FE	Yes	Yes	Yes	Yes
Observations	122082801	122082801	92799508	92799508
AUROC	0.879	0.8795	0.8618	0.8625
GINI	0.758	0.759	0.7236	0.725
Pseudo-R^2	0.2009	0.2037	0.1815	0.1842

Figure 5: Correlation All

The graph displays the distribution of default correlation ρ (y-axis) by PD Long-run (x-axis). The correlation is calculated on a synthetic portfolio generated by combinations of a subset of variables available in the sample ad used in the logistic regression. This is due to data-imaging constraints. As some dimensions are kept fixed, the range of default correlation ρ is reduced and spans from 0% up to 7%. The graph clearly shows the lack of proportional relationship between correlation ρ and PD Long-run. Moreover, increasing risk seems to bring to lower correlation ρ, in a similar fashion as per corporate portfolios, where correlation ρ is set to be a decreasing function of implied risk.

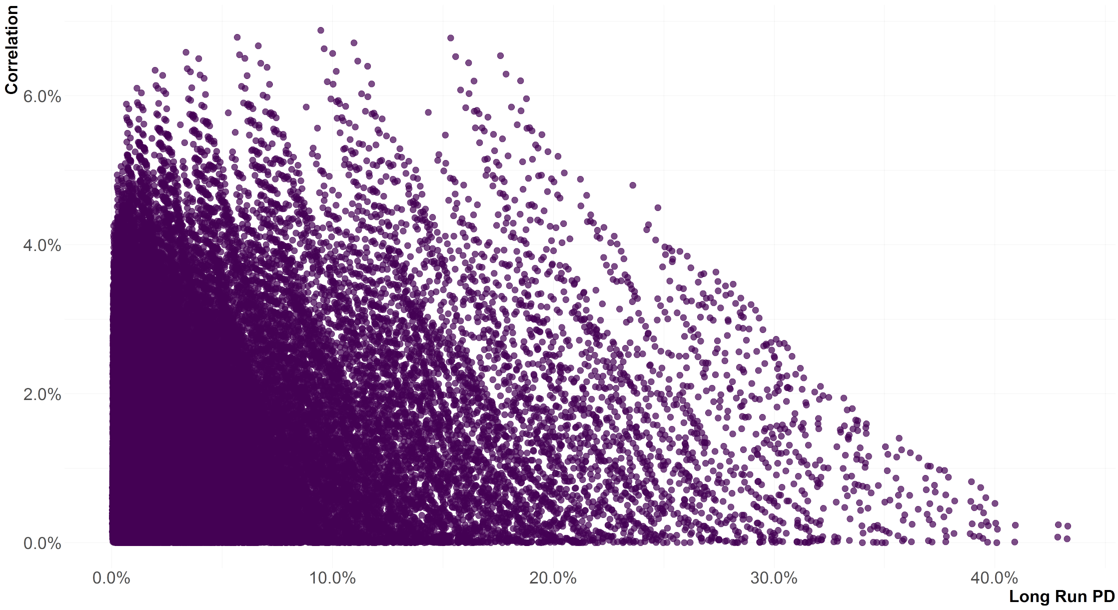

Figure 6: BEA Territories and Updated LTV

The graph unfolds the relationship between default correlation ρ (y-axis) by PD Long-run (x-axis) after breaking down by Updated LTV and BEA Territories. Updated LTV is discretised from 30% to 120% by 30, while BEA regions range across all possible values. The graph highlights the evident relationship between Updated LTV and increasing correlation, which however needs to be carefully considered as the increasing Updated LTV is a consequence of the falling house prices that have characterised the aftermath of GFC on housing market. Given that the effect of housing market is captured by Updated LTV, the graph highlights that default correlation ρ differentiation is also dependant on geography.

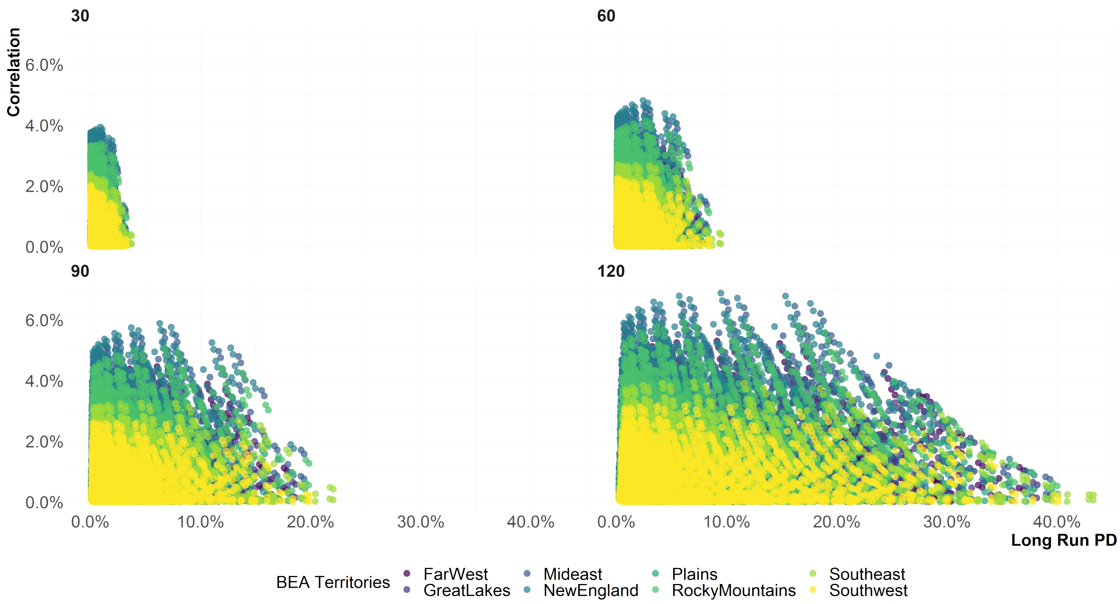

Figure 7: Credit Score and Number of Borrowers
The graph unfolds the relationship between default correlation ρ (y-axis) by PD Long-run (x-axis) after breaking down by Credit Score and Number of Borrowers. Single-borrower mortgages experience a higher PD long-run and a generally lower default correlation ρ, compared with Joint applications. Credit Score ρ pattern is characterised by some differences between Number of Borrowers. Correlation values tend to increase with decreasing credit score for Joint mortgages, while for Single borrowers it is characterised by a hump shape with lower values for sub-prime mortgages. This is most probably linked to the deteriorating long-run PD for these observations, which makes mortgages less sensitive to adversely changing economic conditions.

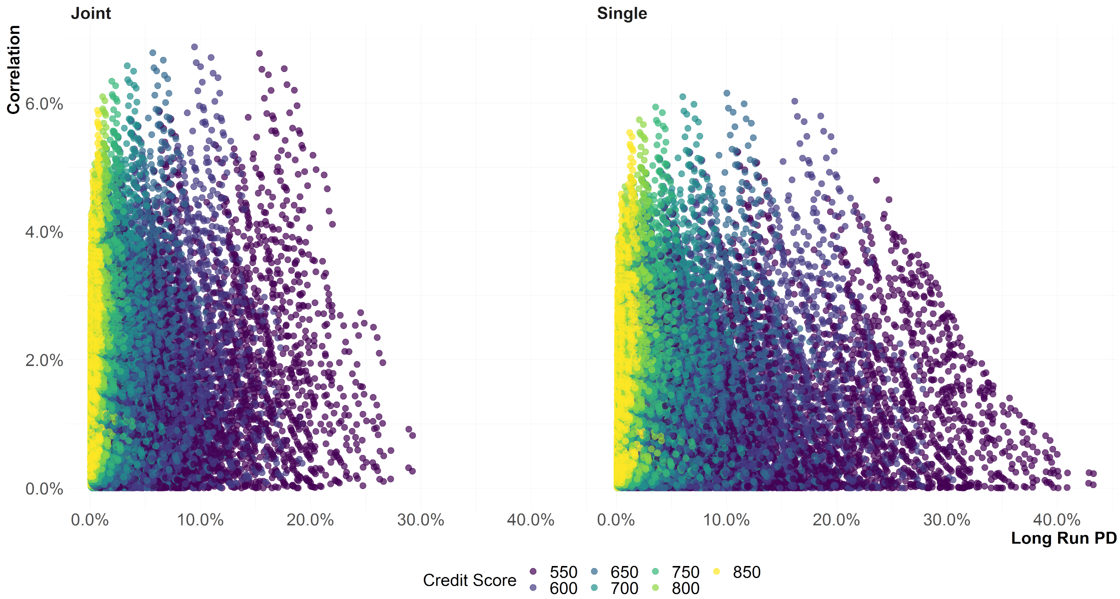

Figure 8: BEA Territories and Sellers
The graph unfolds the relationship between default correlation ρ (y-axis) by PD Long-run (x-axis) after breaking down by Bureau of Economic Analysis (BEA) Regions and Sellers. While BEA regions behave similarly, Sellers mortgage default correlations ρ display different patterns in relation with PD long-run. Bank of America, Branch banking, Chase, Citi, Sun trust and US bank have generally higher correlations compared with smaller banks, characterised by a different concentration across US States.

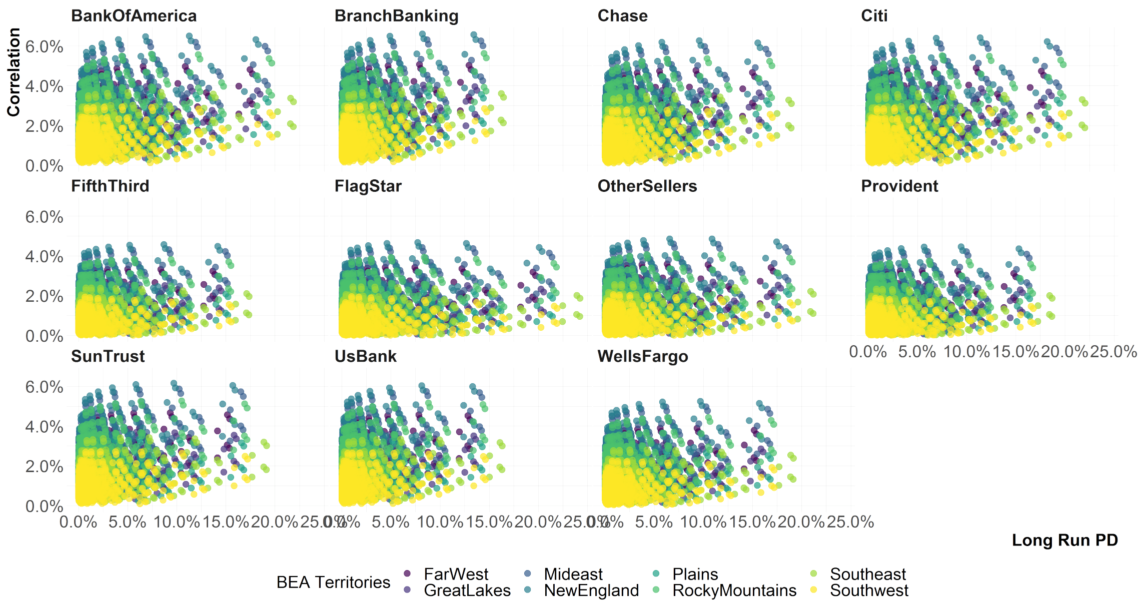

Table 9: Average Default Correlation
The table displays the average default correlation ρ across different segments and combined by characteristics. The x-dimension includes Credit Score, Debt-to-Income and Loan Purpose (Cash-out Refinance (C), No Cash-out Refinance (N), Purchase (P)). The y-dimension include BEA Territories, Updated Loan-to-Value, 1^{st} time homebuyer flag, Non-recourse, Property type and Seller. The correlations are calculated on a synthetic portfolio created with all possible combinations of mortgage characteristics entering the Logistic regression model. Continuous variables are discretised, while categorical variables keep all possible values. The variables and respective values that are not included in the x-axis and y-axis contribute to the variance in observed default correlation ρ.

		Credit Score				Debt-to-Income				Purpose		
Variable	Segments	550	650	750	850	30	40	50	60	C	N	P
BEA Territories	FarWest	1.405	1.528	1.605	1.685	1.502	1.751	2.03	2.34	1.674	1.743	1.26
	GreatLakes	1.01	1.137	1.232	1.327	1.127	1.332	1.564	1.825	1.27	1.332	0.933
	Mideast	1.882	2.012	2.06	2.111	1.96	2.261	2.593	2.957	2.164	2.241	1.661
	NewEngland	2.02	2.139	2.177	2.222	2.082	2.394	2.739	3.116	2.294	2.371	1.77
	Plains	0.868	0.995	1.095	1.193	0.991	1.178	1.392	1.634	1.123	1.182	0.815
	Rocky Mountains	1.627	1.763	1.827	1.892	1.723	1.998	2.305	2.643	1.911	1.985	1.451
	Southeast	0.878	1.013	1.113	1.211	1.007	1.198	1.417	1.663	1.141	1.202	0.827
	Southwest	0.64	0.757	0.86	0.963	0.762	0.918	1.098	1.304	0.873	0.926	0.619
	US Territories	1.99	2.197	2.247	2.279	2.128	2.449	2.801	3.185	2.341	2.425	1.801
Updated LTV	0	0.986	1.083	1.189	1.304	1.102	1.268	1.454	1.662	1.214	1.267	0.936
	20	1.09	1.177	1.275	1.383	1.187	1.372	1.581	1.816	1.315	1.371	1.005
	40	1.219	1.296	1.381	1.479	1.292	1.502	1.739	2.007	1.44	1.499	1.089
	60	1.374	1.444	1.514	1.597	1.422	1.661	1.933	2.241	1.594	1.656	1.194
	80	1.548	1.626	1.679	1.742	1.582	1.855	2.166	2.518	1.78	1.848	1.323
	100	1.692	1.841	1.885	1.922	1.771	2.081	2.427	2.808	1.987	2.067	1.476
	120	1.672	2.064	2.133	2.148	1.972	2.3	2.652	3.022	2.174	2.275	1.638
1^{st} Time Homebuyer	Yes	1.27	1.401	1.481	1.56	1.378	1.61	1.872	2.162	1.538	1.605	1.152
	No	1.467	1.608	1.678	1.747	1.573	1.829	2.115	2.431	1.748	1.819	1.322
Recourse	Yes	1.048	1.179	1.27	1.36	1.166	1.374	1.61	1.875	1.311	1.374	0.967
	No	1.69	1.83	1.889	1.947	1.786	2.066	2.376	2.718	1.976	2.05	1.508
Property Type	CO	1.912	2.051	2.103	2.156	1.998	2.307	2.648	3.022	2.208	2.287	1.691
	CP	0.391	0.495	0.6	0.709	0.512	0.628	0.765	0.923	0.596	0.64	0.409
	MH	0.691	0.839	0.955	1.062	0.84	1.016	1.219	1.451	0.965	1.024	0.678
	PU	2.218	2.347	2.377	2.412	2.28	2.616	2.986	3.39	2.507	2.589	1.941
	SF	1.632	1.79	1.862	1.929	1.749	2.033	2.348	2.696	1.942	2.019	1.468
Seller	BankOfAmerica	1.653	1.812	1.874	1.931	1.766	2.045	2.355	2.695	1.955	2.03	1.489
	BranchBanking	1.762	1.876	1.929	1.988	1.833	2.115	2.429	2.774	2.026	2.099	1.553
	Chase	1.598	1.733	1.797	1.861	1.694	1.964	2.264	2.594	1.878	1.95	1.428
	Citi	1.619	1.779	1.843	1.902	1.734	2.01	2.316	2.652	1.921	1.996	1.46
	FifthThird	1.024	1.146	1.237	1.328	1.136	1.338	1.568	1.826	1.277	1.338	0.942
	FlagStar	0.997	1.144	1.239	1.327	1.129	1.336	1.57	1.834	1.273	1.336	0.932
	Provident	1.086	1.226	1.316	1.402	1.209	1.424	1.666	1.938	1.358	1.422	1.003
	SunTrust	0.978	1.105	1.197	1.289	1.095	1.294	1.519	1.773	1.234	1.295	0.907
	UsBank	1.558	1.695	1.76	1.826	1.657	1.923	2.219	2.545	1.838	1.91	1.395
	WellsFargo	1.548	1.664	1.73	1.8	1.631	1.892	2.182	2.504	1.811	1.88	1.376
	Others	1.236	1.37	1.452	1.534	1.348	1.578	1.837	2.125	1.507	1.573	1.125

Table 10: Standard Deviation Correlation
The table displays the standard deviation default correlation across different segments and combined by characteristics. The x-dimension includes Credit Score, Debt-to-Income and Loan Purpose (Cash-out Refinance (C), No Cash-out Refinance (N), Purchase (P)). The y-dimension include BEA Territories, Updated Loan-to-Value, 1^{st} time homebuyer flag, Non-recourse, Property type and Seller. The correlations are calculated on a synthetic portfolio created with all possible combinations of mortgage characteristics entering the Logistic regression model. Continuous variables are discretised, while categorical variables keep all possible values. The variables and respective values that are not included in the x-axis and y-axis contribute to the variance in observed default correlation ρ.

		Credit Score				Debt-to-Income				Purpose		
d	Segments	550	650	750	850	30	40	50	60	C	N	P
BEA Territories	FarWest	1.18	1.16	1.09	1.02	0.97	1.09	1.22	1.37	1.17	1.19	0.93
	GreatLakes	0.95	0.95	0.91	0.87	0.80	0.91	1.03	1.16	0.97	0.99	0.77
	Mideast	1.44	1.42	1.31	1.21	1.17	1.31	1.46	1.63	1.40	1.42	1.14
	NewEngland	1.50	1.47	1.36	1.24	1.22	1.36	1.51	1.68	1.46	1.47	1.18
	Plains	0.86	0.88	0.85	0.81	0.74	0.84	0.95	1.08	0.90	0.92	0.70
	Rocky Mountains	1.31	1.30	1.21	1.12	1.07	1.21	1.35	1.51	1.29	1.31	1.04
	Southeast	0.88	0.90	0.87	0.83	0.75	0.86	0.98	1.11	0.92	0.94	0.72
	Southwest	0.70	0.73	0.72	0.71	0.62	0.71	0.81	0.93	0.76	0.78	0.58
	US Territories	1.55	1.55	1.45	1.33	1.29	1.44	1.61	1.79	1.54	1.55	1.25
Updated LTV	0	0.84	0.82	0.81	0.8	0.75	0.74	0.74	0.73	1.43	1.45	1.16
	20	0.95	0.91	0.88	0.86	0.86	0.84	0.83	0.81	1.43	1.45	1.16
	40	1.08	1.02	0.97	0.94	0.97	0.95	0.93	0.91	1.37	1.39	1.11
	60	1.24	1.16	1.09	1.03	1.10	1.07	1.04	1.01	1.42	1.43	1.14
	80	1.41	1.33	1.23	1.15	1.25	1.20	1.16	1.12	1.05	1.07	0.83
	100	1.54	1.52	1.41	1.3	1.40	1.35	1.29	1.24	1.09	1.10	0.85
	120	1.59	1.71	1.62	1.47	1.57	1.51	1.44	1.38	1.12	1.14	0.88
1^{st} Time Homebuyer	Yes	1.22	1.22	1.16	1.09	1.04	1.17	1.32	1.47	1.03	1.05	0.81
	No	1.35	1.35	1.27	1.18	1.15	1.29	1.44	1.6	1.36	1.37	1.09
Recourse	Yes	1.04	1.05	1.01	0.96	0.89	1.01	1.15	1.29	1.32	1.34	1.07
	No	1.43	1.42	1.32	1.23	1.20	1.34	1.49	1.66	1.18	1.20	0.95
Property Type	CO	1.30	1.28	1.19	1.09	1.02	1.15	1.28	1.43	1.27	1.28	1.02
	CP	0.45	0.48	0.49	0.49	0.41	0.47	0.54	0.62	0.51	0.53	0.39
	MH	0.73	0.73	0.72	0.69	0.60	0.69	0.79	0.91	0.76	0.77	0.58
	PU	1.42	1.41	1.29	1.18	1.12	1.25	1.40	1.55	1.38	1.39	1.12
	SF	1.18	1.18	1.11	1.02	0.95	1.06	1.19	1.34	1.17	1.19	0.94
Seller	BankOfAmerica	1.43	1.43	1.35	1.24	1.21	1.36	1.51	1.68	1.43	1.45	1.16
	BranchBanking	1.46	1.43	1.33	1.23	1.21	1.35	1.51	1.67	1.43	1.45	1.16
	Chase	1.38	1.38	1.29	1.19	1.16	1.30	1.45	1.62	1.37	1.39	1.11
	Citi	1.41	1.42	1.33	1.23	1.20	1.34	1.50	1.67	1.42	1.43	1.14
	FifthThird	1.03	1.03	0.99	0.94	0.88	0.99	1.12	1.26	1.05	1.07	0.83
	FlagStar	1.05	1.06	1.03	0.97	0.90	1.03	1.16	1.32	1.09	1.10	0.85
	Provident	1.08	1.10	1.05	0.99	0.93	1.06	1.19	1.34	1.12	1.14	0.88
	SunTrust	1.00	1.01	0.97	0.93	0.86	0.98	1.11	1.25	1.03	1.05	0.81
	UsBank	1.36	1.36	1.27	1.18	1.14	1.28	1.44	1.6	1.36	1.37	1.09
	WellsFargo	1.34	1.32	1.23	1.15	1.12	1.25	1.40	1.56	1.32	1.34	1.07
	Others	1.17	1.17	1.11	1.05	1.00	1.12	1.26	1.41	1.18	1.20	0.95

Table 11: Maximum Default Correlation
The table displays the maximum default correlation across different segments and combined by characteristics. The x-dimension includes Credit Score, Debt-to-Income and Loan Purpose (Cash-out Refinance (C), No Cash-out Refinance (N), Purchase (P)). The y-dimension include BEA Territories, Updated Loan-to-Value, 1^{st} time homebuyer flag, Non-recourse, Property type and Seller. The correlations are calculated on a synthetic portfolio created with all possible combinations of mortgage characteristics entering the Logistic regression model. Continuous variables are discretised, while categorical variables keep all possible values. The variables and respective values that are not included in the x-axis and y-axis contribute to the variance in observed default correlation ρ.

		Credit Score				Debt-to-Income				Purpose		
Variable	Segments	550	650	750	850	30	40	50	60	C	N	P
BEA Territories	FarWest	8.8	8.56	8.19	7.55	6.2	7.01	7.88	8.8	8.63	8.8	7.01
	GreatLakes	7.44	7.2	6.9	6.4	5.06	5.79	6.57	7.44	7.44	7.4	5.78
	Mideast	11.17	10.19	9.79	8.99	7.58	8.48	9.43	11.17	11.17	10.42	8.48
	NewEngland	10.8	10.55	10.12	9.27	7.9	8.82	9.79	10.8	10.61	10.8	8.81
	Plains	7.99	6.68	6.42	5.99	4.65	5.33	6.08	7.99	7.99	6.87	5.32
	Rocky Mountains	11.06	9.41	9.04	8.33	6.91	7.77	8.68	11.06	11.06	9.63	7.77
	Southeast	11.43	6.84	6.59	6.16	4.77	5.47	7.05	11.43	11.43	8.15	5.44
	Southwest	7.51	5.73	5.5	5.18	3.88	4.49	5.16	7.51	7.51	5.88	4.47
	US Territories	23.27	10.95	10.64	9.95	8.24	12.57	17.72	23.27	23.27	19.12	12.56
Updated LTV	0	5.7	5.34	5.14	5.12	3.97	4.44	5.04	5.7	5.63	5.7	4.45
	20	6.54	6.05	5.69	5.5	4.45	5.07	5.77	6.54	6.46	6.54	5.1
	40	7.47	6.92	6.4	6.03	5.11	5.82	6.61	7.47	7.36	7.47	5.85
	60	8.4	7.93	7.28	6.75	5.84	6.63	7.49	8.4	8.27	8.4	6.65
	80	9.29	8.99	8.34	7.63	6.62	7.46	8.36	9.29	9.15	9.29	7.46
	100	10.23	9.99	9.51	8.7	7.41	8.3	9.24	10.23	10.07	10.23	8.29
	120	23.27	10.95	10.64	9.95	8.24	12.57	17.72	23.27	23.27	19.12	12.56
1^{st} Time Homebuyer	Yes	20.47	10.12	9.81	9.15	7.52	10.22	15.09	20.47	20.47	16.45	10.2
	No	23.27	10.95	10.64	9.95	8.24	12.57	17.72	23.27	23.27	19.12	12.56
Recourse	Yes	20.69	8.54	8.27	7.76	6.17	10.33	15.25	20.69	20.69	16.63	10.31
	No	23.27	10.95	10.64	9.95	8.24	12.57	17.72	23.27	23.27	19.12	12.56
Property Type	CO	17.63	10.08	9.79	9.18	7.49	8.35	12.62	17.63	17.63	13.86	8.3
	CP	4.29	4.21	4.06	3.89	2.69	3.16	3.7	4.29	4.17	4.29	3.14
	MH	23.27	6.21	6.02	5.75	8.05	12.57	17.72	23.27	23.27	19.12	12.56
	PU	17.58	10.95	10.64	9.95	8.24	9.15	12.6	17.58	17.58	13.83	9.1
	SF	20.33	9.35	9.1	8.57	6.83	10.29	15.07	20.33	20.33	16.39	10.28
Seller	BankOfAmerica	22.22	10.89	10.64	9.95	8.11	11.8	16.8	22.22	22.22	18.16	11.79
	BranchBanking	13.55	10.95	10.62	9.78	8.24	9.15	10.07	13.55	13.55	11.08	9.1
	Chase	17.26	10.53	10.26	9.51	7.85	8.71	12.23	17.26	17.26	13.49	8.68
	Citi	22.39	10.78	10.53	9.85	8.01	11.94	16.95	22.39	22.39	18.31	11.92
	FifthThird	12.48	8.27	8.02	7.45	5.92	6.69	8	12.48	12.48	9.12	6.64
	FlagStar	23.27	8.51	8.28	7.83	8.05	12.57	17.72	23.27	23.27	19.12	12.56
	Provident	18.97	8.72	8.5	7.96	6.26	9.06	13.74	18.97	18.97	15.06	9.05
	SunTrust	14.47	8.15	7.92	7.38	5.81	6.56	9.71	14.47	14.47	10.9	6.53
	UsBank	17.44	10.4	10.13	9.4	7.73	8.59	12.39	17.44	17.44	13.65	8.56
	WellsFargo	12.67	10.2	9.89	9.11	7.59	8.47	9.36	12.67	12.67	10.33	8.42
	Others	16.29	9.22	8.97	8.34	6.71	7.51	11.34	16.29	16.29	12.58	7.48

Table 12: Excess Interest Rate: Linear Regression
The table reports the estimated coefficients for the linear regressions and their robust standard errors. The bottom of the table reports number of observations, R^2, Adjusted-R^2 and AIC. The dependent variable is the Excess (delta) IR from the average Interest Rate by quarter of origination. The frequency of observation is quarterly. ρ is the mortgage default correlation derived from the Logistic Regression on the reduced sample. 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. The Full sample includes mortgages originated from 2012 to 2017. *** $p<0.01$; ** $p<0.05$; * $p<0.1$.

Variables	Model 1	Model 2	Model 3
Credit Score	-0.0019***	-0.0019***	-0.0019***
Original LTV	0.0060***	0.0061***	0.0061***
Original DTI	0.0018***	0.0024***	0.0022***
HPI12	0.7355***	0.6789***	0.6842***
Ump12	-0.0172***	-0.0180***	-0.0182***
First-Time Homebuyer	0.0110***	0.0060***	0.0068***
Joint	-0.0522***	-0.0464***	-0.0478***
Non Recourse	-0.0389***	-0.0291***	-0.0320***
ρ	3.8089***	1.4315***	0.2446*
ρ*BranchBanking			-0.0213
ρ*Chase			-1.1298***
ρ*Citi			-1.0295***
ρ*FifthThird			4.0675***
ρ*Others			2.5920***
ρ*Provident			-2.9902***
ρ*SunTrust			2.0013***
ρ*UsBank			2.3045***
ρ*WellsFargo			1.5893***
Constant	1.3802***	1.3526***	1.3762***
Bank FE	No	Yes	Yes
BEA Territories FE	Yes	Yes	Yes
N. Units FE	Yes	Yes	Yes
Channel FE	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes
Occupancy FE	Yes	Yes	Yes
Observations	7677015	7677015	7677015
R^2	0.2289	0.2353	0.2355
Adjusted-R^2	0.2289	0.2352	0.2355
AIC	7329391.087	7266292.611	7263753.205

Figure 9: Excess Interests Paid

The graph shows the maximum difference amongst banks in pricing the effect of mortgage default correlation ρ. 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 default correlation ρ 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 \$ 7,000, suggesting that banks do not really differentiate in pricing default correlation ρ. However, the right tail of the distribution shows that the difference in pricing outstretches up to \$ 40,000. This implies that banks tend to price in a different way specific segments in the mortgage market based on default correlation risk experienced by specific clusters.

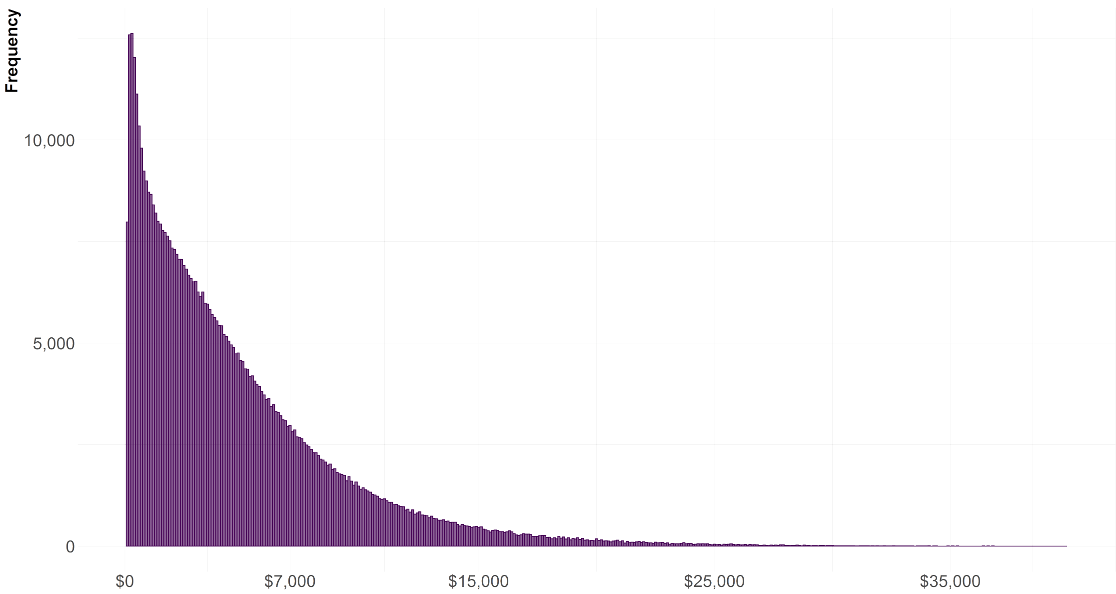

Figure 10: Excess Interest Paid by BEA Territories

The graph shows the maximum difference amongst banks in pricing the effect of mortgage default correlation ρ, 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 default correlation ρ 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 clearly shows geographic patterns exists in pricing mortgage default correlation ρ, especially where the lending is concentrated in Midwest, New England, Rocky mountains and US territories. In these territories, financial institutions tend to price default correlation ρ hence giving an incentive to borrowers in shopping around, as total interests charged can differ significantly.

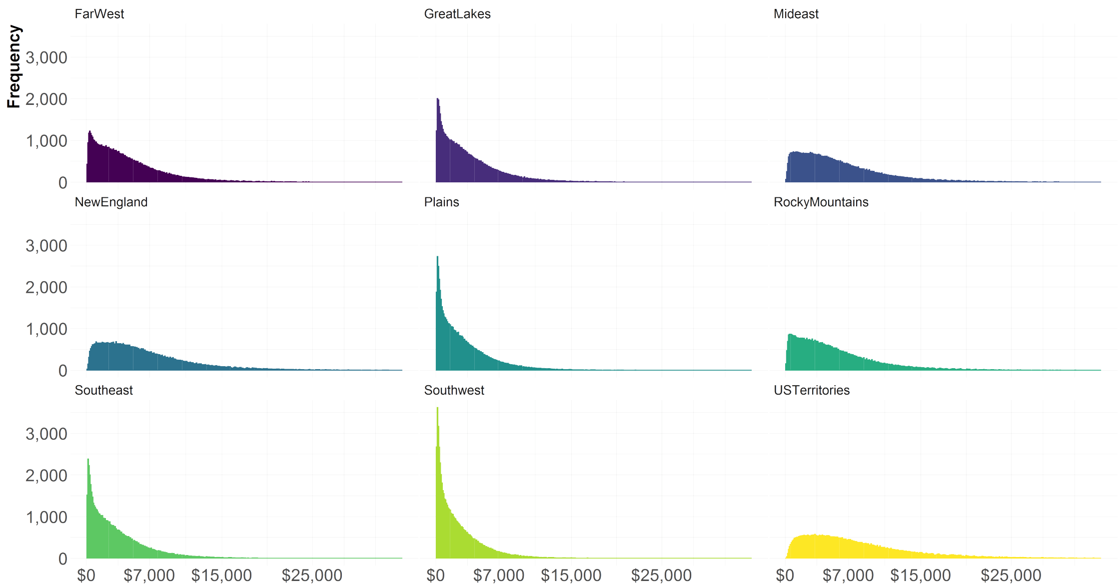

Figure 11: Excess Interest Paid by Original LTV

The graph shows the maximum difference amongst banks in pricing the effect of mortgage default correlation ρ, 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 default correlation ρ 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 financial institutions price mortgage default correlation ρ by increasing Original LTV in a somehow similar way. Financial institutions are therefore aware that mortgages with higher LTVs tend to cluster and experience a contagion effect under adverse economic conditions, hence pricing it quite similarly amongst them. Borrowers might not significantly benefit if they want to be priced differently based on their Original LTV.

Figure 12: Excess Interest Paid by Credit Score
The graph shows the maximum difference amongst banks in pricing the effect of mortgage default correlation ρ, 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 default correlation ρ 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 financial institutions price mortgage default correlation ρ by decreasing Credit Score in a somehow similar way. Financial institutions are therefore aware that mortgages with lower Credit Scores tend to cluster and experience a contagion effect under adverse economic conditions, hence pricing it quite similarly amongst them. Borrowers might not significantly benefit if they want to access lower interest charge based on their Credit Score.

Figure 13: Excess Interest Paid by Non-Recourse

The graph shows the maximum difference amongst banks in pricing the effect of mortgage default correlation ρ, by breaking down by Recourse and Non-Recourse states. 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 default correlation ρ 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 graphs highlights that financial institutions experience a higher variability in pricing mortgage default correlation ρ for non-recourse states. Most likely it can be attributed to a different business capillarity in non-recourse states, resulting on a disparity when pricing of mortgage default correlation ρ.

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*

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*

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*

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* .

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**.

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

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.

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

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**.

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.

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*

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.

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.

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*

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*

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*

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.