FICO® SCORE

Scoring Solutions

FICO® Score 10 T Materially Outperforms VantageScore 4.0

White Paper

Executive Summary

FICO® Score 10 T Decisively Beats VantageScore 4.0 on Predictability

An analysis by FICO data scientists has found that FICO® Score 10 T significantly outperforms VantageScore 4.0 in mortgage origination predictive power. An independent third-party study found that VantageScore 4.0 generates a minimal predictive improvement over Classic FICO—so minor that, when properly adjusting the comparison due to a truncation anomaly described below, it should be questioned if VantageScore 4.0 even beats the 20-year-old Classic FICO at all. At the same time, FICO Score 10 T far exceeds Classic FICO in detecting loan losses. In fact, FICO Score 10 T's improvement over Classic FICO was shown to be five times better than VantageScore 4.0's improvement, with FICO Score 10 T detecting 18% more defaulters in the critical score decile commonly used for mortgage originations versus just 3.4% for VantageScore 4.0.

Truncated Data: VantageScore's Illusion of Performance

Importantly, both Classic FICO and VantageScore 4.0 have been compared on conforming mortgage data, which truncates the performance calculation of Classic FICO below the 620 cut off. While the performance of Classic FICO should be measured through its entire score range from 300 to 850, the historical database only allows for measuring the strength of Classic FICO from 620 and 850. This truncation effect negatively impacts the calculation of Classic FICO's strength by 15% or more, artificially boosting VantageScore 4.0's performance in comparison—a boost big enough that it calls into serious question whether VantageScore now actually beats Classic FICO at all.

VantageScore 4.0 Punishes Non-Homeowners—FICO® Score 10 T Does Not

FICO® Score 10 T handily wins despite VantageScore's attempt to improve the performance of VantageScore 4.0 by including mortgage-specific variables in its model. But, including mortgage-specific variables penalizes people who have never owned a home. Under VantageScore millions of Americans—including young people, members of the military, and people from disadvantaged groups—will have lower scores than they otherwise would, merely because they have never owned a home. That's unfair. Forcing lenders to tell people from disadvantaged groups through adverse action notices that they have been rejected for a mortgage because they don't currently own a home is irresponsible. FICO Score 10 T includes rental data while not penalizing people for not owning a home.

Impact of VantageScore's Changes to VantageScore 4.0 After Submission to GSEs

Nearly two years after both FICO and VantageScore submitted their models for evaluation by the GSEs and FHFA for accuracy and reliability, VantageScore decided to modify VantageScore 4.0 to exclude a large category of predictive credit file information—Medical Debt Collection Records. VantageScore's post validation change raises serious questions of a potential breach of the integrity of the FHFA and GSE Validation and Approval of Credit Scores process that would impact the VantageScore/Classic FICO comparison cited in this white paper. Not only does it undermine determining whether VantageScore 4.0 now even outperforms Classic FICO, it raises serious doubt of whether the VantageScore 4.0 model evaluated by the GSEs and FHFA under the Joint Enterprise Credit Score Solicitation process is even the same VantageScore model used to generate scores in the historical dataset made available by the GSEs for industry modeling and transition. Most troubling, it creates concerns about whether the current new model was even approved through the required process. If model developers are allowed to discretionarily change their models after GSE validation, no constraints would remain for any model developers to refrain from altering their models to curry favor with stakeholders, including lenders who transfer the risk to the GSEs and taxpayers. In fact, they would be highly incented to do so. The race to the bottom is no longer theoretical and, if allowed to continue, could have significant and irreversible impacts on the mortgage ecosystem.

More loans, better pricing with FICO® Score 10 T

Based on enhanced prediction, use of FICO® Score 10 T rather than VantageScore 4.0 by market participants throughout the mortgage industry will drive significantly more loan approvals for prospective borrowers and, due to better model performance for mortgage insurers, investors and others, improve mortgage pricing and lower costs for borrowers—benefiting millions of Americans.

Given this superior performance, it's no surprise that lenders representing over $**300 billion in annual mortgage originations** and **$1.5 trillion in mortgage servicing portfolios** have already adopted and started trading loans with FICO® Score 10 T outside of the GSE conforming market.

The Bottom Line

In head-to-head competition, VantageScore is the loser and it's not even close. The mortgage industry deserves better than a runner-up score. FICO® Score 10 T isn't just better—it's the trusted standard. It wins on predictive accuracy. It wins on fairness. For lenders, investors, and consumers, and any stakeholder concerned about safety and soundness, FICO Score 10 T is the clear choice.

FICO® Score 10 T Materially Outperforms VantageScore 4.0

Since 2022, when the Federal Housing Finance Agency ("FHFA") announced that both FICO® Score 10 T and VantageScore 4.0 were approved for use for conforming mortgages sold to the government sponsored enterprises, Fannie Mae or Freddie Mac ("GSEs"), stakeholders may have assumed both scores' predictive power are nearly the same. They are not. An analysis by FICO data scientists reveals material differences between the predictiveness of the models. FICO Score 10 T, built using our proprietary and time-tested FICO® Score design methodology, is materially more effective at assessing likelihood of default in conforming mortgage originations than VantageScore 4.0. Notably, even when comparing VantageScore to our 20-plus year old FICO® Score models (commonly referred to in the mortgage industry as "Classic FICO"), studies have found that VantageScore barely beats (or in some score ranges is no more predictive than) Classic FICO for conforming mortgage originations. Further, this weak performance of VantageScore compared to Classic FICO is true even though VantageScore benefits from a significant performance measurement anomaly relative to Classic FICO due to a truncation effect discussed below. At the same time, FICO Score 10 T materially outperforms the Classic FICO Score for conforming mortgage originations. In this paper, we provide data and analysis that supports the conclusion that FICO Score 10 T materially outperforms VantageScore 4.0 in predictive power for conforming mortgage originations.

This paper also highlights that FICO® Scores are built to last. As the gold standard in credit risk assessment, FICO takes seriously the responsibility to help protect the safety and soundness of consumer credit industries. Our focus on ensuring strong model performance over time and protecting against model performance deterioration are but a few good examples. Many developers design their models in a way that is "overfit" to the current environment or take measures that reduce model robustness over time. Although in some cases these models can get an edge in predictive accuracy assessments closer to their development periods, problems will emerge unexpectedly over time. Because organizations and whole industries want to plan for and coordinate mission critical credit score replacement at a time of their choosing rather than when they are forced to by significant model performance degradation, ensuring the long-term performance of models over years and, even decades, becomes paramount. From Classic FICO to FICO® Score 10 T, stability and long-term performance have been key considerations built into FICO Score's proprietary design blueprint. In fact, Classic FICO has been depended on in mortgage originations for more than two decades and has proven to be highly reliable and robust predictors of credit risk through several economic cycles, including the Great Recession and the COVID-19 pandemic. That said, notable changes over decades, including newly available data in the credit file, new analytic techniques, and changes in consumer behavior, have enabled FICO to develop a new FICO Score in FICO Score 10 T that significantly outperforms Classic FICO, and, in turn, as we show below, VantageScore 4.0.

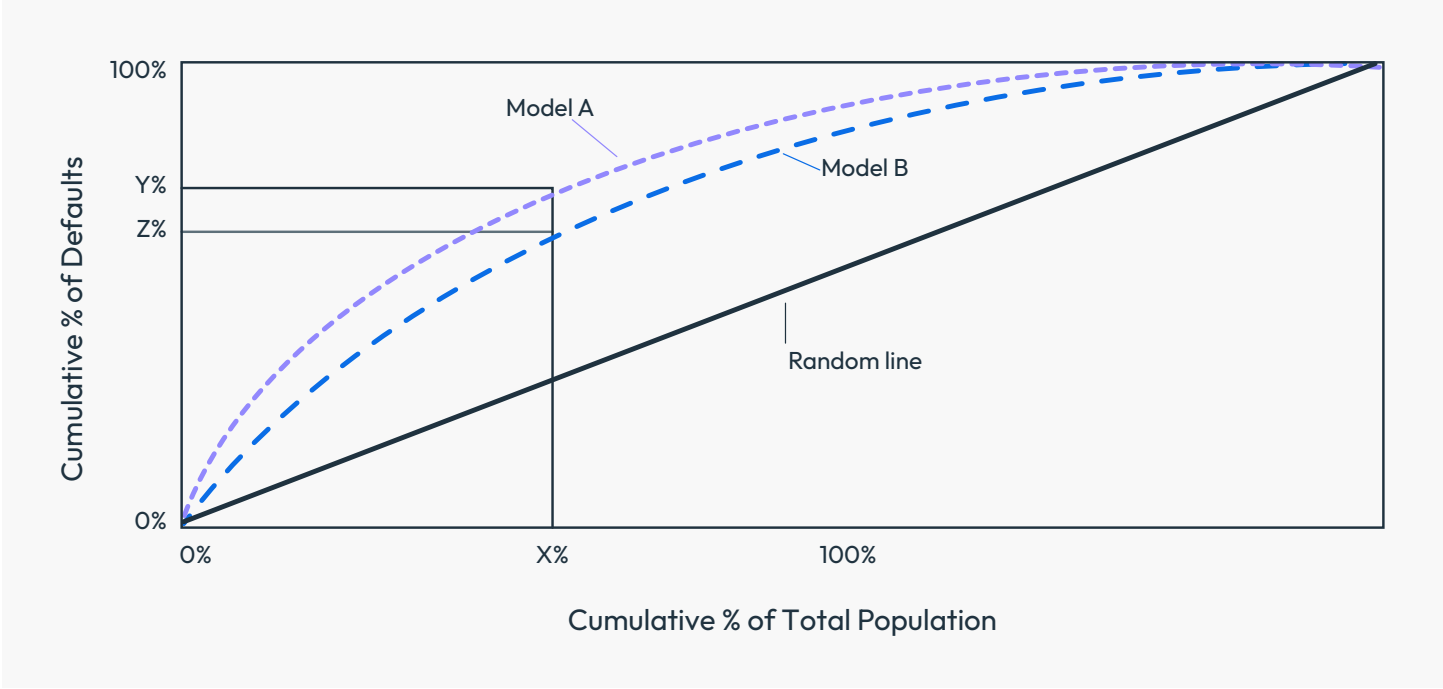
Comparing the Performance of Multiple Credit Scoring Models

How is model performance or the predictive strength of one credit score vs. another measured? Two of the most widely used measures that data scientists use to compare the rank ordering effectiveness of predictive models are the Receiver Operating Characteristic ("ROC") curve and the Kolmogorov-Smirnov ("K-S") statistic. Both measures quantify the degree of differentiation a predictive model provides between two distinct outcome groups (e.g., defaulters and non-defaulters). Appendix A contains additional details regarding these two measures and how they are calculated.

The ROC curve can be further analyzed to reveal the percentage of defaulters captured at or below a score cutoff corresponding to the lowest scoring X% of the population. As shown in Figure 1 below, Model A identifies Y% of cumulative defaulters, and Model B identifies Z% of cumulative defaulters, at the same score cutoff. Model A is, therefore, a more effective model as it identifies a greater percentage of defaulters in the lowest scoring X% of the population. To put numbers behind this example, if Model A captures 42% of cumulative defaulters and Model B captures 40% of cumulative defaulters, each at the cutoff corresponding to the same lowest scoring X% of the total population, Model A offers 5% (i.e., 42% minus 40%, divided by 40%) relative improvement in defaulters identified at that cutoff.

In the conforming mortgage space, comparison of ROC curves often focuses on the lowest scoring decile of the population, the critical score decile commonly used for mortgage originations.

Figure 1



To best interpret how meaningful differences in performance between Classic FICO, FICO® Score 10 T, and VantageScore 4.0 really are, it is important to understand how different levels of relative improvement in the defaulters captured at a particular score cutoff in a ROC curve are generally viewed in credit risk scoring. For purposes of this discussion, it is important to recognize that Classic FICO and FICO Score 10 T are different generations of credit scoring models, but FICO Score 10 T and VantageScore 4.0 are credit scoring models built in much closer time periods.

- The typical relative improvement in defaulters captured in the lowest decile from one generation of a broad-based credit scoring model (e.g., a FICO® Score model) version to a future generation model is in the *2.5–5% range*.[1]
- Observing relative improvement of *5–10%* from one broad-based credit scoring model to the next is much more uncommon, and is more typically observed only for specific subsegments of the population and/or specific use cases.[2] This high level of performance improvement from one credit score model to another can also be observed when the new model is optimized to a different outcome measure or incorporates entirely new sources of data. For example, the UltraFICO® Score, which leverages a new set of predictive information in consumer contributed checking and savings data, offers relative improvement of approximately 10% over the base FICO® Score for thin file/new-to-credit borrowers. Such a difference would mean that reductions in portfolio loss rates and/or increases in approval rates (holding loss rates fixed) would be realized by utilizing the better performing score. In mortgage originations, the implications are even more pronounced. Improved model accuracy would translate to improvements in approvals at origination, and lower pricing for borrowers, improved mortgage insurance underwriting, better mortgage insurance rates, improved prepayment model performance, lower mortgage investment pricing, or more favorable mortgage security pricing. Capital allocation requirements, collection operations, and servicing costs could all be improved with this increased model performance.
- Relative improvement of over *10%* is rare between broad-based credit scoring scores models, especially for those built in the same "generation" and for similar purposes.[3] Similar to the discussion above, such a large difference would equate to material reductions in portfolio loss rates and/or material increases in approval rates (holding loss rates fixed) realized by utilizing the better performing score. Again, in mortgage originations the implications are even more pronounced. The high degree of improved model accuracy would translate not just to substantial improvements in approvals at origination, but also substantially lower pricing for large numbers of borrowers, and significantly improved mortgage insurance underwriting, mortgage insurance rates, prepayment model performance, mortgage investment pricing, and/or more favorable mortgage security pricing. Again, such high performance improvements could even improve capital allocation requirements, collection operations, and servicing costs.

---

[1] Similarly, discussing model improvement in terms of K-S, the typical improvement from one credit scoring model to another would be in the 1–2 K-S point range in absolute terms, and also in the <5% range in terms of relative K-S improvement.

[2] Likewise, when evaluating one credit scoring model to another, improvement of 2–5 K-S points, or in the 5–10% range on a relative basis, is not common, and would also typically be seen only for specific subsegments of the population and/or specific use cases.

[3] K-S for this "rare" level of improvement from one credit scoring model to another would generally be shown in more than 5 K-S points in absolute terms, and in over 10% improvement on a relative basis.

In summary, even a relative improvement in defaulters identified of up to 5% is meaningful. Lenders will be willing to incur costs and make efforts to move to scores offering such improvements. Relative improvements in default identification of 5–10% are much less common and would represent a large efficiency opportunity for lenders who discover such new scores. Relative improvements of more than 10% represent a rare opportunity to materially improve risk management and other vital operations. In our experience, risk takers across the mortgage ecosystem would demand any score that could deliver relative performance improvements approaching or exceeding 10%.[4]

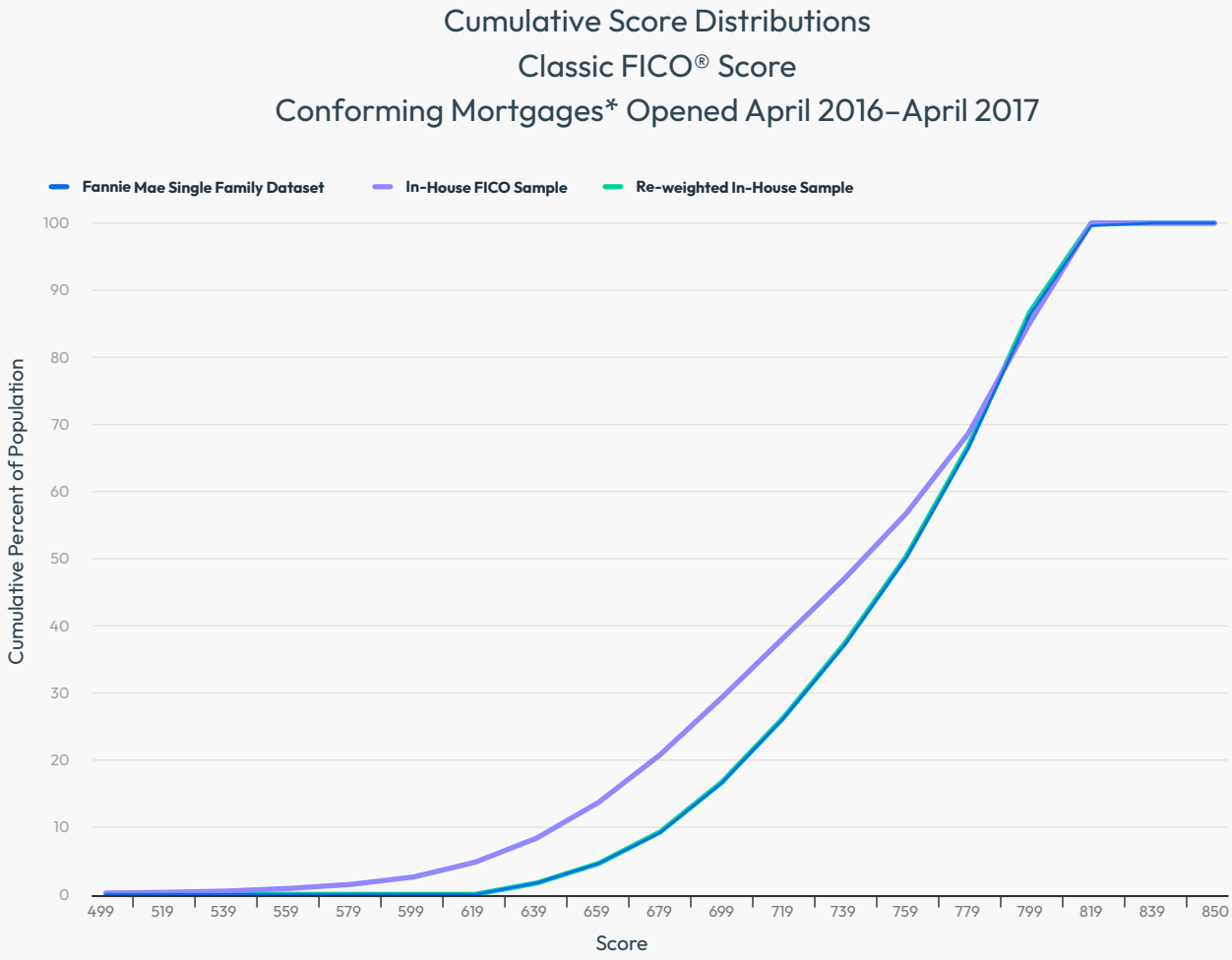
FICO Study: Comparing FICO® Score 10 T to Classic FICO on "GSE-like" Portfolios

Our study was designed to mimic what would be observed when comparing FICO® Score 10 T with Classic FICO on the loan-level historical datasets published by Fannie Mae and Freddie Mac.

To accomplish this, we started with a dataset composed of nationally representative sample of millions of credit files, and then filtered the dataset to only credit files containing newly opened conforming mortgages. We then re-weighted the records in the sample such that the Classic FICO® Score distribution of the newly opened conforming mortgages matched the Classic FICO Score distribution observed in the Fannie Mae Single Family Historical Loan Performance Dataset on newly opened mortgages, from that same time period. Figure 2 below plots the Classic FICO Score distributions of vintages of conforming mortgages from April 2016–April 2017. Note that after the re-weighting of the credit files, the resulting cumulative Classic FICO distribution from the in-house FICO data sample is virtually identical to the Classic FICO distribution observed on the Fannie Mae Single Family Historical Loan Performance Dataset, from that same time period.

[4] Similarly, risk takers across the mortgage ecosystem would, in our experience, demand any score that could deliver the "rare" relative performance improvements levels approaching or exceeding 5 K-S points in absolute terms, or 10% improvement on a relative basis.

Figure 2



*All scorable, non-jumbo mortgage originations

Once we had successfully re-weighted our sample of credit files with newly booked conforming mortgage loans from April 2016–April 2017 to mimic "GSE-like" portfolios, we then assessed the effectiveness of Classic FICO and FICO® Score 10 T on this data sample using a 24 month, 90+ days past due performance outcome. This approach is largely consistent with the study design proposed by the GSEs as part of the model accuracy assessment within the Joint Enterprise Credit Score Solicitation process (see page 10 of the Joint Enterprise Credit Score Solicitation).[5] Note that the Joint Enterprise Credit Score Solicitation explicitly calls out K-S as the model performance measure to be utilized when assessing accuracy of any models submitted under the Joint Enterprise Credit Score Solicitation (page 12 of the Joint Enterprise Credit Score Solicitation).

[5] Pursuant to the FHFA's final rule issued pursuant to Section 310 of the Economic Growth, Regulatory Relief, and Consumer Protection Act of 2018 (Pub. L. 115-174, Section 310), the GSEs published a Joint Enterprise Credit Score Solicitation on February 18, 2020 ("Joint Enterprise Credit Score Solicitation").

The results of our study are shown in Figures 3, 4, and 5 below. On a sample of mortgages opened between April 2016 and April 2017 and substantially similar to a vintage examined by the GSEs as part of the "Credit Score Models and Reports Initiative", FICO® Score 10 T exhibited substantial lift over Classic FICO:

- 18.1% relative improvement in defaulters captured in the lowest decile when comparing ROC curves (see Figure 4: 51% vs 43%)
- 8.3 K-S points improvement in absolute terms, or 18.8% relative improvement in K-S (see Figure 5: 52.7 K-S vs 44.3 K-S)

Figure 3

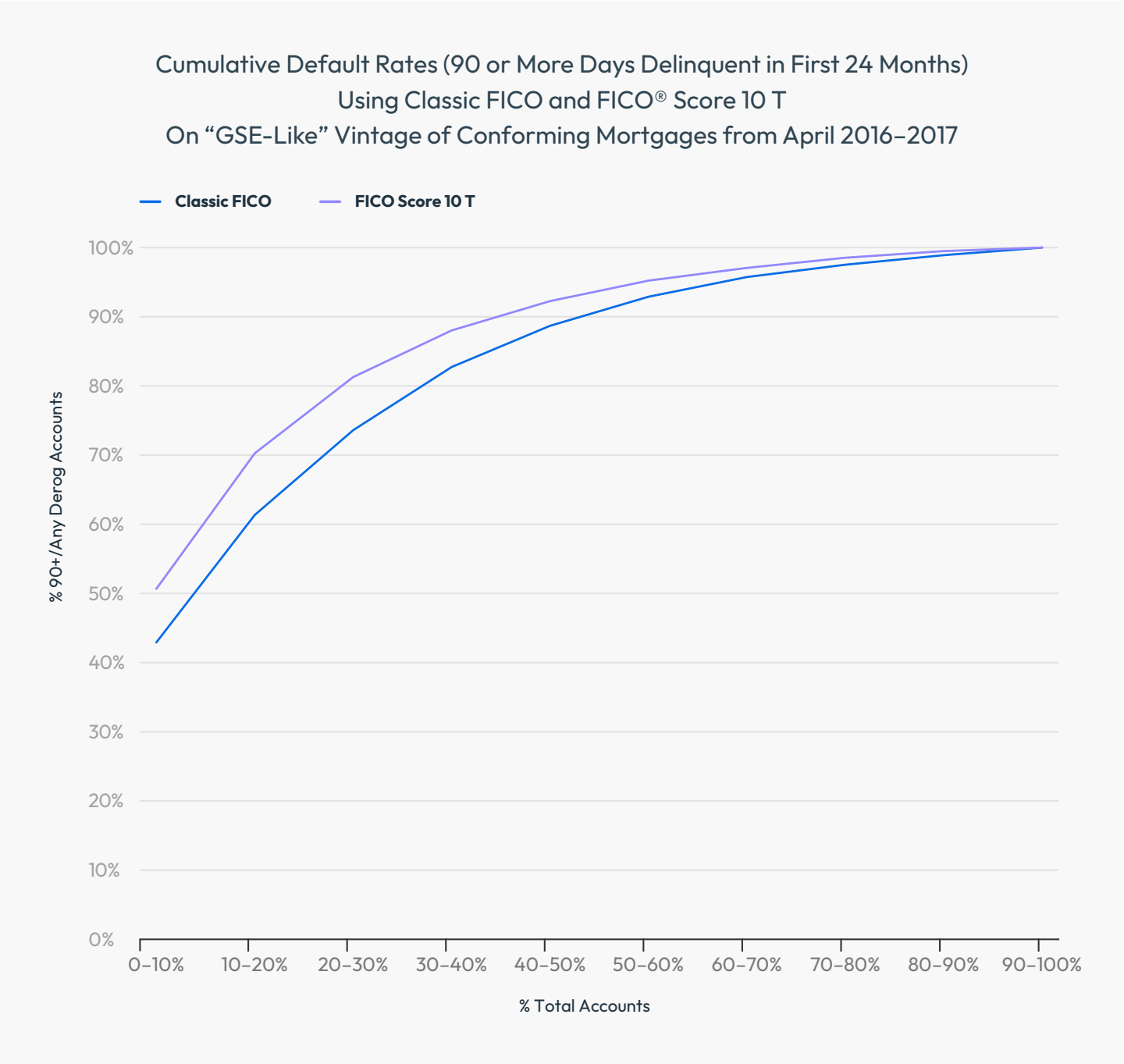


Figure 4
Cumulative Default Rates (90 or More Days Delinquent in First 24 Months), by Credit Score Percentile, and Credit Score Version

Percentile of credit scores	Classic FICO, April '16–'17 Vintage	FICO® Score 10 T, April '16–'17 Vintage	% Relative Improvement
0–10%	43%	51%	18%
0–20%	61%	70%	15%
0–30%	74%	81%	10%
0–40%	83%	88%	6%
0–50%	89%	92%	4%
0–60%	93%	95%	3%
0–70%	96%	97%	1%
0–80%	98%	99%	1%
0–90%	99%	99%	1%
0–100%	100%	100%	0%

Figure 5

		Mortgage Vintage
	Score	April 2016–April 2017
K-S Statistic	Classic FICO	44.3
	FICO® Score 10 T	52.7
K-S Difference—Absolute		8.3
K-S Difference—Relative		18.8%

“Truncation: The Homecourt Disadvantage”

Because the GSEs generally utilize a Classic FICO cutoff of 620 for loan approval, the effect of re-weighting the data sample to mimic the distribution of Classic FICO® Score on the loans observed in the Fannie Mae Single Family Historical Loan Performance Dataset was to in essence truncate the sample with respect to Classic FICO. This can be seen in Figure 2 above, where the percent of approved loans with a Classic FICO Score below 620 on the Fannie Mae Single Family Dataset is effectively zero. Said another way, because of the cutoff, the historical performance database published by Fannie Mae has a distribution of Classic FICO Scores that stops or is truncated at 620. While the performance of Classic FICO should be measured through its entire score range from 300 to 850, the historical database only allows for measuring the performance of Classic FICO between 620 and 850. Because K-S measures the ability of a score to separate a population into the widest possible segments of the defaulters and non-defaulters (i.e., the purpose of a score), Classic FICO’s K-S is assessed lower than it really is when only considering loans from the 620–850 score range and not the entire 300–850 score range.

We have previously highlighted the adverse impact that such truncation, aka “homecourt disadvantage” can have on the incumbent score model (i.e., in this case Classic FICO) when it comes to conducting head-to-head model performance validations between the incumbent and a “challenger” model on the population of loans that were booked using the incumbent model. This adverse impact can be clearly observed in Figure 6 below, which shows the K-S statistic calculated on the 2016–2017 vintage of approved mortgages in our study both before and after the weighting adjustment was applied to mimic “GSE-like” portfolios. After the re-weighting is applied, the relative reduction in K-S value for Classic FICO is almost double that of the relative reduction in the K-S value for FICO® Score 10 T (red cells highlighted in Figure 6 below). In turn, the relative lift offered by FICO Score 10 T over Classic FICO nearly doubles in the re-weighted sample (green cells highlighted in Figure 6 below), from a 9.7% relative increase in K-S on the initial sample to 18.8% after re-weighting. Note that this material impact is observed even though the more correlated the “challenger” score is to the incumbent score, the less pronounced the effect of this truncation is expected to be. That is because the more similar the two scores are, the more likely they are to both “agree” in scoring consumers below 620 (the cutoff in our example), and therefore, in that case, the more likely the truncation affecting the incumbent score is to also impact the challenger score. And yet, in spite of the many similarities in design blueprint between Classic FICO and FICO Score 10 T, the truncation impact between those two FICO Score versions is still very clearly visible. Importantly, any negative truncation impact on Classic FICO when compared to FICO Score 10 T is less than the expected negative impact of truncation on Classic FICO when comparing it to VantageScore, especially given the differences in design blueprint).

Figure 6

April 2016–April 2017 Vintage				
	Score	Sample of All Conforming Mortgages ("Before Re-weighting")	Sample of GSE-like Conforming Mortgages ("After Re-weighting")	K-S Difference—Relative
K-S Statistics	Classic FICO	52.0	44.3	-14.7%
	FICO® Score 10 T	57.0	52.7	-7.6%
K-S Difference—Absolute		5.0	8.3	
K-S Difference—Relative		9.7%	18.8%	

This analysis underscores the disadvantage that Classic FICO faces in any head-to-head tests conducted on datasets that reflect (or are re-weighted to reflect) the GSE's loan approval population. As a result of the GSE's reliance on Classic FICO® Score as a key threshold for approve/decline, these datasets are materially truncated with respect to Classic FICO, placing new "challenger" scores (including FICO® Score 10 T) at a distinct advantage when conducting a comparison of model performance metrics. Therefore, any VantageScore 4.0 comparison to Classic FICO using the Fannie Mae Single Family Historical Loan Performance Dataset would also clearly benefit VantageScore. In fact, given that VantageScore 4.0 is almost certainly less correlated to Classic FICO than FICO Score 10 T is, the impact of truncation could be even more pronounced than the 3.3 points of K-S seen for FICO Score 10 T vs. Classic FICO (see Figure 6: 8.3 K-S after re-weighting vs. 5 K-S before re-weighting). If so, any VantageScore 4.0 performance of less than 3.3 K-S points over Classic FICO would mean that VantageScore actually underperforms Classic FICO on an untruncated sample. This truncation effect negatively impacts the calculation of Classic FICO's strength by 15% or more, artificially boosting VantageScore 4.0's performance in comparison—a boost big enough that it calls into serious question whether VantageScore actually outperforms Classic FICO at all.

Third-party Studies Comparing Classic FICO to VantageScore 4.0

While FICO has not conducted any analyses directly comparing Classic FICO or FICO® Score 10 T to VantageScore 4.0 on the same dataset (as FICO does not have access to VantageScore 4.0 data), a number of studies have been published by third parties that compare Classic FICO to VantageScore 4.0 using datasets such as the GSE loan level historical datasets.

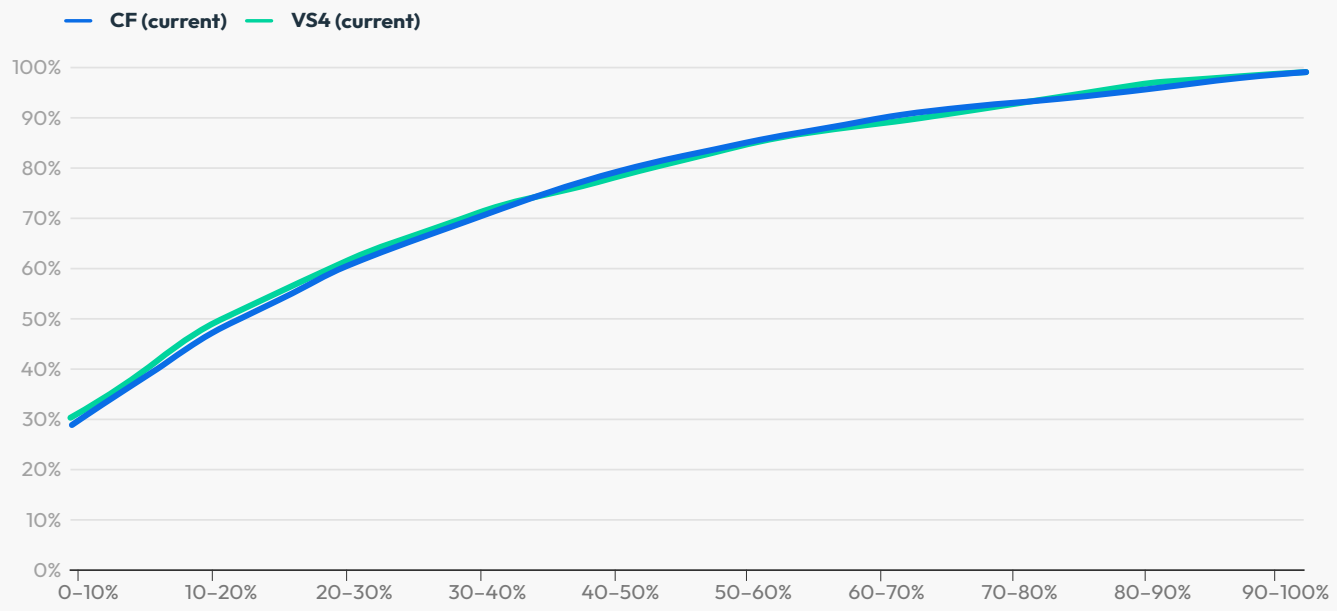
Of those third-party studies, it appears that only the Urban Institute conducted a direct comparison of the predictive strength of Classic FICO and VantageScore 4.0 via standard model performance metrics. Examining loans from 2013 to 2023 in the Fannie Mae Single-Family Loan Performance Database, and utilizing an ever 90+ days past due performance definition, Urban published ROC curves and the data underlying these ROC curves comparing the two scores. The key quote in the executive summary of its findings was as follows:

"Both credit scoring models effectively distinguish between high-risk and low-risk borrowers. VantageScore 4.0 is marginally more effective at identifying high-risk borrowers from among those with the lowest credit scores, though the differences are small."

How small were those marginal differences and how were they quantified? A visual quantification can be found in the slight "white space" between the ROC curve for "CF" (Classic FICO) and "VS4" (VantageScore 4.0) shown in Figure 7 below. An empirical quantification can be taken from the table of data underlying the Urban Institute's ROC curve calculation (see Figure 8 below), which shows in the lowest scoring decile of consumers that VantageScore 4.0 captured 30% of subsequent defaulters while Classic FICO captured 29%, for a relative difference of 3.4% more defaulters captured in the lowest decile.

Let's contrast those results comparing Classic FICO and VantageScore 4.0 to our study comparing Classic FICO and FICO® Score 10 T. Visually, the difference in "white space" in the Classic FICO and FICO Score 10 T ROC curves (See Figure 3 above) is dramatically more pronounced than what was observed in the Urban Institute's comparison of ROC curves for Classic FICO and VantageScore 4.0 (see Figure 7). And while the Urban Institute quantified the relative improvement in subsequent defaulters captured in the lowest scoring decile of the population at a mere 3.4% for VantageScore 4.0, our calculations—based on a data sample of mortgages opened between April 2016 and April 2017 and substantially similar to a vintage examined by the GSEs as part of the "Credit Score Models and Reports Initiative"—find that FICO Score 10 T offers an impressive 18.1% relative improvement (see Figure 4 above). This clearly highlights that FICO Score 10 T offers significant relative improvement over VantageScore 4.0.

Figure 7



Source: Urban Institute calculations from Fannie Mae data.
Note: CF= Classic FICO; VS4= VantageScore 4.0

Figure 8
Cumulative Default Rates (Ever 90 or More Days Delinquent), by Credit Score Percentile and Scoring Method

Percentile of credit scores	CF (Current)	CF (Ave.)	CF (Max.)	VS4 (Current)	VS4 (Ave.)	VS4 (Lowest)	VS4 (Median)	VS4 (Highest)
0–10%	29%	31%	31%	30%	32%	33%	32%	32%
0–20%	47%	50%	49%	49%	51%	51%	51%	50%
0–30%	61%	63%	63%	62%	63%	63%	63%	63%
0–40%	71%	73%	73%	72%	73%	73%	73%	72%
0–50%	80%	81%	81%	79%	81%	80%	80%	80%
0–60%	86%	87%	87%	86%	86%	86%	86%	86%
0–70%	91%	91%	91%	90%	91%	91%	91%	90%
0–80%	94%	95%	95%	94%	95%	95%	95%	94%
0–90%	97%	98%	98%	98%	98%	98%	98%	98%
0–100%	100%	100%	100%	100%	100%	100%	100%	100%

Source: Urban Institute calculations from Fannie Mae data.
Note: CF= Classic FICO; VS4= VantageScore 4.0

Keep in mind, the Urban Institute found that VantageScore 4.0 was only "marginally more effective" than Classic FICO—a score built over 20 years ago—in spite of the fact that its study was conducted on the Fannie Mae Single Family Loan Performance dataset that is materially truncated with respect to and negatively impacting Classic FICO. As noted above, this truncation has the effect of weighing down Classic FICO model performance metrics, which makes it easier for a "challenger" score to exhibit more significant improvement in risk prediction. Importantly, as discussed above, adjusting for this truncation effect could wipe out any observed lift of VantageScore 4.0 over Classic FICO.

Impact of VantageScore's Use of Mortgage Variables on the Predictive Comparison

When comparing the Classic FICO Score or FICO® Score 10 T to VantageScore 4.0 for mortgage origination, it is important to note that the VantageScore results not only include the truncation benefits discussed above, but also include the benefits of using mortgage-specific variables in the VantageScore models. Without the use of such predictive variables, the predictive power of VantageScore 4.0 on mortgage originations would be even lower. In contrast, for reasons discussed below, FICO deliberately avoided the use of mortgage-specific variables in the development of FICO Score 10 T. The Classic FICO Score also excludes mortgage-specific variables. Given this absence of mortgage-specific variables, the performance of Classic FICO and FICO Score 10 T compared to VantageScore 4.0 in predicting mortgage origination losses is even more impressive.

For background, VantageScore's documentation regarding VantageScore 4.0 reason codes reveals the inclusion of multiple mortgage-specific variables. Among several mortgage-specific reason codes,[6] VantageScore lists for VantageScore 4.0 are "No open first mortgage accounts in your credit file" and "Lack of first mortgage account information." This means that borrowers who do not currently have an active mortgage and borrowers who have never had a mortgage have lower VantageScore 4.0 scores than they otherwise would and lower VantageScore 4.0 scores than otherwise identical borrowers with such mortgage history.

[6] In the U.S., credit risk models provide reason codes to inform borrowers in the context of adverse actions (such as loan declination) as to the principal reasons or key factors that negatively impacted the consumer's credit score considered in the adverse action.

FICO data scientists designed FICO® Score 10 T with improvement in predictive performance in mind for all segments, including the mortgage industry. But, rather than letting mortgage-specific variables into the model (and requiring Americans to have mortgage history on file to get a higher score), FICO data scientists intentionally left out those variables. Our new score, just like Classic FICO before it, was carefully constructed from the standpoint of palatability to ensure that the factors affecting the score are reasonable to consumers and lenders alike. For example, while credit mix is still an important dimension of the FICO® Score calculation, FICO data scientists uncovered additional methods to effectively capture this dimension by assessing and differentiating between revolving and installment account experience more generally, and across loan types, rather than utilizing mortgage-specific variables. During model development, FICO data scientists carefully considered the impact and palatability of each variable included in these scoring models and the associated reason codes that may be utilized by lenders in adverse action notices in relation to millions of mortgage applicants every year.

Because FICO does not include mortgage-specific variables in these models, FICO® Scores for first-time home buyers are not lower simply by virtue of not owning a home. Conversely, VantageScore's use of mortgage-specific variables in VantageScore 4.0 could lead to a lower score for a potential first-time home buyer with an otherwise solid breadth of credit experience, and result in a mortgage loan application being rejected, because the applicant did not already have a mortgage loan. Even more likely, because of the highly refined and segmented nature of mortgage pricing, using a lower VantageScore 4.0 for an applicant who did not already have a mortgage loan could result in a potential first-time home buyer paying a higher rate than a similarly situated borrower who already owns a home. Worse yet, despite VantageScore's claims of inclusion and access to credit, its use of mortgage-specific variables in the VantageScore could actually have the opposite impact on homeownership. If VantageScore's inclusion of mortgage-specific variables results in lower scores for Americans who don't currently or have never owned a home, such VantageScore treatment could detrimentally impact the chances of approval for those applicants, likely disproportionately impact young people, members of the military, and others from traditionally disadvantaged groups, seeking first-time homeownership.

Impressively, FICO® Score 10 T outperforms VantageScore 4.0 in predictive power, while avoiding the use of mortgage-specific variables and the potential for negative impacts that go with it.

Impact of VantageScore's Changes to VantageScore 4.0 After Submission to GSEs

Under the FHFA's Joint Enterprise Credit Score Solicitation process, both FICO and VantageScore submitted applications for consideration of their respective credit scores. The GSEs proceeded to obtain research datasets for FICO® Score 10 T and VantageScore 4.0 and evaluated these scores for accuracy and reliability. Subsequently, on August 10, 2022, nearly two years after both FICO and VantageScore submitted their models for evaluation by the GSEs and FHFA, VantageScore announced that it was modifying VantageScore 4.0 to exclude a large category of predictive credit file information—Medical Debt Collection Records.

We acknowledge that medical collections are a legitimate topic for debate by policy makers and others on their appropriateness for inclusion in credit files and credit scores. We further acknowledge that certain states have designed laws to prohibit the inclusion of medical collections in credit files for residents of their states. At the same time, medical collections are still reported to and included in the credit files of the three large nation-wide credit bureaus, Transunion, Experian and Equifax and are available for use in the majority of states. VantageScore's change to VantageScore 4.0, while applauded by some, still raises questions of a potential serious breach of the integrity of the FHFA and GSE Validation and Approval of Credit Scores process that would impact the comparison not only of FICO® Score 10 T and VantageScore 4.0, but the comparison of Classic FICO® Score and VantageScore 4.0 for purposes of determining whether VantageScore 4.0 now even outperforms Classic FICO.

VantageScore, itself, admitted that this change would impact the performance of its model when it announced the change on August 10, 2022 as follows: "Impact to the VantageScore models' predictive performance is expected to be minimal for a large segment of the population." Our research has similarly and consistently shown predictive value from the inclusion of medical collections on credit scores. While VantageScore attempts to downplay any impact on predictive performance, it could be material: after all, what is the effect of a "minimal" impact on a VantageScore 4.0 already only "marginally more effective"? And this leads to the problem of understanding the comparison of VantageScore 4.0 to FICO® Score 10 T.

As noted above, the historical database used to compare VantageScore 4.0 to Classic FICO® Score was published by the GSEs. But, which version of VantageScore 4.0 does it include—the one that includes medical collections or the one that doesn't? No stakeholder seemingly has any way of knowing. If it includes the version that considers medical collections, then the performance of FICO® Score 10 T over VantageScore 4.0 would be even greater than the already superior performance by FICO Score 10 T discussed in this paper.

Not only does it undermine determining whether VantageScore 4.0 now even outperforms Classic FICO, it raises serious doubt of whether the VantageScore 4.0 model evaluated by the GSEs and FHFA under the Joint Enterprise Credit Score Solicitation is even the same VantageScore model used to generate scores in the historical dataset made available by the GSEs for industry modeling and transition. Most troubling, it creates concerns about whether the current new model was even approved through the required process.

Changes to models destined for the mortgage industry after submission to the GSEs for evaluation is especially troubling because the risk is transferred to the GSEs and taxpayers. Typical constraints on model developers by lenders are distorted in the conforming market. Under such circumstances, if models can be changed solely at the discretion of the model developer after the fact, there would be virtually no constraints for any model developers to refrain from altering their models to curry favor with stakeholders, including lenders who transfer the risk to others. The race to the bottom is no longer theoretical. This highlights just one early example of the "race to the bottom" behavior that many market participants have come to fear and if allowed to continue could have significant and irreversible impacts on the mortgage ecosystem.

Conclusion

In this paper, we have discussed the results of a study that we conducted to quantify the lift offered by FICO® Score 10 T over Classic FICO on a sample of credit files designed to be substantially similar to the GSE's loan level historical datasets. Our findings are clear: FICO Score 10 T offers dramatic improvements in risk prediction, as much as a remarkable 18.8% relative improvement in K-S and 18.1% relative improvement in defaulters captured in the lowest decile of the population when comparing ROC curves.

At the same time, third party studies of Classic FICO and VantageScore 4.0 on the same GSE loan level data have found that VantageScore 4.0 is only "marginally more effective" than Classic FICO—a score built over 20 years ago. And this finding is in spite of the fact that Classic FICO is at a real disadvantage in any head-to-head comparison on GSE loan level data due to the effects of truncation and the absence of mortgage-specific variables.

Combining our findings on Classic FICO relative to the strong improvements shown by FICO® Score 10 T together with the findings of third-party studies comparing Classic FICO and VantageScore 4.0 leads to the straightforward conclusion that FICO Score 10 T is considerably more powerful than and significantly outperforms VantageScore 4.0 in predicting mortgage default risk. Impressively, FICO Score 10 T's improvement over Classic FICO was shown to be five times better than VantageScore 4.0's improvement, with FICO Score 10 T detecting 18% more defaulters in the critical score decile commonly used for mortgage originations versus just 3.4% for VantageScore 4.0.

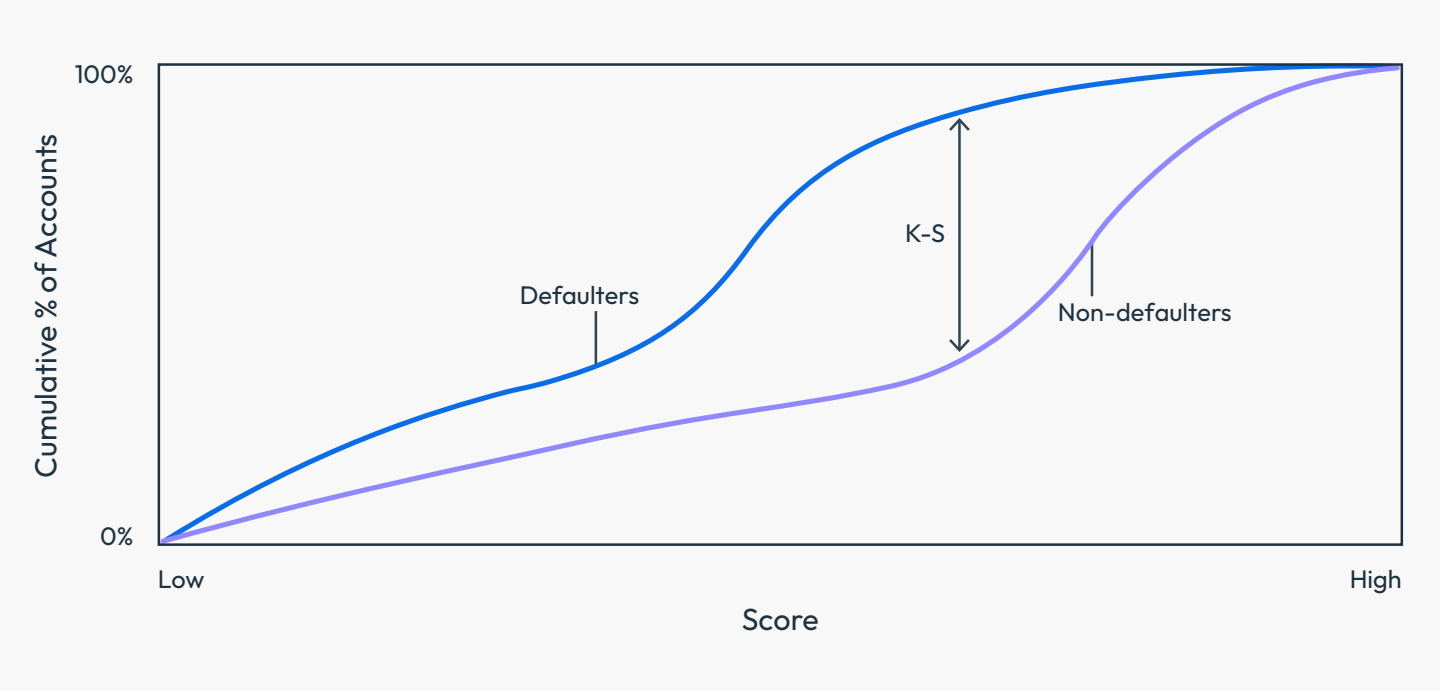
Based on enhanced prediction, use of FICO® Score 10 T rather than VantageScore 4.0 by market participants throughout the mortgage industry will drive significantly more loan approvals for prospective borrowers and, due to better model performance for mortgage insurers, investors and others, improve mortgage pricing and lower costs for borrowers—benefiting millions of Americans.

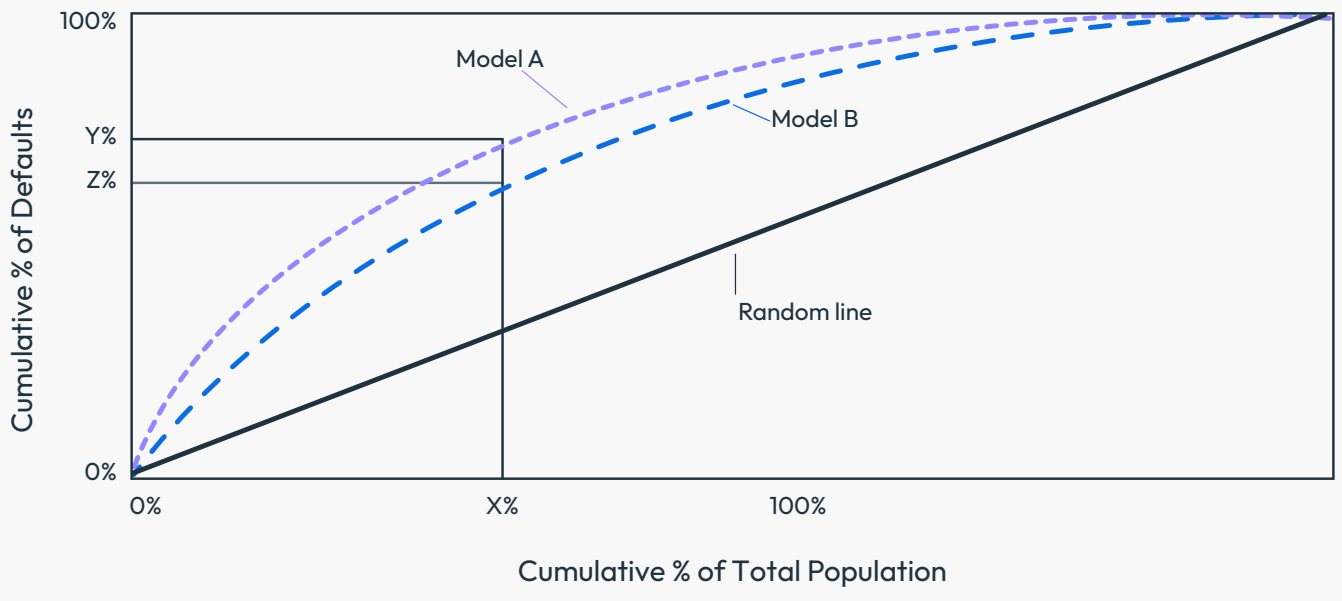
And it is not just FICO who recognizes the superior performance of FICO® Score 10 T versus the competition: many sophisticated participants in the mortgage industry are already adopting FICO Score 10 T. Since 2023, clients representing over $300 billion in annualized mortgage originations and approximately $1.5 trillion in eligible mortgage portfolio servicing have entered into agreements for the FICO Score 10 T early adopter program enabling them to pull FICO Score 10 T at origination.

At FICO, we know that we compete on predictive performance, but it doesn't stop there. Our scores are used by lenders because they are the trusted, reliable, independent, recognized around the world, and time-tested through multiple economic cycles scores. But, predictive power counts, too. And, as these results show, FICO® Score 10 T wins hands down on this dimension as well, and there is every reason to believe that it's not even close.

Appendix A: Measures for Evaluating and Comparing Predictive Models

Kolmogorov-Smirnov (K-S) Statistic: The K-S statistic represents the maximum difference between the cumulative distributions of two groups (e.g. non-defaulters and defaulters). A zero value indicates that the credit score fails to differentiate between defaulters and non-defaulters; a value equal to 100 indicates that the credit score perfectly differentiates defaulters from non-defaulters.





Receiver Operating Characteristic (ROC) Curve: Also known as the Trade-Off Curve, Lorenz curve, or Lift curve, the ROC curve is a plot of ascending accumulation of one group of accounts (e.g., defaulters) vs. another group (e.g., non-defaulters or the "total" population). The ROC curve visually shows a model's effectiveness, as opposed to a summary statistic. It is useful for visually comparing multiple models' effectiveness at a particular operating point or across the spectrum of the score distribution. For example, in the graph below, at the lowest scoring X% of the cumulative total population, Model A identifies Y% of cumulative defaults, and model B identifies Z% of cumulative defaults. Model A is the more effective model as it identifies a greater percentage of defaults.

To learn more, visit

www.fico.com

FICO® SCORE

5374P 7/25