2022

Understanding the Risks of Multiple Credit Scores in Mortgage Lending

Dr. Clifford Rossi
Chesapeake Risk Advisors, LLC
8/1/2022

Understanding the Risks of Multiple Credit Scores in Mortgage Lending

Executive Summary

For decades, credit scores have been a mainstay in consumer lending, including mortgages. During much of this time credit investors have relied on a single credit provider, FICO, for credit scores. VantageScore's entry into the credit scoring market in 2006 introduced competition in this area and over time realignment of their credit scores into the same 300-850 score range as FICO has furthered the potential for score interchangeability among score users.

Claims by VantageScore that their VantageScore 4 score could effectively expand access to credit for historically unscorable consumers heightened interest by policymakers and industry on how to leverage credit scores for more equitable lending. One idea that has been circulated by the Federal Housing Finance Agency (FHFA) has been to consider allowing one of four options for multiple credit scores used in the GSEs' mortgage automated underwriting scoring systems. Each of the options would introduce a degree of competition in the provision of credit scores that until now has been absent from GSE-eligible mortgage underwriting.

While competition is inherently beneficial to markets and consumers, it can present a variety of risks that if not accounted for could be detrimental to consumers and credit investors alike. Of interest to this study is understanding the effects of alternative credit scoring models on credit risk, profitability and effectiveness at expanding access to credit. Specifically, leveraging a large loan level sample of GSE loans originated between 1999-2015, the analysis focused on demonstrating how mortgage scores developed by two hypothetical mortgage score providers would affect these outcomes.

The implications for credit risk managers and policymakers are the following:

1/ Credit risk may differ significantly across scores from different providers despite being depicted in the same score range.

2/ Alignment of credit risk across different credit scores is possible at a point in time but can deteriorate over time and with varying economic conditions.

3/ Scores developed on limited performance history, while perhaps fulfilling minimal model performance requirements, may not result in a significant uptake in acceptable loans based on model error and credit risk considerations.

4/ Financial performance can deviate widely by score. Profitability of mortgage borrowers with much less performance history than typical borrowers is much worse (negative) for the same policy score cutoff used for typical borrowers, where profitability is positive.

Reliance on scores that are highly effective in distinguishing between good and bad loans (not just minimally viable) is the best way of ensuring credit risk is managed prudently while expanding access to credit. Also, when using multiple scores, there are other drawbacks such as operational challenges in business and risk systems, and the potential to dilute predictive performance. To mitigate third-party and credit risk, the analysis suggests that credit investors need to perform their own analysis of alternative scores and confirm the claims of their providers to expand credit while managing credit risk.

Understanding the Risks of Multiple Credit Scores in Mortgage Lending

Background

For decades, credit scores have been an integral part of credit risk management in the mortgage industry. The adoption of automated underwriting scorecards by Freddie Mac in 1996 and followed closely by Fannie Mae ushered in a more analytically driven risk management framework that revolutionized the way credit scores are used to manage mortgage credit risk. Statistically based underwriting models were demonstrated to effectively rank order borrower credit risk consistently and accurately compared with manual underwriting. Integrating credit scores with other key underwriting variables such as loan-to-value (LTV) ratio and debt-to-income (DTI) ratio, among others enabled credit investors such as Fannie Mae and Freddie Mac to holistically evaluate a borrower's credit profile. Even the Federal Housing Administration (FHA), an agency skeptical of credit scores, adopted automated underwriting in 1996. From there, credit scoring became nearly ubiquitous across the mortgage spectrum.

Credit scores are used in a variety of ways by credit investors and lenders. Beyond their use in automated underwriting, credit policies typically incorporate credit scores into their eligibility criteria and are used for risk-based pricing, product development, best execution strategy in secondary markets, pipeline management and hedging, mortgage asset valuation (including mortgage servicing rights (MSRs)) and credit risk transfer (CRT) pricing. In the credit risk management area, credit scores have become a mainstay in conducting loan loss reserve analysis, loss forecasting, economic capital analysis and credit portfolio management, for example.

For most of this period all these activities have been based on a single provider of credit scores, FICO. FICO scores leverage detailed credit information on individuals from the major credit repositories to model credit delinquency. While FICO develops several credit scores for different applications, the mortgage industry for nearly two decades has relied on what has become known as Classic FICO. Classic FICO is not mortgage-specific but has been proven to be highly significant from a statistical perspective over time when modeling mortgage delinquency.

In 2006, the credit repositories TransUnion, Experian and Equifax established VantageScore, a company also focused on building analytically based credit scores. To that point, FICO dominated the credit score market. However, heightened interest by policymakers in expanding access to credit for borrowers with nontraditional credit or no credit history has the potential to disrupt the way the mortgage industry has relied on and used credit scores in credit decisioning. Consequently, finding ways to broaden the reach of credit scores for these consumer segments has important implications for the industry and consumers.

A study by the Consumer Financial Protection Bureau (CFPB) in 2015 examined credit records of 5 million individuals in their Consumer Credit Panel (CCP) sample.[1] From this analysis the CFPB determined that 26 million individuals were determined to be so-called "credit invisibles," or people without any credit record. An additional 19 million were determined to be unscorable by credit scoring models due to insufficient credit history or a relatively new credit history. Both FICO and VantageScore have conducted separate analyses of these credit segments as well and it has been well-established now in research that large numbers of credit invisibles and unscorable populations exist and merit focus on expanding their access to credit.[2]

Of particular interest is VantageScore's analysis of unscorable populations as a relatively new market entrant to the third-party credit score market. VantageScore's approach to analyzing unscorable consumers is to relax the criteria FICO applies in generating a credit score. FICO will only produce a credit score if there is at least 6 months' worth of credit usage history on file and the credit file has been updated within the last 6 months. VantageScore claims that an additional model in their scoring suite enables them to accurately evaluate consumers with thin credit files. In doing so, they further claim that they can assign a credit score for an additional 30-35 million individuals beyond FICO.[3] For the newly scored consumer segment, VantageScore claims that their model performance is well above industry standard benchmarks of a model's discriminatory power, i.e., ability to distinguish between events (defaults) and nonevents (nondefaults).[4] Further, VantageScore has produced their scores in the same 300-850 score range as FICO.

As part of a plan to consider replacing the GSEs' use of Classic FICO with either a FICO Score such as FICO 10 T or VantageScore 4, the FHFA offered 4 credit score options from which it solicited industry feedback.[5] These options were described by the FHFA as follows:

1/ Option 1 – Single Score – Either a FICO Score or VantageScore 4 would be required by the GSEs.

2/ Option 2 – Require Both Scores – the GSEs in this case would require that both a FICO Score and VantageScore 4 be delivered on each borrower.

3/ Option 3 – Lender Choice with Constraints – lenders would be allowed to select either a FICO Score or VantageScore 4 and deliver that score to the GSEs for some period of time.

4/ Option 4 – Waterfall – in this configuration the GSEs would permit multiple scores to be delivered. A primary score would be established and if that were not available, the system would revert to a secondary score; again, either a FICO Score or VantageScore 4.

In 2019, the Federal Housing Finance Agency (FHFA) issued a final rule, Validation and Approval of Credit Score Models used by Fannie Mae and Freddie Mac.[6] This rule sets forth the process by which the GSEs would assess and approve third-party credit scores used in their various activities. This rule set up the possibility for more than one third-party credit score to be used in evaluating GSE-eligible mortgage credit risk which has significant implications for the industry and mortgage risk management practices. More recently, the FHFA held a Listening Session on credit scores to explore ways to expand mortgage credit, including efforts to incorporate nontraditional forms of credit into the credit evaluation process. It is important to note that both FICO and VantageScore claim that their models are able to account for this type of information.

Study Objectives and Approach

Regulators and industry have for years discussed how to prudently expand access to mortgage credit. With improved modeling and competition among third-party credit score providers, the possibility exists to support this goal. However, the industry has had little experience managing credit risk in an environment where multiple credit scores for a borrower from different credit score providers can be used. This possibility has significant implications for credit investors, whether GSEs, portfolio lenders, private mortgage insurance companies or CRT credit investors. The focus of this study, therefore, is to present an empirical analysis comparing three different statistically based mortgage scores leveraging Classic FICO augmented with a number of other traditional credit risk attributes for different borrower cohorts.

The analysis is based on a scenario where two providers of different mortgage scores for a typical borrower exist (Score Provider 1 and 2).[7] Scores 1 and 2 are developed by Score Providers 1 and 2, respectively. These two mortgage scores are developed and validated from loans considered to be within the traditional set of historical GSE-eligible population of mortgage borrowers; specifically, borrowers with DTIs less than or equal to 43%, which aligns to the CFPB's original Qualified Mortgage Rule DTI threshold.

Score Provider 2 is assumed to have developed a new mortgage score (Score 3) that claims can substantially expand the market for a class of stretch borrowers that have historically not been served as effectively as the traditional borrower cohort. Score 3 is developed and validated from another set of GSE loans for borrowers having DTIs over 43%. The purpose of selecting the 43% DTI threshold is to provide a proxy of a borrower segment such as unscorables where loan performance history is not nearly as extensive as it is for traditional borrowers.[8] Score 3 thus serves as a proxy of third-party credit scoring models used to score unscorable consumers.

Each score is transformed into a common score range of 300-850 at a point in time (1999-2004). From there, analysis is conducted to demonstrate differences in model performance over time between scores and their implications on credit risk as well as a profitability analysis for each score. The analysis shows that while performance of Score 3 may exceed certain industry model performance benchmarks, the nature of that data and limited historical performance show demonstrably weaker model performance than either Scores 1 or 2. Moreover, even though all three scores are presented in the same score range, these differences in performance result in significant variations in credit risk. In other words, a score of 660 for Score 1-3 does not translate into the same risk. In fact, a score of 660 for Score 3 is shown to exhibit considerably higher credit risk than for either a 660 score for Scores 1 or 2. Further, once all three scores are calibrated in the initial 1999-2004 period, Scores 2 and 3 show marked deviation from Score 1's performance range over time

From purely a credit risk management perspective, great care must be taken in interpreting and using credit scores from different score providers. Users must assess the performance of different credit scores and set credit parameters in a consistent manner. That may result in tradeoffs between risk and volume. As exemplified by Score 3, setting the same policy score cutoff for Score 1 and 2 results in a significantly lower percent of approved >43% DTI borrowers than for <=43% DTI borrowers applying the same cutoff score for Score 1 and 2. In other words, the higher credit risk of >43% borrowers coupled with the poorer performance of the Score 3 model leads to a low pull-though rate of acceptable credit quality borrowers. These results imply that prima facie score comparability may mask important differences in risk across borrower segments that do not necessarily translate well into sustainable and prudent expansion of credit to nontraditional borrowers. This does not mean that credit scores cannot be developed to expand access to credit for nontraditional borrowers, but rather great care must be taken in evaluating the performance of third-party models.

Methodology and Data

Model Building and Score Transformation

The analysis presented in this study is based on an industry standard credit scoring methodology. Specifically, three statistically based logistic regression models are estimated using publicly available loan level credit performance data from the GSEs.[9] Each model is developed to predict the likelihood of ever becoming 90 days delinquent or worse (Ever D90+). This definition of delinquency is commonly used in developing mortgage credit risk models, including automated underwriting scorecards. Candidate risk factors considered in the analysis include those described in Table 1. Additional transformations for some of these risk factors were made to better represent their underlying relationship to Ever D90+. Additional details on these transformations, the data and the modeling can be found in the appendix.

Two models were estimated to generate Score 1 and Score 2. Both models leveraged the same GSE loan sample for loans originated between 1999-2015 with performance through 2021. Loans were randomly sampled from both Freddie Mac and Fannie Mae data and combined to reflect historical market shares for both GSEs during this period. Loans in this sample had DTIs less than or equal to 43%. When the original CFPB QM Rule was implemented, it featured a 43% maximum DTI to be considered QM-eligible. Historically, loans with DTIs greater than 43% comprised a smaller share of GSE loans. In the sample used in the analysis from the 1999-2015 period, these loans comprised approximately 22 percent of all loans sold to the GSEs.

Table 1

Candidate Risk Factors

Borrower:
Credit Score
LTV
DTI
Loan Balance
Number of Borrowers
Loan Purpose
First-time Homebuyer
Property:
Number of Units
Occupancy Type
Property Type
Other:
Origination Channel
Loan Term

The full sample for all DTIs had 144,458 loan level observations across these origination years after data cleaning. This sample was used to proxy for consumers with sufficient credit history to effectively develop a credit score. Score 1 was estimated intentionally to have the highest discriminatory power among all other scores to establish a benchmark of performance indicative of a leading mortgage score by Score Provider 1. Score 2, while built from the same sample as Score 1 leveraged the same candidate risk factors but in a different specification in order to represent a competitor score (Score Provider 2). The discriminatory power of Score 2 was slightly less than Score 1 so that both scores could be viewed as comparable from a user's perspective.

Score 3 was estimated using GSE data on 38,491 loans with DTIs above 43% between 1999-2015. Across this origination period 10.4% of the loans were Ever D90+ compared with 4.4% in the Score 1 and 2 sample. The set of candidate risk factors common to Score 1 and 2 were also used in specifying this model. Score 3's discriminatory power was considerably lower than Score 1 or 2, despite a large number of alternative model specifications. Score 3's Gini coefficient and other model performance metrics were, however, above industry standards of performance, a result consistent with industry consumer unscorable model performance as noted earlier.

Once the models were developed, they were validated against a different sample from the development data but comparable to both samples, i.e., <= 43% and >43% DTIs. From there, an estimated probability of Ever D90+ for every loan in the sample was produced for all three mortgage scoring models. Common with industry credit scoring techniques, these estimated probabilities were transformed into mortgage scores that ranged from 300 (highest Ever D90+ rates) to 850 (lowest Ever D90+ rates). In this framework every 50 points of mortgage score would double the odds of default.[10] For example, a borrower with a score of 600 would be twice as likely to become Ever D90+ than a borrower with a score of 650.

Score Alignment Process

One of the areas of interest for this study was to understand how scores change over time and their implications on credit risk. To conduct this portion of the study, all three scores were aligned to the 1999-2004 origination period. However, in order to have comparable seasoning between origination periods, the performance window was set as any loan ever becoming D90+ within 120 months of origination. In addition to 1999-2004, the pre-crisis period 2005-2007 along with the crisis period 2008-2010 were selected for comparison by scores in terms of credit performance and comparability.[11] The process meant aligning Scores 2 and 3 (competitor scores) with Score 1. The result was to ensure that the relationship between log odds and score (i.e., slope and intercept) were the same between Score 1, 2 and 3 for the 1999-2004 period.[12] The process used to align scores was the following:

Step 1: Group loans from the 1999-2004 period for each score into score buckets in 25-point increments. For example, 300-325, 325-350, etc. Calculate log odds and average score in each bucket.

Step 2: Estimate a simple linear regression for Score 1 (baseline score) with log odds of each score bucket as the dependent variable and average score in the bucket as the independent variable.

Step 3: Estimate linear regressions for Score 2 and 3 similar with Step 2.

Step 4: Calculate the alignment parameters for Score 2 and 3. These would be defined as the following:

$$\alpha_{21} = \frac{\alpha_2 - \alpha_1}{\beta_1}$$

$$\alpha_{31} = \frac{\alpha_3 - \alpha_1}{\beta_1}$$

$$\beta_{21} = \frac{\beta_2}{\beta_1}$$

$$\beta_{31} = \frac{\beta_3}{\beta_1}$$

In this form, α_{21} and α_{31} are the alignment intercept parameters and β21 and β31 are the alignment slope parameters for Scores 2 and 3, respectively. Likewise, α_1, α_2, α_3, are the intercept parameters from Steps 2 and 3 and β_1, β_2, and β_3 are the slope parameters from those steps for Scores 1-3, respectively.

Step 5: For each loan's score, calculate $\alpha_{i1} + \beta_{i1}$Score for Score i (2 or 3).

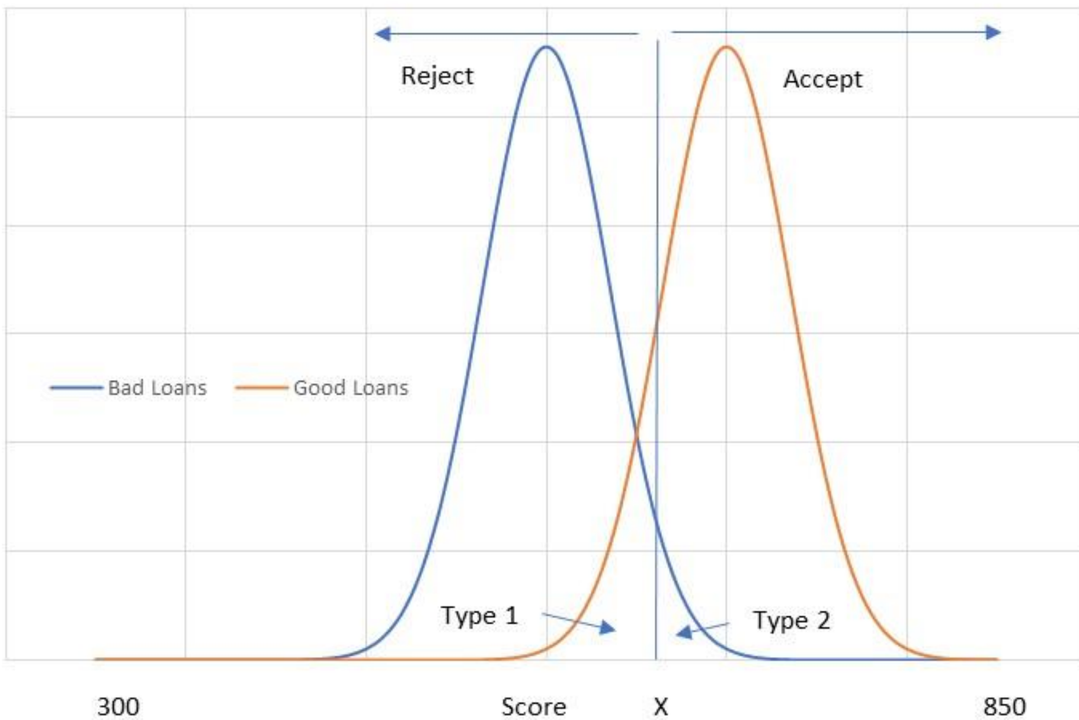
Step 6: Aggregate scores by bucket and rerun Steps 2 and 3 with these scores. This will ensure that the relationship between log odds and scores are identical between Scores 1-3.

Step 7: For the origination periods 2005-2007 and 2008-2010, separately apply the alignment parameters from Step 4 to loans in those two periods and rerun Step 5 and 6 for these loans.

Profitability Analysis

In addition to comparing score performance across scores and origination periods, a profitability analysis was designed to examine how using different scores that have the same score range could result in significantly different credit and financial outcomes across borrower cohorts. The analysis is based on the assumption that any model will result in some level of Type 1 (false positive) or Type 2 error (false negative). For the purposes of this analysis, a Type 1 error is one where an otherwise good (never default) loan is rejected and a Type 2 error occurs when a bad (default) loan is accepted against an imposed credit policy cutoff score. The amount of Type 1 and Type 2 error varies across models. Models with higher discriminatory power result in less Type 1 and Type 2 error than other models. The effect on credit investors from these errors is depicted in Figure 1. Figure 1 shows two loan distributions displayed against a mortgage score on the X-axis; the blue distribution represents all loans that default at some point and the red distribution represents loans that never default. Each loan in these distributions is scored against the model based on the collection of risk attributes described earlier. A

Figure 1: Effect of Type 1 and 2 Errors on Credit Risk



credit policy cutoff score X is applied against the scorecard which determines whether a loan is accepted or rejected. Type 1 error manifests under the portion of the nondefault distribution to the left of X, whereas Type 2 errors occur under the default distribution to the right of X. The analysis thus requires an estimate of the cost associated with both errors. The cost associated with a Type 2 error for a credit investor is the loss associated with that loan. Type 1 error is represented as the opportunity cost of foregone profit that would have been earned on each nondefault loan that should have been accepted but was not.

Credit loss was defined as the product of unpaid principal balance ($ UPB) for loan i, the loan's estimated probability of default and the loss severity rate (in percent). Data derived from Fannie Mae's Data Dynamics tool on historical mortgage loans in their credit portfolio were used to establish the loss severity estimates for <=43% and >43% DTI loans. Over time, loss severities vary but exhibited little deviation between DTI cohorts. The average loss severity over the 1999-2004 period was approximately 25% and this value was used in the analysis.

Annualized profitability (expressed in yield (%)) was defined as the following:

Annual Profit = Mortgage Note Rate – Credit Cost – Option-adjusted Spread – Servicing Cost – Origination Cost – Cost of Funding.

Credit Costs reflect estimated credit losses, Option-adjusted Spread (OAS) captures the effect of prepayment costs on the mortgage, and Cost of Funding reflects a blended debt and equity mix of financing (assumed for this analysis to follow the assumptions of Goodman and Zhu's analysis of 90% equity and 10% debt).[13] Assumed returns on equity (15%) and debt (6%) were also drawn from the Goodman and Zhu analysis. All-in funding costs were assumed to be 3.066%. The note rate assumed was 5.7% from the June 2022 Freddie Mac Primary Mortgage Market Survey rate for fixed-rate 30-year mortgages. Other assumptions from the Goodman and Zhu analysis used included servicing cost (5bps) and origination costs (50 bps). The OAS estimates of 115bps (<=43% DTIs) and 85bps (>43% DTIs) were drawn from June 2022 industry model estimates of OAS for similar pools of mortgages in GSE mortgage-backed securities. Finally, a lifetime profitability estimate for each loan was derived by multiplying the Annual Profit estimate by an estimated duration. The estimated duration for <43% and >43% DTI loans was 4.5 and 5.5 years, respectively. These estimates were also drawn from recent industry estimates on comparable mortgages.

Results

To gain a sense of the data used in the analysis, Table 2 presents some summary statistics on the three origination periods of interest. Default loan counts are defined as the number of loans in each period that were ever D90+ within 10 years of origination.[14] The data reveal two important patterns in the sample; first, it confirms the high default period of the 2005-2007; second it shows that the >43% DTI sample exhibits consistently higher default rates than <=43% DTI loans on an uncontrolled basis.

The three mortgage scores were estimated on their respective full samples' origination period 1999-2015 (i.e., the <=43% DTI sample for Scores 1 and 2 and >43% DTI for Score 3) and validated on samples of different loans from similar periods. Table 3 presents a summary of the model performance for each score. Scores 1 and 2 are very close across all measures and the results suggest a high degree of discriminatory power between default and nondefault loans. Score 3's model performance metrics indicate

Table 2: Summary Sample Statistics by Origination Period

>43% DTI Sample				
Period	Total (#)	# of NonDefault	# Default	Default %
1999-2004	11,912	11,474	438	3.68%
2005-2007	11,946	10,506	1,440	12.05%
2008-2010	8,149	7,575	574	7.04%
	32,007	29,555	2,452	
<=43% DTI Sample				
Period	Total (#)	# of NonDefault	# Default	Default %
1999-2004	42,267	41,246	1,021	2.42%
2005-2007	23,752	21,936	1,816	7.65%
2008-2010	29,753	29,090	663	2.23%
	95,772	92,272	3,500	

Table 3

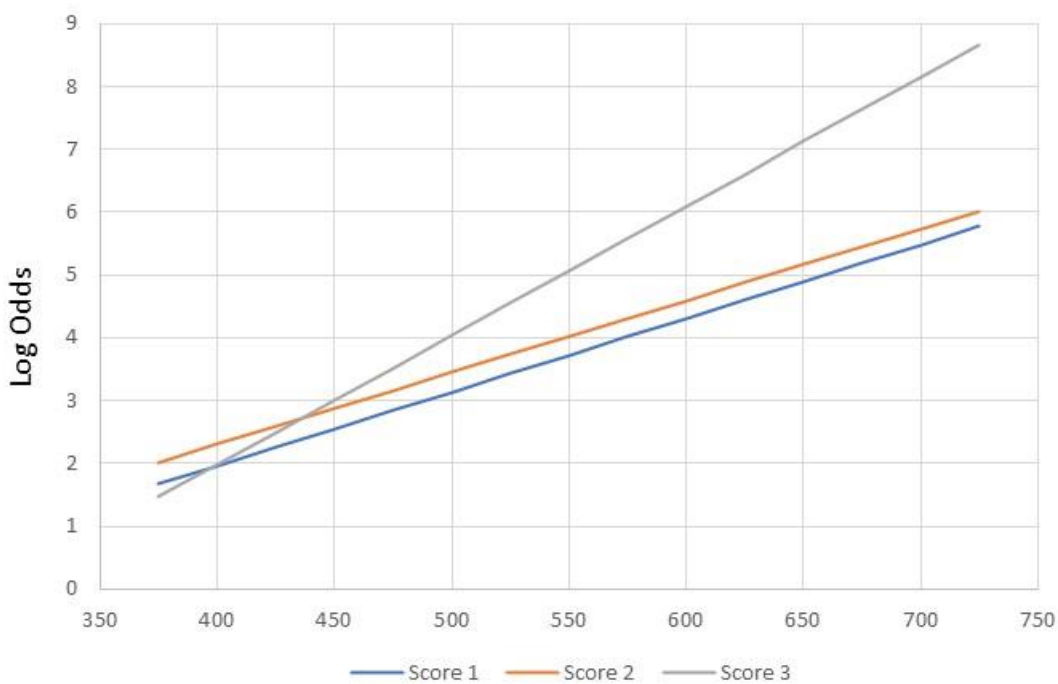
	Score 1	Score 2	Score 3
KS	0.501	0.496	0.383
Gini	0.647	0.642	0.512
ROC	0.823	0.821	0.756

that the model can distinguish between default and nondefault loans but Score 3's discriminatory power is clearly much weaker than Score 1 and 2. Note that Score 3's Gini coefficient is comparable to the VantageScore model's Gini coefficient of 52.3 on unscorable accounts. We can now observe that although Score 3 may exceed industry standard performance of 45 for the Gini coefficient, the model does generate more Type 1 and Type 2 error than Score 1 and 2 for traditional borrower segments. Hence, using Score 3 to generate more loans will come at the expense of trading off these two errors.

This effect will be made even clearer from a business perspective in the profitability analysis discussed below.

The origination period 1999-2004 was used as a baseline from which to make comparisons by score over time. To see how each score aligns to credit performance for the 1999-2004 period before applying the alignment parameters described earlier,

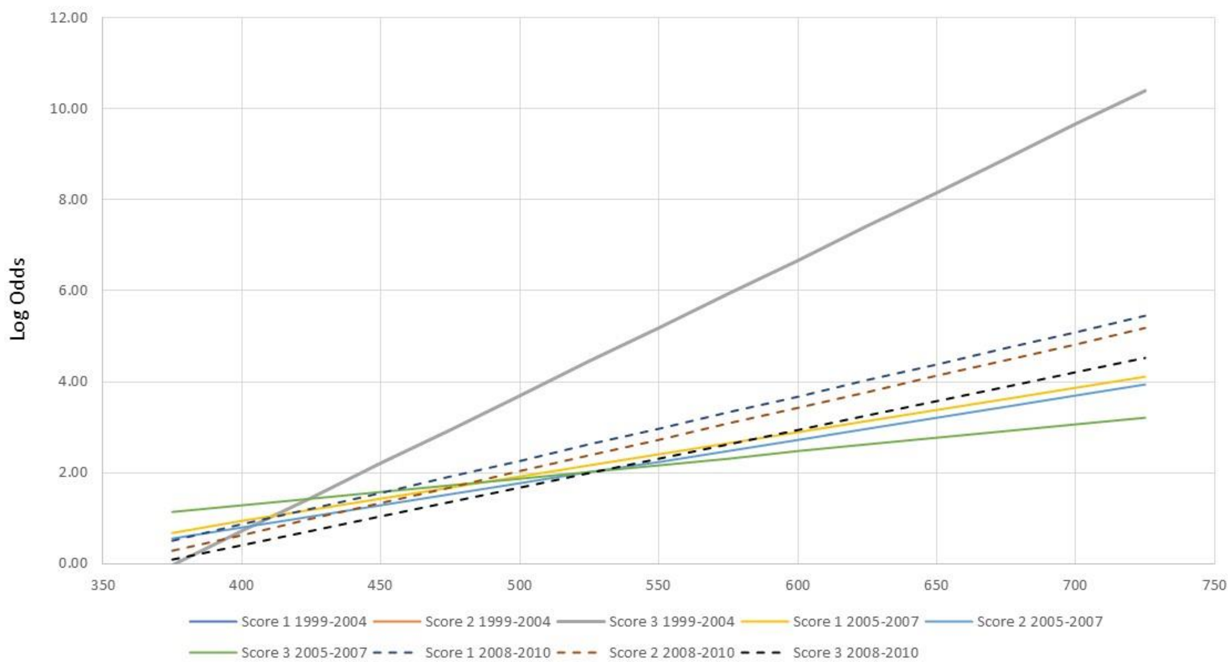
Figure 2: Pre-aligned Scores by Log Odds, 1999-2004 Originations



consider Figure 2. One immediate takeaway from Figure 2 is that all three scores rank order risk during this time period. That is, log odds (log of the ratio of good loans to bad loans in each score bucket) rises in a linear fashion with mortgage score. Scores 1 (Score Provider 1 for <=43% DTIs) and 2 (Score Provider 2 for <=43% DTIs) are much more closely aligned which is not surprising given that both were estimated on the same sample with comparable variables but a slightly different specification. Score 3 (Score Provider 2 >43% DTIs) shows a considerable difference in the score to performance relationship from either Score 1 or 2, owing in large part to the different loan sample used in the estimation. Another clear takeaway from Figure 2 is that despite all three scores falling within the same score range, the relative credit risk differs across them.

Scores from one model can be calibrated or aligned to those of another model using a variety of techniques even if as in this case they happen to already be in the same score range.[15] Applying the steps for score alignment described in the methodology section, an interesting set of results emerges for the three scores across the three origination periods. This is shown in Figure 3.

Figure 3: Log Odds by Score (Aligned for 1999-2004)



Due to the intentional alignment for Scores 1, 2, and 3 for 1999-2004, all three scores fall exactly on top of each other in Figure 3 for that period which is reflected by the steepest of the lines in Figure 3. This result is consistent with the overall performance differences between origination periods which indicated that the 1999-2004 period exhibited generally better credit performance than the other two periods.[16] Note that even after applying the 1999-2004 alignment parameters to Scores 1-3 the scores begin to diverge somewhat in 2005-2007 and 2008-2010. The three scores in the 2005-2007 period flatten significantly compared with 1999-2004 due to the higher default rates of those origination years. Scores 1 and 2 show close alignment still but Score 3 flattens even more than the other two scores for 2005-2007. An implication for this result is that risk managers need to be careful in leveraging scores from other periods where alignment to a specific period has been made but has not been reassessed over time. By 2008-2010, all three scores move upward from the 2005-2007 period as defaults generally decline but still reflect a heavy influence of the first half of 2008 when default rates were still very high. During the 2008-2010 period, Score 3's performance to score relationship diverges from both Scores 1 and 2. Other than for the 1999-2004 period when all three scores were designed to be fully aligned, these scores are not interchangeable in credit risk management applications even though they are represented to be in the same score range. Score 3's performance divergence from Scores 1 and 2 reflects the difference in sample performance of the >43% DTI borrowers and the weaker performance of that model. These differences become even more apparent in terms of their implications for risk management by examining the profitability analysis.

The results of that analysis are shown in Table 4. A separate sample of <=43% DTI loans (72,160) originated between 1999-2015 were drawn for Scores 1 and 2 and the same size sample for >43% DTI loans was drawn for Score 3 over the same period in order to maintain comparability for the analysis across scores. The number of loans for each sample are shown in Table 4 along with the percentages of defaulted and nondefaulted loans in each score bin (as well as their counts). The analysis then proceeds to calculate the credit loss to all loans above the credit policy cutoff for a particular score range. If, for example, the score range is 600-650, then the endpoint of 650 would be the cutoff score. Any loans above a score of 650 would be accepted. A tally of the percent of nondefault and default loans made above each cutoff associated with each score bin is displayed. The last two columns of Table 4 show the total cost of making a bad loan (Type 2 error) and not making an otherwise good loan (Type 1 error) using the loan level estimates for each of these two errors described in the methodology

Table 4: Profitability Analysis by Score (All Origination Periods Combined)

Score 1												Above Score Cutoff	
Total Loans	72160	% of All Loans			Good Loans	Bad Loans	Good Loans Made	% of Good Loans	Bad Loans Made	% of Bad Loans	Profit on Goods	Loss on Bads	Net Profit
Score Range	# Loans	Over Cutoff	0	1				Over Cutoff		Over Cutoff			
<=400	748	100%	0.7326	0.2674	548	200	69194	100%	2966	100%	$ 197,342,925	$ 185,351,500	$ 11,991,425
400-450	3087	99%	0.7969	0.2031	2460	627	68646	99%	2766	93%	$ 195,780,068	$ 172,850,550	$ 22,929,518
450-500	6339	95%	0.8755	0.1245	5550	789	66186	96%	2139	72%	$ 188,764,042	$ 133,664,944	$ 55,099,098
500-550	9193	86%	0.9333	0.0667	8580	613	60637	88%	1349	46%	$ 172,935,982	$ 84,339,600	$ 88,596,382
550-600	11037	73%	0.9677	0.0323	10681	356	52057	75%	736	25%	$ 148,466,246	$ 46,016,281	**$ 102,449,965**
600-650	13351	58%	0.9836	0.0164	13132	219	41376	60%	380	13%	$ 118,005,359	$ 23,735,338	$ 94,270,021
650-700	14480	39%	0.9914	0.0086	14355	125	28244	41%	161	5%	$ 80,552,663	$ 10,050,563	$ 70,502,101
700-750	9796	19%	0.9967	0.0033	9764	32	13889	20%	36	1%	$ 39,610,740	$ 2,267,563	$ 37,343,177
750-800	3664	6%	0.9992	0.0008	3661	3	4125	6%	4	0%	$ 11,764,664	$ 247,138	$ 11,517,527
>800	465	1%	0.9978	0.0022	464	1	464	1%	1	0%	$ 1,323,266	$ 63,938	$ 1,259,329
	72160				69194	2966							

Score 2													
Total Loans	72160	% of All Loans			Good Loans	Bad Loans	Good Loans Made	% of Good Loans	Bad Loans Made	% of Bad Loans	Profit on Goods	Loss on Bads	Net Profit
Score Range	# Loans	Over Cutoff	0	1				Over Cutoff		Over Cutoff			
<=400	854	100%	0.7354	0.2646	628	226	69196	100%	2964	100%	$ 197,346,716	$ 185,268,438	$ 12,078,278
400-450	3169	99%	0.7993	0.2007	2533	636	68568	99%	2738	92%	$ 195,555,564	$ 171,145,413	$ 24,410,152
450-500	6242	94%	0.8794	0.1206	5489	753	66033	95%	2104	71%	$ 188,327,689	$ 131,477,331	$ 56,850,358
500-550	9672	86%	0.9362	0.0638	9055	617	60544	87%	1351	46%	$ 172,672,404	$ 84,428,256	$ 88,244,148
550-600	12078	72%	0.9675	0.0325	11685	393	51489	74%	734	25%	$ 146,847,680	$ 45,861,156	**$ 100,986,524**
600-650	13471	56%	0.9853	0.0147	13273	198	39804	58%	341	12%	$ 113,520,638	$ 21,327,719	$ 92,192,920
650-700	14246	37%	0.9926	0.0074	14141	105	26531	38%	143	5%	$ 75,666,002	$ 8,951,237	$ 66,714,764
700-750	9816	17%	0.9966	0.0034	9783	33	12390	18%	38	1%	$ 35,336,953	$ 2,362,463	$ 32,974,491
750-800	2580	4%	0.9988	0.0012	2577	3	2608	4%	4	0%	$ 7,436,825	$ 276,563	$ 7,160,263
>800	32	0%	1	0	32	0	31	0%	1	0%	$ 87,474	$ 83,063	$ 4,411
	72160				69196	2964							

Score 3													
Total Loans	72160	% of All Loans			Good Loans	Bad Loans	Good Loans Made	% of Good Loans	Bad Loans Made	% of Bad Loans	Profit on Goods	Loss on Bads	Net Profit
Score Range	# Loans	Over Cutoff	0	1				Over Cutoff		Over Cutoff			
<=400	6144	100%	0.6996	0.3004	4298	1846	64638	100%	7522	100%	$ 178,961,418	$ 1,192,522,802	$ (1,013,561,383)
400-450	13029	91%	0.8112	0.1888	10569	2460	60340	93%	5676	75%	$ 167,060,581	$ 899,915,464	$ (732,854,884)
450-500	16672	73%	0.8852	0.1148	14758	1914	49770	77%	3216	43%	$ 137,797,076	$ 509,921,454	$ (372,124,378)
500-550	16066	50%	0.9427	0.0573	15145	921	35012	54%	1302	17%	$ 96,935,674	$ 206,482,532	$ (109,546,858)
550-600	12475	28%	0.9721	0.0279	12127	348	19866	31%	382	5%	$ 55,003,134	$ 60,537,325	**$ (5,534,191)**
600-650	5728	11%	0.9954	0.0046	5702	26	7739	12%	34	0%	$ 21,426,659	$ 5,356,791	$ 16,069,869
650-700	1771	3%	0.9958	0.0042	1764	7	2037	3%	7	0%	$ 5,639,346	$ 1,179,199	$ 4,460,147
700-750	273	0%	1	0	273	0	273	0%	0	0%	$ 756,732	$ -	$ 756,732
750-800	0	0%	0	1	0	0	0	0%	0	0%	$ -	$ -	$ -
>800	0	0%	0	1	0	0	0	0%	0	0%	$ -	$ -	$ -
	72160				64638	7522							

section. With this information, business and risk management analysts can identify the optimal cutoff where profitability is maximized by score.

Looking at Score 1, applying a credit policy cutoff of 600 maximizes profitability where the firm realizes $102.4 million in profit. This profitability reflects the lifetime return associated with all good loans made net of the costs of Type 1 and 2 errors. In this case, 75% of the good loans are made and by contrast 25% of bad loans are accepted. It is worth highlighting that all score bins are profitable.

Not surprising, Score 2 shows similar results to Score 1. Score 2's profitability is maximized at the same 600 policy cutoff with total profit of about $101 million. And as with Score 1, all score ranges are profitable. These results reflect the fact that the samples used in the analysis are the same and the models are very similar in structure and performance.

Score 3 presents a very different picture of profitability than either Scores 1 or 2. Score 3 is used against a sample of loans of the same size as the samples for Score 1 and 2 but feature only loans with DTIs >43%. If the firm were to apply the 600 credit policy cutoff used in the Score 1 and 2 profitability analysis, it would not maximize profitability for these loans. And profitability would be negative at -$5.5 million. For this sample, profit would be maximized at a cutoff of 650 when profits would be about $16 million, far less than the optimal profit for Score 1 or 2. For both cutoffs, the firm would have difficulty creating a viable product as less than a third of good loans would be made at a cutoff of 600 and it gets worse at 12% pull through for a cutoff of 650.

An important takeaway from this analysis is that scores (i.e., Score 3 in this analysis) designed to expand a particular segment of the market may pose greater credit risk ($60.5 million at a cutoff of 600) than scores used for borrower segments where performance history is more extensive for the same cutoff (Scores 1 and 2 Type 2 costs of $45-46 million). Issues in building scores that reduce Type 1 and Type 2 errors contribute to this outcome that reduce profitability and product effectiveness for borrowers. This latter point is also important as the industry would not find relatively high reject rates to be a viable offering. These results have important implications for credit investors and policymakers as it relates to the potential for multiple credit scores to be available in mortgage lending activities.

Conclusions and Summary Observations

Expanding access to credit for millions of consumers with limited or no credit experience has been a long sought-after goal among policymakers and consumer advocates for years. Doing so in a manner that balances market access with effective credit risk management is critical for credit investors and consumers alike. The foundation of effective credit risk management is sound underwriting practices, which for most consumer lending products is based on statistically based automated underwriting scoring systems. These scorecards rely heavily on borrower credit history either leveraging detailed credit attributes and/or a credit score such as FICO or VantageScore, along with other noncredit-related attributes. The reliability and credibility of the statistical properties of these scores over time have allowed them to become a key measure of a borrower's creditworthiness and thus have found widespread use among credit providers, insurers, and other market segments as a result. Those scores, however, are only as effective as the credit data on which they are built. FICO, the dominant provider of credit scores over the years has relied on minimum requirements for generating a credit score. The rationale for that has been that without at least 6 months of credit history, the statistical reliability of credit scores is diminished. In an automated underwriting environment when credit decisions are made in seconds, credit investors and their lender partners must have confidence in the statistical accuracy of their scores.

VantageScore's announcement that their latest models could effectively score a large segment of the consumer market that had been viewed as unscorable due to limited or no credit history, has created the potential to disrupt the way credit scores are developed and used across consumer lending markets. In particular, the potential to expand access to this segment of the market has rekindled the multiple credit score discussion within the mortgage industry as it relates to the two largest mortgage credit investors on the planet, Fannie Mae and Freddie Mac. The FHFA's consideration of four different options for using credit scores in both GSEs' AUS processes brings this issue into greater focus as it has significant implications on consumers and credit investors alike.

This analysis sheds some empirical light on the implications for credit risk management and consumer access to credit. Using a large sample of actual GSE mortgage loans the analysis sought to compare the effects of three different mortgage credit scores on credit risk, credit investor profitability and impact on borrower access to credit. The approach assumed a market with two mortgage score providers; one provider had extensive experience in credit scoring and developed Score 1 on a traditional segment of the market, i.e., borrowers with DTIs <=43% and another provider that also developed a similar score (Score 2) for that same borrower segment but also developed a new score (Score 3) for a segment of the mortgage market that traditionally been less represented (>43% DTI loans).

The analysis demonstrated that although Score 3 met minimum requirements for industry model performance (e.g., Gini coefficient >45), that performance was substantially less that either Score 1 or 2. This result is consistent with VantageScore's claim that their VantageScore model's performance exceeded industry benchmarks for performance with a Gini coefficient of 52.3 on unscorable accounts. Score 3 was comparable in performance with a Gini coefficient of 51.2. By comparison, Score 1 and 2 had significantly higher Gini coefficients of 64.7 and 64.2, respectively, though these scores were based on the traditional borrower segment. These results show that while a score may be minimally viable, it exposes credit investors to risk from higher Type 1 and Type 2 errors that result from a model's greater difficulty in discerning between good and bad loans with less performance history. Using >43% DTI loans as the proxy for consumers with less credit history turns out to be a less conservative segment for study from a credit history perspective than unscorable consumers since >43% DTI loans are represented in the GSE historical record. In reality, developing statistically reliable scoring models of unscorables is more difficult given the paucity of data for these consumers and potentially poses more Type 1 and 2 error to credit investors.

Credit investors and policymakers should therefore exercise great caution in assuming the interchangeability of credit scores as a mechanism to expand access to credit. The problem is that on the surface, multiple scores, even if presented on the same score continuum are not the same from a risk perspective. That was clearly and unequivocally demonstrated in this analysis showing that while all 3 scores were aligned for the benchmark period of 1999-2004, scores over time began diverging as economic conditions changed, with differing levels of credit risk for the same score.

These differences in scores, despite being placed in the same score range, became even more apparent in the profitability analysis showing that the same credit policy cutoff used in optimizing profitability for Score 1 and 2 would destroy shareholder value using Score 3 while having limited value in actually increasing access to credit for the stretch borrower segment.

The implications for credit risk managers and policymakers are the following:

1/ Credit risk may differ significantly across scores from different providers despite being depicted in the same score range.

2/ Alignment of credit risk across different credit scores is possible at a point in time but can deteriorate over time and with varying economic conditions.

3/ Scores developed on limited performance history, while perhaps fulfilling minimal model performance requirements, may not result in a significant uptake in acceptable loans based on model error and credit risk considerations.

4/ Financial performance can deviate widely by score. Profitability of mortgage borrowers with much less performance history (Score 3 borrowers) than typical borrowers (Score 1 and 2 borrowers) is much worse (negative) for the same policy score cutoff used for Score 1 and 2 samples where profitability is positive and optimized.

Reliance on scores that are highly effective in distinguishing between good and bad loans (not just minimally viable) is the best way of ensuring credit risk is managed prudently while expanding access to credit. Also, when using multiple scores, there are other drawbacks such as operational challenges in business and risk systems, and the potential to dilute predictive performance. To mitigate third-party and credit risk, the analysis suggests that credit investors need to perform their own analysis of alternative scores and confirm the claims of their providers to expand credit while managing credit risk.

Technical Appendix

The three mortgage scores estimated in this analysis were developed from the following GSE loan level samples:

Table TA1: Mortgage Score Summary Details

Mortgage Score	Proxy of Scorable/Unscorable Accounts	Proxy Definition	Sample Period
Score 1	Traditional Mortgage Borrowers	<=43% DTIs	1999-2015
Score 2	Traditional Mortgage Borrowers	<=43% DTIs	1999-2015
Score 3	Nontraditional Mortgage Borrowers	>43% DTIs	1999-2015

Summary statistics for each sample are found in Tables TA2 A and B below.

Table TA2A: Summary Statistics for <=43% DTI Sample (144,458 observations)

Variable	Mean	Std Dev	Minimum	Maximum
Mortgage Rate at Origination	5.277	1.260	2.875	8.375
Original UPB ($)	188,314.780	107,792.040	37,000.000	550,000.000
Original Combined LTV	70.772	17.503	22.000	97.000
DTI	29.638	8.742	9.000	43.000
Credit Score	746.282	52.845	595.000	816.000
Investor-owned (=1)	0.065	0.247	0.000	1.000
Primary (=1)	0.893	0.309	0.000	1.000
Cashout Refi (=1)	0.302	0.459	0.000	1.000
Purchase (=1)	0.358	0.479	0.000	1.000
Correspondent Channel (=1)	0.358	0.479	0.000	1.000
Broker Channel (=1)	0.146	0.353	0.000	1.000
Condo (=1)	0.079	0.269	0.000	1.000
Single Family (=1)	0.737	0.440	0.000	1.000
Coop (=1)	0.006	0.076	0.000	1.000
Manufactured Housing (=1)	0.006	0.079	0.000	1.000
>= 2 Borrowers (=1)	0.598	0.490	0.000	1.000
>2 Units (=1)	0.022	0.146	0.000	1.000
Estimated Ever D90+	0.042	0.059	0.001	0.739

Table TA2B: Summary Statistics for >43% DTI Sample (38,491 observations)

Variable	Mean	Std Dev	Minimum	Maximum
Mortgage Rate at Origination	5.801	1.085	2.875	8.375
Original UPB ($)	194,641.630	103,881.510	37,000.000	550,000.000
Original Combined LTV	72.197	16.787	3.000	134.000
DTI	50.096	5.574	44.000	64.000
Credit Score	726.545	55.208	482.000	830.000
Investor-owned (=1)	0.087	0.282	0.000	1.000
Primary (=1)	0.869	0.338	0.000	1.000
Cashout Refi (=1)	0.364	0.481	0.000	1.000
Purchase (=1)	0.394	0.489	0.000	1.000
Correspondent Channel (=1)	0.384	0.486	0.000	1.000
Broker Channel (=1)	0.179	0.384	0.000	1.000
Condo (=1)	0.087	0.281	0.000	1.000
Single Family (=1)	0.748	0.434	0.000	1.000
Coop (=1)	0.004	0.063	0.000	1.000
Manufactured Housing (=1)	0.008	0.086	0.000	1.000
>= 2 Borrowers (=1)	0.504	0.500	0.000	1.000
>2 Units (=1)	0.037	0.188	0.000	1.000
Estimated Ever D90+	0.114	0.098	0.002	0.616

Each model was estimated using a standard credit scoring regression methodology, logistic regression, a form on binary choice dependent variable specification. The dependent variable was specified as 1 if a loan had ever reached 90 days past due or worse in its performance experience and 0 otherwise. This definition of an event is consistent with mortgage industry credit scoring models.

Both categorical and continuous variables were used in specifying each model. Continuous variables included borrower credit score, LTV, DTI and UPB. In order to capture any nonlinearity between Ever D90+ and the risk factor, these continuous variables were set up as spline effects in the model specification. The variables were tested to ensure monotonicity in default.

A variety of model specifications were tested for all three mortgage scoring models. The estimated coefficients and statistical significance for the models are shown in Table TA3A-C. All coefficients conform to prior expectations with default (e.g., negative signs for credit score effects). In addition, all coefficients are statistically significant. Figures TA1A-C display the receiver operating characteristic curve (ROC) results for each mortgage score based on the validation sample.

Table TA3A: Mortgage Score 1 Model Results

Mortgage Score 1

Parameter	DF	Estimate	Standard Error	Wald Ch-Square	Pr>ChiSq
Intercept	1	7.1642	0.284	636.392	<.0001
Credit Score Base Spline	1	-0.0155	0.00041	1424.6624	<.0001
Credit Score Spline 720 Knotpoint	1	-0.00532	0.001	28.2369	<.0001
Original Combined LTV Spline 45% Knotpoint	1	0.0295	0.0012	603.1989	<.0001
DTI Spline 20% Knotpoint	1	0.0221	0.00193	131.6412	<.0001
Broker Channel	1	0.3742	0.0373	100.4714	<.0001
Correspondent Channel	1	0.2854	0.0309	85.5581	<.0001
Manufactured Housing Home	1	0.9893	0.1221	65.6617	<.0001
Single Family Home	1	0.1683	0.0356	22.3795	<.0001
Cashout Refinance	1	0.522	0.0353	218.1475	<.0001
Purchase	1	-0.2392	0.0389	37.7434	<.0001
Primary Owner	1	-0.3951	0.0463	72.9437	<.0001
2 or more Borrowers	1	-0.5427	0.0279	378.0749	<.0001

Table TA3B: Mortgage Score 2 Model Results

Mortgage Score 2

Parameter	DF	Estimate	Standard Error	Wald Ch-Square	Pr>ChiSq
Intercept	1	7.5896	0.2283	1104.92	<.0001
Credit Score Base Spline	1	-0.0161	0.000329	2391.7096	<.0001
Credit Score Spline 750 Knotpoint	1	-0.0083	0.00149	31.1089	<.0001
Original Combined LTV Spline 65% Knotpoint	1	0.035	0.00152	527.919	<.0001
DTI Spline 28% Knotpoint	1	0.0295	0.0026	129.1186	<.0001
Broker Channel	1	0.3794	0.0373	103.4487	<.0001
Correspondent Channel	1	0.2894	0.0308	88.1036	<.0001
Manufactured Housing Home	1	1.014	0.1221	68.987	<.0001
Single Family Home	1	0.1737	0.0354	24.0798	<.0001
Cashout Refinance	1	0.6341	0.0316	403.3464	<.0001
Investor-owned	1	0.4644	0.0535	75.4168	<.0001
2 or more Borrowers	1	-0.5256	0.0278	356.8705	<.0001

Table TA3C: Mortgage Score 3 Model Results

Mortgage Score 3

Parameter	DF	Estimate	Standard Error	Wald Ch-Square	Pr>ChiSq
Intercept	1	1.0787	0.4582	5.5425	0.0186
Credit Score Base Spline	1	-0.012	0.000545	487.3955	<.0001
Credit Score Spline 720 Knotpoint	1	-0.0031	0.00133	5.3913	0.0202
Original Combined LTV	1	0.0336	0.00145	535.5562	<.0001
DTI Baseline Spline	1	0.0621	0.00527	138.5603	<.0001
DTI Spline 55% Knotpoint	1	-0.0551	0.0137	16.0996	<.0001
Single Family Home	1	-0.1571	0.0421	13.9336	0.0002
Cashout Refinance	1	0.3897	0.0466	70.03	<.0001
Purchase	1	-0.3582	0.0501	51.0755	<.0001
Investor-owned	1	0.2461	0.0636	14.952	0.0001
2 or more Borrowers	1	-0.3988	0.0356	125.6761	<.0001

Figure TA1A: Mortgage Score 1 ROC Results

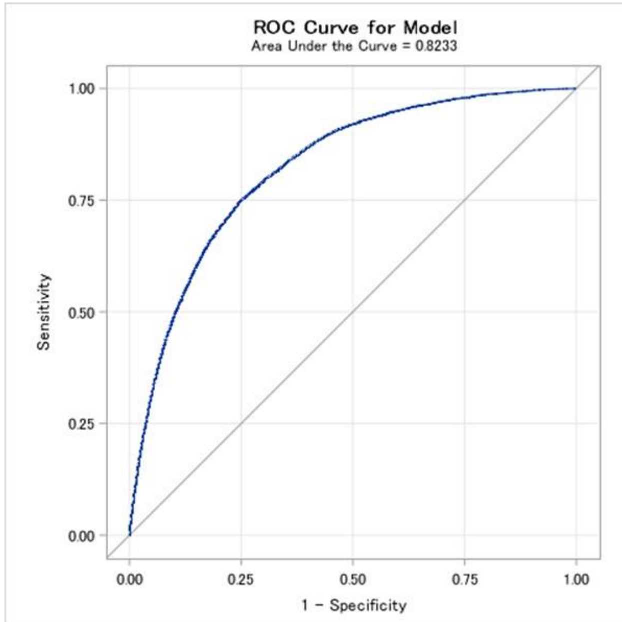


Figure TA1B: Mortgage Score 2 ROC Results

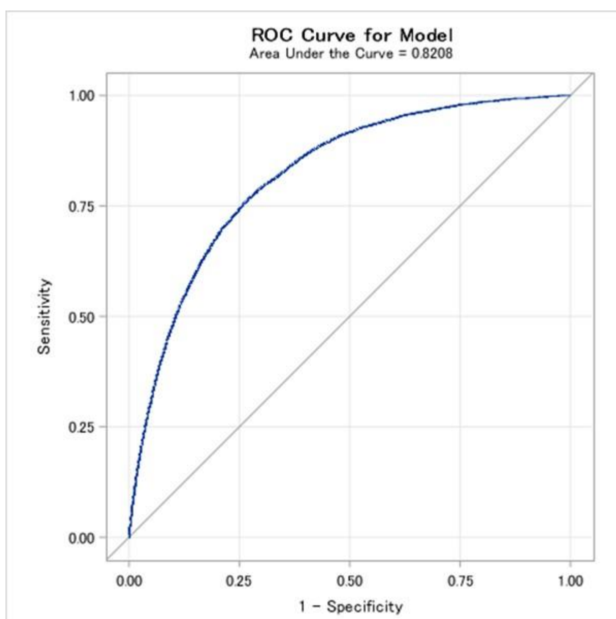
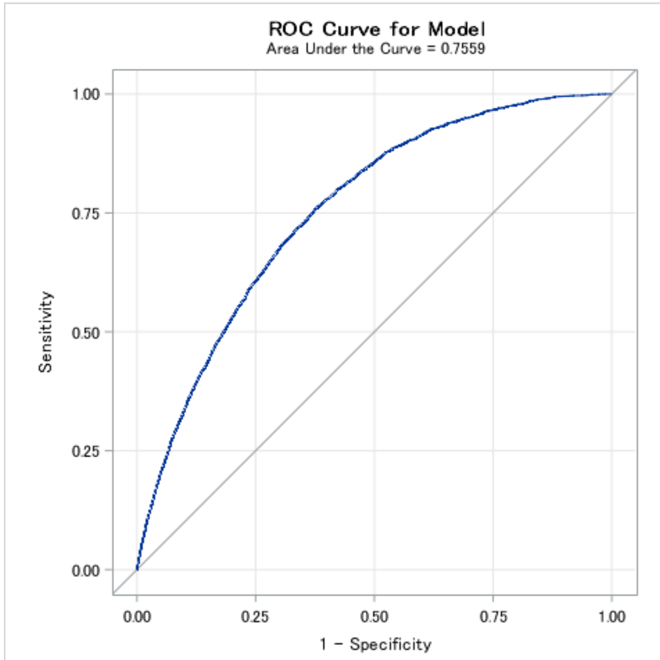


Figure TA1C: Mortgage Score 3 ROC Results



The alignment and associated regressions parameters for Scores 1-3 in 1999-2004 are found below in Table TA4

Table TA4: Score Alignment Parameters

Scoring Model	Alignment Parameters		1999-2004 Log Odds to Score Parameters	
	b	a	b	a
Score 1	1	0	0.012	-2.730
Score 2	0.975	39.579	0.011	-2.266
Score 3	1.753	-299.265	0.021	-6.238

References

1. Consumer Financial Protection Bureau, Office of Research, Data Point: Credit Invisibles, May 2015.
2. Federal Housing Finance Agency, 12 CFR Part 1254, Validation and Approval of Credit Score Models, Final Rule, 2019.
3. Federal Housing Finance Agency, Public Listening Session on Credit Score Models, March 1, 2022, https://www.fhfa.gov/Videos/Pages/Public-Listening-Session-on-Credit-Score-Models.aspx.
4. Federal Housing Finance Agency, FHFA Needs Your Feedback on Credit Score Requirements, blog, January 18, 2018, https://www.fhfa.gov/Media/Blog/Pages/FHFA-Needs-Your-Feedback-on-Credit-Score-Requirements.aspx.
5. Goodman Laurie S. and Jun Zhu, The GSE Reform Debate: How Much Capital is Enough? Urban Institute, Housing Finance Policy Center, October 24, 2013.
6. FICO, To Score Or Not To Score, White Paper, 2018, https://www.fico.com/sites/default/files/inline-files/FICO_70_Insights_To_Score_or_Not_To_Score_3009WP_US_EN.pdf.
7. Kaul, Karan, Adopting Alternative Data in Credit Scoring Would Allow Millions of Consumers to Access Credit, Urban Institute, March 15, 2021.
8. Parrent, Tom and George Haman, Risks and Opportunities in Expanding Mortgage Credit Availability Through New Credit Scores, Quantilytic, LLC, December 2017.
9. Sepp, Pete and Thomas Aiello, Risky Road: Assessing the Costs of Alternative Credit Scoring, National Taxpayers Union, March 21, 2019.
10. Thompson, Sandra L., Federal Housing Finance Agency, Statement of Sandra L. Thompson, FHFA Director, Before the House Committee on Financial Services "Housing in America: Oversight of the Federal Housing Finance Agency", 7/20/2022.
11. VantageScore, Are credit scores assigned to conventionally unscoreable consumers considered reliable?, June 24, 2020, https://vantagescore.com/newsletter/are-credit-scores-assigned-to-conventionally-unscoreable-consumers-considered-reliable/.
12. VantageScore, Credit Score Basics, Part 3: Achieving the Same Risk Interpretation from Different Models with Different Score Ranges, September 2011.

Endnotes

---

[1] Consumer Financial Protection Bureau, Office of Research, Data Point: Credit Invisibles, May 2015.

[2] FICO, To Score Or Not To Score, White Paper, 2018, https://www.fico.com/sites/default/files/inline-files/FICO_70_Insights_To_Score_or_Not_To_Score_3009WP_US_EN.pdf. Karan Kaul, Adopting Alternative Data in Credit Scoring Would Allow Millions of Consumers to Access Credit, Urban Institute, March 15, 2021.

[3] VantageScore, Are credit scores assigned to conventionally unscoreable consumers considered reliable?, June 24, 2020, https://vantagescore.com/newsletter/are-credit-scores-assigned-to-conventionally-unscoreable-consumers-considered-reliable/.

[4] VantageScore uses the Gini coefficient to measure model performance citing an estimate of 45 as the minimum industry standard of performance. They claim that the Gini coefficient on a validation sample of newly scored consumers is 52.3.

[5] FHFA, FHFA Needs Your Feedback on Credit Score Requirements, blog, January 18, 2018, https://www.fhfa.gov/Media/Blog/Pages/FHFA-Needs-Your-Feedback-on-Credit-Score-Requirements.aspx.

[6] Federal Housing Finance Agency, 12 CFR Part 1254, Validation and Approval of Credit Score Models, Final Rule, 2019.

[7] A mortgage score is a statistically based model that incorporates a number of borrower, property, loan and other risk attributes in addition to credit score in predicting mortgage default or delinquency.

[8] Borrowers with DTIs >43% represented between 17.4% (1999-2004) and 25.7% (2005-2007) of all mortgage purchased by Fannie Mae (according to their Data Dynamics tool) for the origination period used in this study of 1999-2015. One could argue that this definition of a stretch borrower is a far less restrictive proxy of credit unscorables with the implication that the results from the analysis of Score 3 may present a more favorable depiction of the impact on credit risk and profitability of Score 3 compared with Scores 1 and 2 than the effects of credit scores developed from true unscorable consumer segments to traditional scoring cohorts.

[9] Logistic regression techniques are often used in credit scoring as they permit better specification of the probability of an event by restricting the probability space to the 0-1 domain, among other statistical benefits of the specification.

[10] An industry standard transformation from PD to score is to multiply the $X\beta$ from the logistic regression results by the ratio (-PTD/ln(2)) where PTD represents the points to double the odds of an event. For this analysis PTD equals 50.

[11] The first half of 2008 performed very differently from the second half of that year due to severe credit tightening as the crisis unfolded. While the crisis extended beyond 2008, loan performance of 2009 and 2010 was very good by comparison so the 2008-2010 origination period can be viewed as being affected by the abrupt change in credit policy at the onset of the mortgage crisis.

[12] Note that credit scoring outcomes are typically presented by the log of the odds of some event, in this case Ever D90+. The odds of default are not a probability but rather

defined to be the ratio of the number of nonevents to events. Taking the log of the odds is a mathematical transformation to present the results in a more symmetrical way for depicting event and nonevent outcomes.

[13] Laurie S. Goodman and Jun Zhu, The GSE Reform Debate: How Much Capital is Enough? Urban Institute, Housing Finance Policy Center, October 24, 2013.

[14] Loan counts from the 2011-2015 origination period were excluded from Table 1 due to the 10 year loan seasoning requirement. Those loans were used, however, in estimating models for Score 1-3.

[15] VantageScore, Credit Score Basics, Part 3: Achieving the Same Risk Interpretation from Different Models with Different Score Ranges, September 2011.

[16] Again, subject to the earlier comment that the 2009-2010 origination years performed comparably if not better (for 2010) to the 1999-2004 period. The first half of 2008 tends to skew the overall period's results. These comments are corroborated by data provided in Fannie Mae's Data Dynamics tool.
https://datadynamics.fanniemae.com/data-dynamics/#/report/3