Machine Learning–Based Systematic Investing in Agency Mortgage-Backed Securities

Nikhil Arvind Jagannathan and Qiulei(Leo) Bao

Nikhil Arvind Jagannathan
is a portfolio manager and senior mortgage-backed securities research analyst at Franklin Templeton in Singapore.
nikhil.jagannathan@franklintempleton.com

Qiulei(Leo) Bao
is a research analyst/data scientist at Franklin Templeton in New York, NY.
leo.bao@franklintempleton.com

KEY FINDINGS

- The authors introduce a strategy to systematically invest in specified pools within the agency mortgage-backed securities market and display its outperformance against a benchmark index.
- Machine learning (ML) prepayment models with the help of cloud computing demonstrate enhanced pattern recognition without losing interpretability on single variable influence.
- By combining loan-level and pool-level models, the ML prepayment model shows improved accuracy and competitiveness compared with standard industry models.

ABSTRACT

With a total outstanding balance of more than $8 trillion as of this writing, agency mortgage-backed securities (MBS) represent the second largest segment of the US bond market and the second most liquid fixed-income market after US Treasuries. Institutional investors have long participated in this market to take advantage of its attractive spread over US Treasuries, low credit risk, low transaction cost, and the ability to transact large quantities with ease. MBS are made of individual mortgages extended to US homeowners. The ability for a homeowner to refinance at any point introduces complexity in prepayment analysis and investing in the MBS sector. Traditional prepayment modeling has been able to capture many of the relationships between prepayments and related factors such as the level of interest rates and the value of the embedded prepayment option, yet the manual nature of variable construction and sheer amount of available data make it difficult to capture the dynamics of extremely complex systems. The long history and large amount of data available in MBS make it a prime candidate to leverage machine learning (ML) algorithms to better explain complex relationships between various macro- and microeconomic factors and MBS prepayments. The authors propose a systematic investment strategy using an ML-based mortgage prepayment model approach combined with a coupon allocation optimization model to create an optimal portfolio to capture alpha versus a benchmark.

Mortgage-backed securities (MBS) are contracts that entitle their holders to the cash flows of mortgage loans. With a total outstanding balance of more than $12 trillion as of this writing, they constitute the second largest fixed-income sector (behind Treasuries) in the US market. Most of this market, $8.3 trillion, is composed of agency MBS, securities where timely principal and interest payments are guaranteed by the government sponsored entities (GSEs, or agencies) Fannie

Mae (FNMA), Freddie Mac, and Ginnie Mae. The defining features of this market are its large size, low credit risk, and the presence of prepayment risk.

Prepayment risk is the risk to investors of receiving unscheduled, early principal payments. It is a feature of investing in MBS caused by the mortgage borrowers' ability to pay off their mortgages at any time. When principal is returned earlier than expected, investors tend to lose the premium over face value paid and, because mortgage prepayments tend to increase as interest rates decline, investors receive principal in a lower yielding environment. Those cashflows received need to be deployed at lower prevailing rates, leading to reinvestment risk. One might think of agency MBS as akin to Treasuries with embedded call options written to borrowers. To compensate investors, agency MBS often yield 50–100 basis points (bps) in return to Treasuries of similar duration. Mortgage prepayments are at the heart of all agency MBS valuation and analysis.

In recent years some sophisticated investors have sought to take advantage of the enhanced pattern recognition and accuracy of machine learning (ML) models in agency MBS prepayment modeling. You may have heard about ML in various contexts. Models that provide automatic detection of both linear and nonlinear patterns and emphasize predictive accuracy over interpretability are the purview of ML. These models tend to be computationally intensive and require large amounts of data. For many tasks, they are dramatically more accurate than linear models. Examples of such models include gradient-boosted decision trees, neural networks, and support-vector machines.

Given that an MBS is a pool of individual mortgage loans, it would be optimal to have prepayment models that use both loan-level and pool-level data. The availability of vast amounts of data, combined with the computational complexity of processing billions of historical loans across the universe of mortgages, has led to only a few industry-standard prepayment models using loan-level data for prepayment prediction. Modern technological frameworks related to big data and cloud computing have enabled us to incorporate loan-level information to model prepayments.

Previous literature has shown that ML models for mortgage prepayment are proving competitive, if not superior, to the traditional modular mortgage prepayment model.[1] By combining a loan-level gradient-boosted tree model with a pool-level random forest (RF) tree model, we are able to generate more accurate prepayment predictions compared with a third-party model.[2] Over the period of July 2021–March 2022, we examined 2,523 liquid FNMA 30-year MBS pools with 13,014 observations in which the ML prepayment model achieved a root mean square error (RMSE) of 2.2% conditional prepayment rate (CPR) and an industry-standard model achieved an RMSE of 4.0% CPR on the same set of pool observations.

To generate alpha from an ML prepayment model, it is possible to implement a systematic investment strategy using ML-based prepayment predictions to compute horizon returns for each MBS pool and maximize overall portfolio returns using a proprietary optimizer.

Starting in December 2020, the authors constructed such a strategy. The optimizer was calibrated constraining various portfolio parameters including duration, convexity, loan size, bid-ask spread, and premium over to-be-announced (TBA) securities. The resulting optimal portfolio has outperformed the Bloomberg US MBS Fixed Index since the inception of the strategy. We demonstrate a practical way to implement a quantitative ML-driven systematic approach to invest in the agency MBS market.

---

[1] Glenn M. Schultz and Frank J. Fabozzi (2021) showed that ML prepayment models are proving competitive, if not superior, to the traditional modular mortgage prepayment model in *Rise of the Machines: Application of Machine Learning to Mortgage Prepayment Modeling*.

[2] Using Yield Book predictions as comparison.

BUILDING A SYSTEMATIC STRATEGY IN THE MOST SUITABLE INVESTMENT UNIVERSE

Before we identify the investment universe within agency MBS for the systematic strategy, a brief overview of the market in general is needed. Securitization revolutionized the US mortgage market by converting idiosyncratic financial assets into one that is more accessible and easier to understand by investors. By bearing the credit risk of the underlying loans and transforming loans into guaranteed pass-through securities, the agencies improved liquidity in the secondary mortgage market. Once a pass-through is created, it subtracts servicer and guarantee fees and distributes prorated shares of pool cash flows (principal and interest) from the underlying loans to investors.

To further improve liquidity in mortgage pass-throughs, pools that offer similar risks and rewards are considered interchangeable by the standard market convention and, therefore, can substitute for one another. Pools that can be substituted usually include loans guaranteed by the same agency, with the same coupon and original term to maturity. As a result of this convention, substitutable pools trade in the TBA market where pool type, coupon, amount, settlement date, and a price are agreed upon. The advantages of the TBA market are that it allows originators to sell pools that were not formed yet, freeing both buyer and seller from considering pool characteristics.

Agency MBS trade in two channels, the TBA market and the specified pool market. Although the majority of MBS trading happens in the TBA market, we felt a systematic investment strategy that leverages pattern recognition capabilities of an ML model would have the greatest value in the context of generating alpha in the specified pool market. The TBA market is a forward market in which only very general characteristics are specified by investors prior to delivery. Sellers of TBAs have discretion over what pools they can deliver to the buyer as long as the pools satisfy good-delivery guidelines (Killian 2015). A strong incentive exists for the seller of the TBA contract to deliver the cheapest pools available to them; as such, the TBA market is also known as the worst-to-deliver market. On the other hand, the specified pool market enables investors to purchase individual MBS pools with specific favorable collateral and prepayment attributes. Investors will pay a premium over TBA to avoid receiving pools with undesirable prepayment characteristics in the TBA market.

Given the complex and dynamic relationship between pool characteristics and prepayment, an ML model should capture the nuance better, given its ability to detect nonlinear patterns and that it is not confined to a certain functional form.

To use the prepayment models to select pools in a systematic way, we can employ a pool-level prepayment forecast to estimate the constant-spread holding-period return for a given horizon. An optimizer can then be configured to calculate weights on an investment universe to maximize return subject to customized portfolio constraints.

To summarize, to systematically invest in the agency MBS market, we construct a portfolio in the specified pool market with pools selected by a proprietary optimizer leveraging prepayment predictions from an ML model.

PREPAYMENT MODELING

The art and science of building mathematical models to forecast prepayments stretches back to the 1980s (Waldman 1985). The most well-known and consequential driver of prepayment activity is the average level of mortgage rates. Borrowers tend to refinance their mortgages when they can obtain a rate sufficiently below the rate at which they originally borrowed. This relationship between prepayments and interest rates gives MBS their characteristic negative convexity: MBS prices do not rise as much as noncallable bond prices when rates fall. In addition to rate refinances, all the following actions result in prepayments:

- Existing home sales: The sale of a home results in mortgage prepayment.
- Cash-out refinances: Refinancing into a larger loan results in prepayment.
- Defaults: Because of the agency guarantee, a foreclosure registers as prepayment.
- Curtailments: Some borrowers pay excess principal each month to lower future interest payments, resulting in partial prepayment.

A forecasting model must consider a large number of variables at the intersection of industry trends, macroeconomics, and borrower demographics that drive prepayments.

Note that prepayments usually are measured in terms of CPR. Single monthly mortality (SMM) is simply the principal prepayments over the outstanding balance for the month. CPR is an annualized measure of unscheduled balance received over a period. It is defined in Equation 1:

$$
\begin{aligned}
CPR_k &= 1-(1-SMM)^{\frac{12}{k}} \\
&= 1-\left(1-\frac{Unscheduled\ principal\ received\ over\ k\ months}{Scheduled\ current\ balance\ at\ end\ of\ k^{th}\ month}\right)^{\frac{12}{k}}
\end{aligned} \tag{1}
$$

A short-term prepayment model can be used to predict prepayment speeds over a near horizon, where the impact of interest rate volatility should be muted. Although this is not always the case, the assumption is not without foundation: Mortgage refinancing typically closes 30–45 days from application, and borrowers often shop lenders for better rates before locking in. Thus, a significant portion of the information needed to make an accurate short-term forecast is available. As a result, short-term models can be used to estimate plausible holding period returns over limited horizons.

The advantage of a dedicated short-term model is the ability to include a large number of variables that would be impractical to incorporate in longer term models. Long-term models require that all time-varying inputs be forecasted, for example, interest rates, volatility, housing prices, the level of unemployment, consumer confidence, and average wages, etc. As the number of inputs requiring forecasts grow, error accumulates. Given the complexity of long-term forecasts on model inputs, longer term models tend to use a smaller set of variables and can miss out on more nuanced signals.

We will discuss the modeling technique and performance statistics of the ML prepayment model in detail in later sections. Next, we emphasize the two distinct advantages of a prepayment model, utilizing ML techniques compared with traditional linear models and modeling loan-level data.

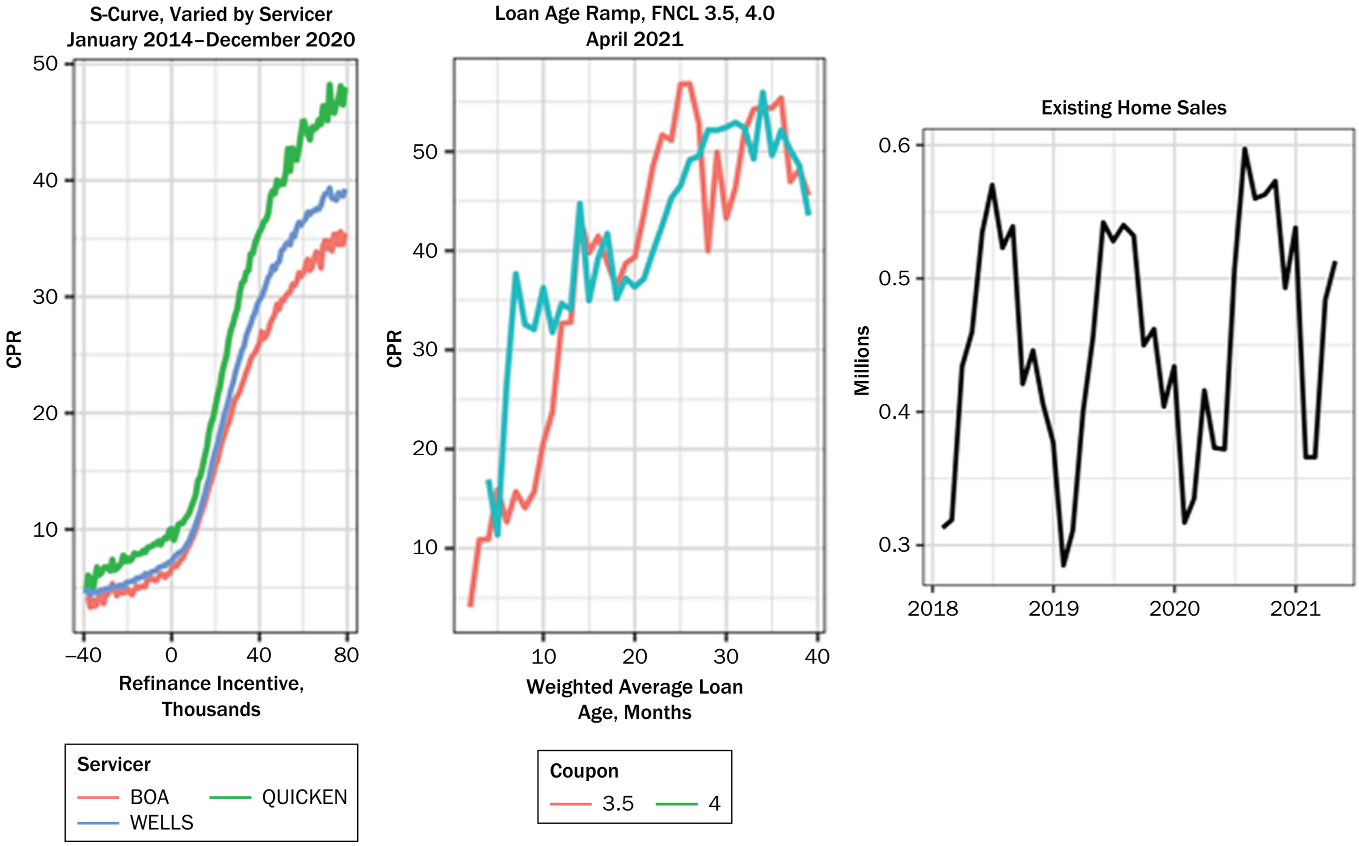
Traditional Linear Model vs. Machine Learning Model

Because of the size of the sector and its long history, many prepayment behaviors are well researched and understood. For example, the relationships between loan age and likelihood to prepay, borrowers' S-shaped response to lower mortgage rates, the seasonality of housing turnover, and so on, are well known.

Exhibit 1 depicts three well-known prepayment patterns. The left graph uses loan-level data to show the relationship between aggregate prepayment activity and

EXHIBIT 1

Three Well-Known Prepayment Patterns



SOURCES: Franklin Templeton Investments, Fannie Mae, and the National Association of Realtors.

the borrower refinance incentive, an estimate of borrowers' monetary incentive to refinance. Because of its shape, this relationship is known as the S-curve. The middle graph uses pool-level data to show the relationship between loan age (in months) and prepayment activity. This relationship is often described as ramplike. The right graph exhibits the seasonal pattern in existing home sales that drives base prepayment speeds.

Prepayment modeling has traditionally fallen under the domain of econometric models, a term used by economists to describe the statistical modeling of economic phenomena. Linear models dominate this field. Linear models are simple but flexible models that quantify the proportional impact of one or several variables on another. For example, they facilitate such statements as "on average, pools of New York mortgages pay five CPR less than other pools," or "on average, for each additional $100,000 dollars of loan balance, pools pay five additional CPR." Most industry prepayment models, such as The Yield Book, Locus, Black Knight, and The Bloomberg Agency Model, are linear models at their core.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon \tag{2}$$

An example of a linear model is shown in Equation 2. The X_is represent independent variables that are used to predict a target variable Y. The β_is reflect numerical values that are estimated by the model.

Linear models have clear advantages and disadvantages. Their main advantage is that they are transparent and interpretable. They also are computationally tractable. Their main flaw lies in that the real world is anything but linear. To incorporate nonlinearity, econometricians must explicitly define the nonlinearity they are interested in and include a transformed variable as an input to their model. For example, to capture the S-shape of borrowers' response to the current level of mortgage rates, an econometrician would create a variable analogous to the arctangent of the difference between borrowers' original mortgage rate and current prevailing rates. Equation 3 shows a linear model with a nonlinear independent variable. The properties of the arctangent function, like the sigmoid function, make it suitable for modeling S-curves, notably that $lim_{x \to \pm\infty} f(x) = \pm c$. This manual process makes it difficult and unwieldly to capture the dynamics of extremely complex systems.

$$Y = \beta_0 + \beta_1 arctan(X_1) + \cdots + \epsilon \tag{3}$$

Compared with traditional linear models, ML models tend to have higher accuracy on prediction and better capability on learning complex relationships between variables. Some of the drawbacks of ML models are the difficulty of clearly interpreting results, intensive computation time, and the related high cost of the required resources.

ML models are used to learn patterns in data in either one of two ways, supervised or unsupervised learning. In supervised learning, the goal is to learn the rules between a set of inputs and outputs, and in unsupervised learning only input data are provided. In the context of prepayment modeling, supervised learning should be adopted, given that we aim to learn the rules between prepayment and other input features.

Here we address the interpretability of ML models. Even with a rather complicated model, there are still ways to understand the patterns the models identified given the input variables. One example would be to use Shapley Additive Explanations (SHAPs).

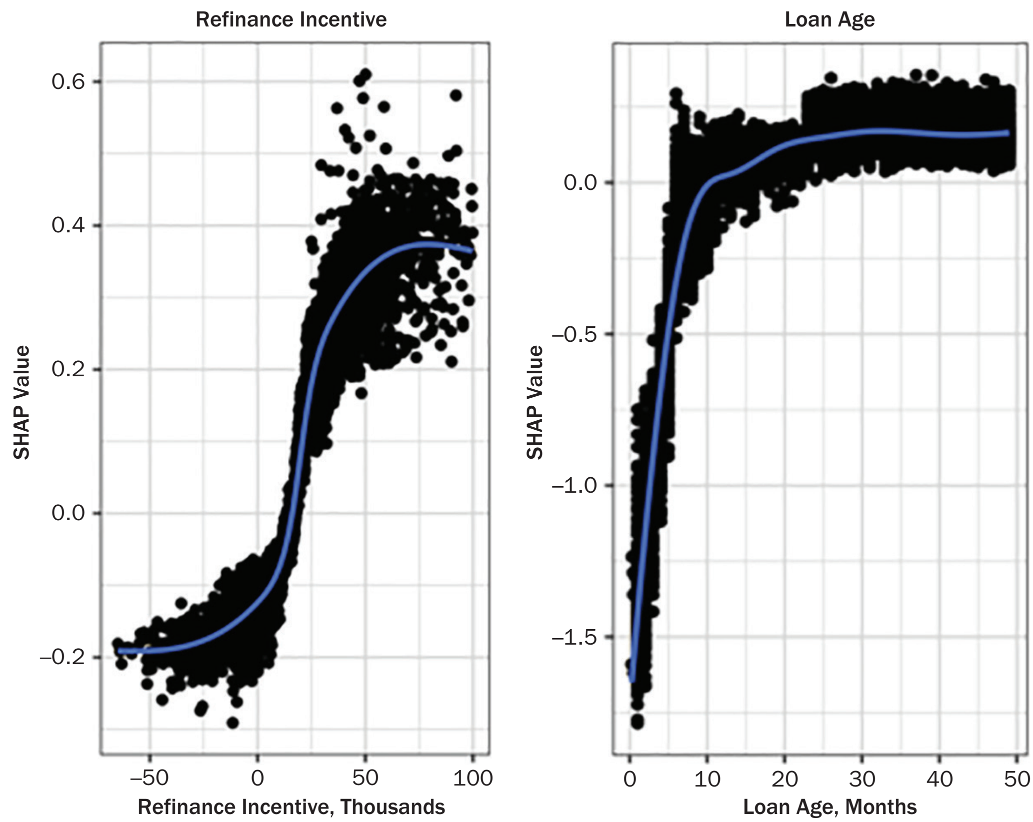
SHAP values use game theory to determine the marginal contribution of each variable to the predicted output; in terms of prepayment modeling, it is the probability of a borrower prepaying within the analysis horizon. The ML model does not provide the contribution of each variable to the prepayment prediction of the loan. A prediction can be explained by assuming that each attribute of a loan is a player in a game in which the prepayment prediction is the payout. SHAP values tell us how to fairly distribute the payout among the features. To make it easier to compare, Exhibit 2 shows two relationships regarding prepayment similar to those in Exhibit 1. The left graph shows the relationship between refinance incentive and borrower likelihood to prepay, and the S-curve is visible. The right graph shows the relationship between loan age and borrower likelihood to prepay. Again, similar to Exhibit 1, the ramp is clearly shown. SHAP values are obtained for each loan-level prediction by perturbing the relevant input variable and observing the resulting change in output. Similar analysis can be applied to all variables in the ML model, which might help us to understand some less obvious relationships.

Loan-Level Modeling

Weighted average collateral information at the pool level has been available since the 1980s and provides the foundation for most prepayment models. Loan-level data became available around 20 years later. Freddie Mac began releasing loan-level data in 2007, and FNMA and Ginnie Mae began doing the same in 2013. The release of loan-level data presents both challenges and opportunities to model MBS pools on a more granular basis.

EXHIBIT 2

Relationships Learned from the ML Model



SOURCE: Franklin Templeton Investments.

Loan-level data constitute a shorter window in MBS history than pool-level data, and the large universe and numerous fields make working with loan-level data (on the order of billions of records) difficult from a computational perspective.

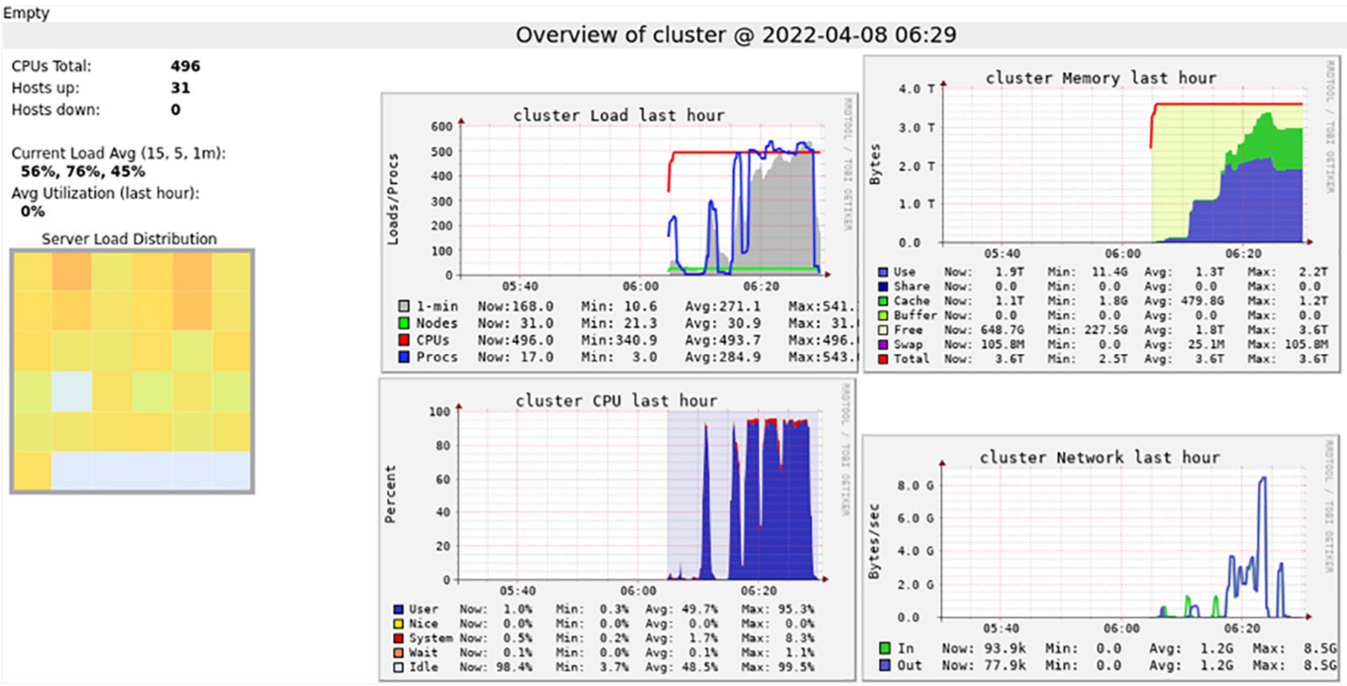
As for the computational challenge, the advancement of cloud computing enables leveraging multiple machines in parallel when training and running a model, which significantly decreases the computing time. Cloud computing is a relatively new paradigm in software development that facilitates broader access to parallel computing via vast, virtual computer clusters. In simple terms, parallel computing can divide larger problems into independent smaller components that can be executed simultaneously by multiple processors communicating via shared memory. Exhibit *3* shows a snapshot of using 31 clusters with 496 central processing units (CPUs) and network and memory loads across the grid when running the ML prepayment model. Given the divide-and-conquer characteristic of the algorithm, we can easily scale the model on more data by adding more clusters.

LOAN-LEVEL MODEL DEEP DIVE

The ML prepayment model has multiple submodels for each of the major fixed-rate collateral types built for short-term horizon prepayment projections: uniform mortgage-backed securities (UMBS) 30-year, UMBS 15-year, and Ginnie Mae II 30-year. Given that the modeling techniques are comparable, we will use the UMBS 30-year

EXHIBIT 3

Snapshot of Cluster Overview for Cloud Computing



SOURCE: Franklin Templeton Investments.[a]

[a]Cluster report from Databricks.

model as an example. The model is trained on more than 2 billion records covering the period from 2009 to the present.

The ML prepayment model uses a light gradient-boosting machine (LightGBM) for loan-level modeling. LightGBM is a distributed gradient-boosting framework for ML that is based on decision tree algorithms. Given that loan-level data contain billions of records, parallel computing is needed in order to train and run an ML model within a reasonable timeframe. The fact that LightGBM is distributed makes it possible for us to leverage parallel computing to use multiple clusters.

LightGBM uses gradient-boosting algorithms, which increases its prediction speed and accuracy, particularly with large and complex datasets. LightGBM relies on not one but multiple decision trees. To aggregate results, we use ensemble methods that combine several decision tree classifiers to produce better predictive performance. There are, however, more than one ensemble method to combine results. Boosting, specifically, is an iterative technique that adjusts the weight of an observation based on the last classification.

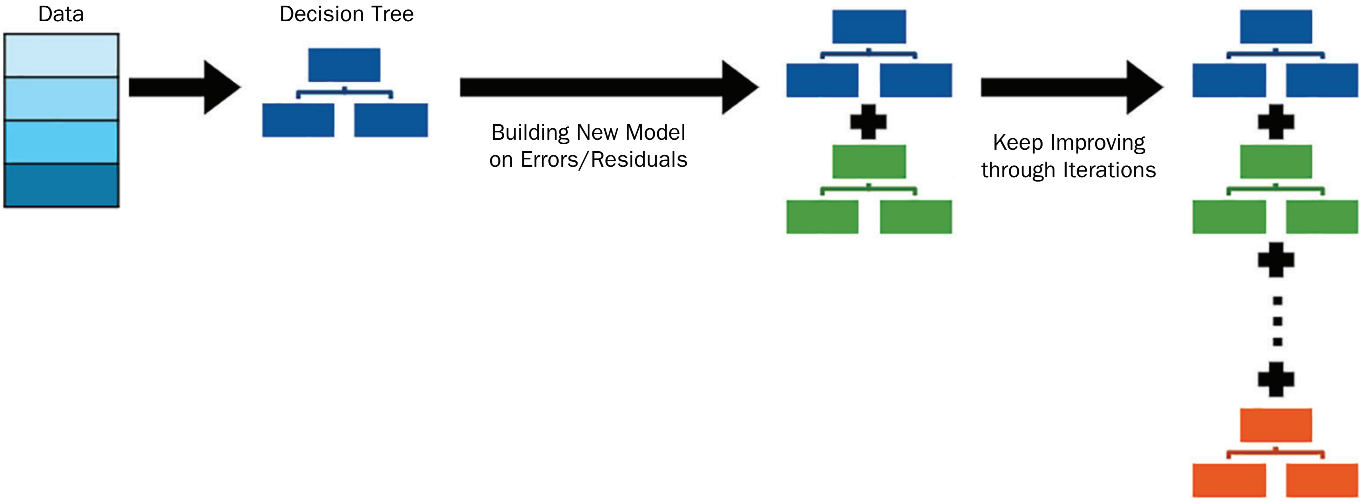
Exhibit 4 is an illustration of the modeling process. The idea of gradient boosting is to build models sequentially, and each subsequent model tries to reduce the errors of the previous model by building a new model on the errors or residuals of the previous model.

In terms of prepayment modeling, we use the gradient-boosting classification because we are trying to predict whether a loan will prepay (classified as 1) or not prepay (classified as 0).

Exhibit 5 shows feature-importance ranking, and we can see that multiple incentives such as payment indicator, refinance indicator, loan age, loan balance, and house price are dominant features of the model.

EXHIBIT 4

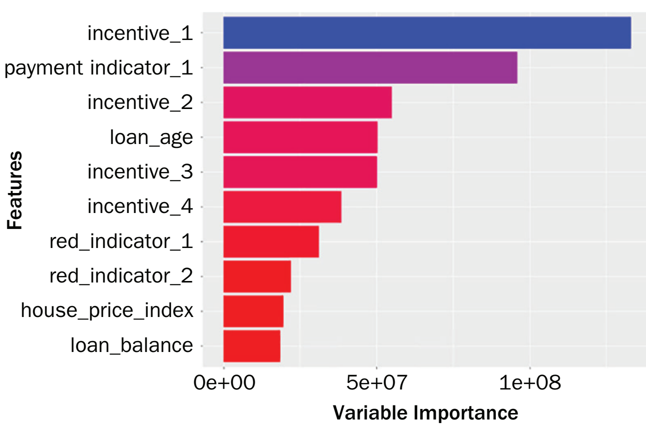
Illustration of the Gradient-Boosted Trees Modeling Process



SOURCE: Franklin Templeton Investments.

EXHIBIT 5

Feature-Importance Ranking for Loan-Level Model



SOURCE: Franklin Templeton Investments.

For classification problems, the receiver operating characteristic (ROC) curve is a common and effective performance measure. ROC is a probability curve and area under the curve (AUC) represents the degree of separability. For each instance, the model generates a probability between 0 and 1. Based on a threshold, the instance could be predicted as 1 if the probability is bigger than the threshold and 0 otherwise. Accordingly, with each threshold, a set of predicted values can be generated, and the true positive rate (TPR) and false positive rate (FPR) can be calculated. Exhibit 6 provides a simple example. Equations 4 and 5 show how TPR and FPR are calculated.

$$TPR = \frac{True\ \ positive}{True\ \ positive + False\ \ negative} \tag{4}$$

$$FPR = \frac{False\ \ positive}{True\ \ negative + False\ \ positive} \tag{5}$$

The ROC curve is plotted with TPR against FPR. The higher the AUC, the better the model's ability to classify. An AUC of exactly 0 has poor separability and indicates the model is predicting all 1s as 0s and 0s as 1s. A poor model has an AUC near 0, which means its predictions are highly misclassified. On the other hand, a good model has an AUC near 1. An AUC of exactly 1 indicates that all predictions are correct, resulting in a perfect separation. When the AUC is 0.5, it means the model has no separability; the performance is similar to that of a coin toss.

Exhibit 7 and Exhibit 8 show that our model has relatively good performance, with AUC = 0.76 on both training and testing data. Consistent performance on both datasets is important because we want the model to generalize well on out-of-sample data. Here the concept of overfitting needs to be introduced. Overfitting happens

EXHIBIT 6
Simple Example of the Calculation of TPR and FPR

Threshold = 0.5

TPR	FPR
$\frac{2}{2+0} = 100\%$	$\frac{1}{2+1} = 33\%$

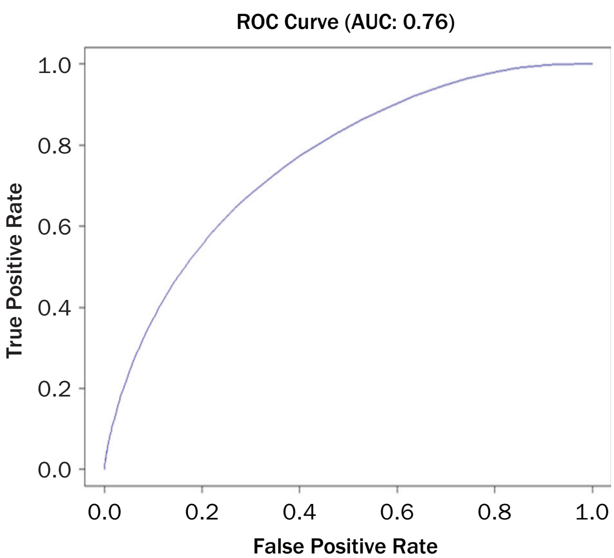
Probability	Actual	Prediction (threshold = 0.5)	Prediction (threshold = 0.8)
0.9	1	1	1
0.8	1	1	1
0.5	0	1	0
0.2	0	0	0

Threshold = 0.8

TPR	FPR
$\frac{2}{2+0} = 100\%$	$\frac{0}{2+0} = 0\%$

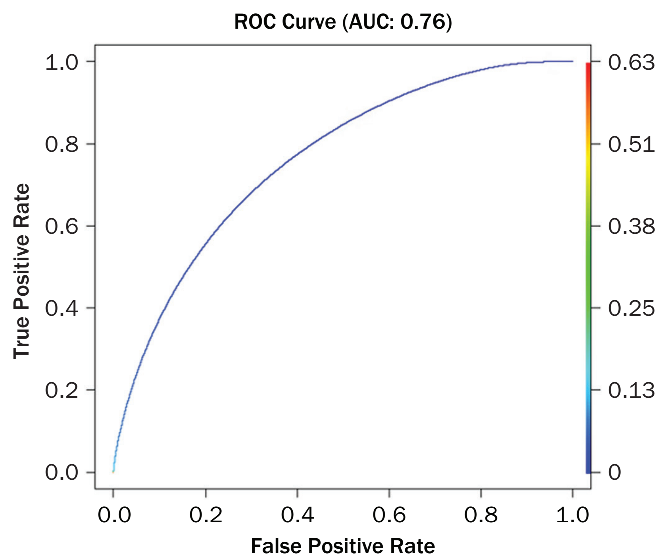
SOURCE: Franklin Templeton Investments.

EXHIBIT 7
Model Performance for Loan-Level UMBS 30-Year Model (ROC on train data)



SOURCE: Franklin Templeton Investments.

EXHIBIT 8
Model Performance for Loan-Level UMBS 30-Year Model (ROC on test data)



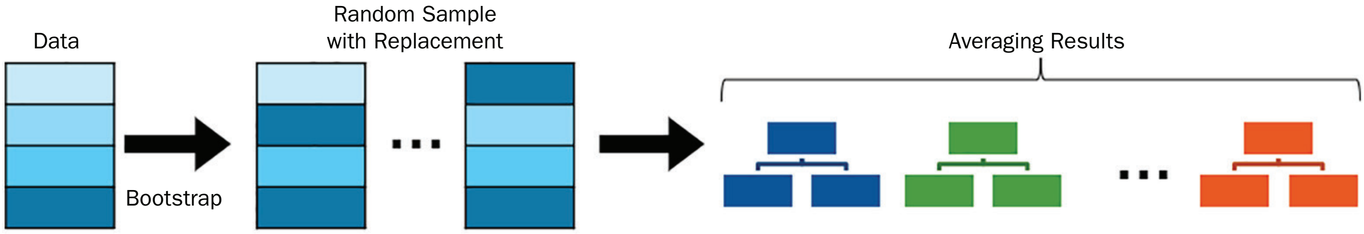
SOURCE: Franklin Templeton Investments.

when a model works well on the training dataset but fails to predict well on the testing dataset. This is a common issue, especially with nonparametric and nonlinear models. Comparable performance across both in-sample and out-of-sample data helps determine if a model is trained properly and performing well.

POOL-LEVEL MODEL DEEP DIVE

The ML prepayment model uses RF for pool-level modeling. We use SMM as the dependent variable given that it is an intuitive way to measure the prepayment rate

EXHIBIT 9
Illustration of the RF Modeling Process



SOURCE: Franklin Templeton Investments.

of an MBS. SMM is simply the principal prepayments over the outstanding balance for the month. Equation 6 shows the formula for SMM and how it is related to CPR.

$$SMM = \frac{Actual\ \ principal\ \ payments - Scheduled\ \ principal\ \ payments}{Beginning\ \ mortgage\ \ balance - Scheduled\ \ principal\ \ payment}$$
$$= \frac{Prepayment}{Oustanding\ \ balance} = 1 - (1 - CPR)^{\frac{1}{12}} \tag{6}$$

Similar to LightGBM, RF is based on a decision tree algorithm. It can handle large datasets efficiently, can provide higher accuracy over a single decision tree, and can be applied to both regression and classification problems. Given that SMM is not binary, RF regression is used for the model. The ensemble method used here to aggregate results is different from LightGBM. RF uses bagging, also known as bootstrap aggregation, to reduce variance and improve prediction accuracy. Bagging is implemented by selecting a number of random samples of data with replacement and averaging the predictions by all the weak models trained on those sample data. A single weak model may not perform well; by averaging results together, RF shows better performance. Compared with boosting, bagging builds weak models in parallel instead of sequentially. Exhibit 9 is an illustration of the modeling process.

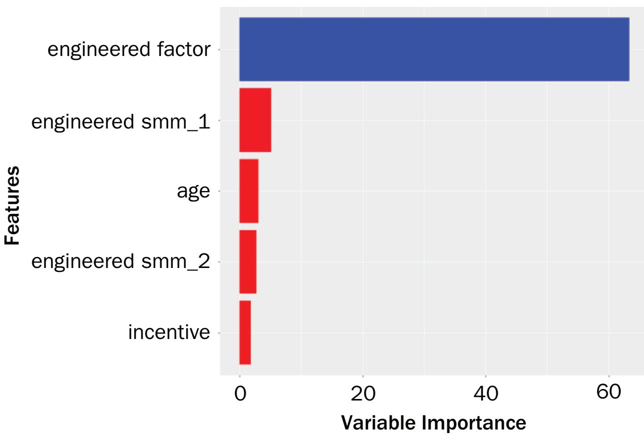
Some engineered factors from the loan-level model results are used in the pool-level model as well. Even though an ML model has less requirement on defining function form for independent variables, feature engineering remains a key component in the model development process. Exhibit 10 shows the feature importance ranking, and we can see that the engineered factor from the loan-level model, SMM, weighted average loan age for the pool, and incentive are important features of the model, with the engineered factor being the most dominant.

For the pool-level model performance, Exhibit 11 shows model-predicted SMMs track actuals well. Exhibit 12 shows actual versus predicted SMM for both linear regression and RF models. The linear model is used for benchmarking purposes. By incorporating loan-level model information, we can see that even the linear regression model has an R^2 of 0.71. With predictions on the *x*-axis and actual on the *y*-axis, however, we can see clearly that the linear model consistently underpredicted. RF model predictions are more in line with actuals and the R^2 also is higher at 0.79.

Normally, the linear model has the advantage of interpretability, but here we show that the RF model can have comparable interpretability. Each independent variable in the linear model has a coefficient, and we can understand easily how the dependent variable will change when a particular independent variable changes by a certain amount.

EXHIBIT 10

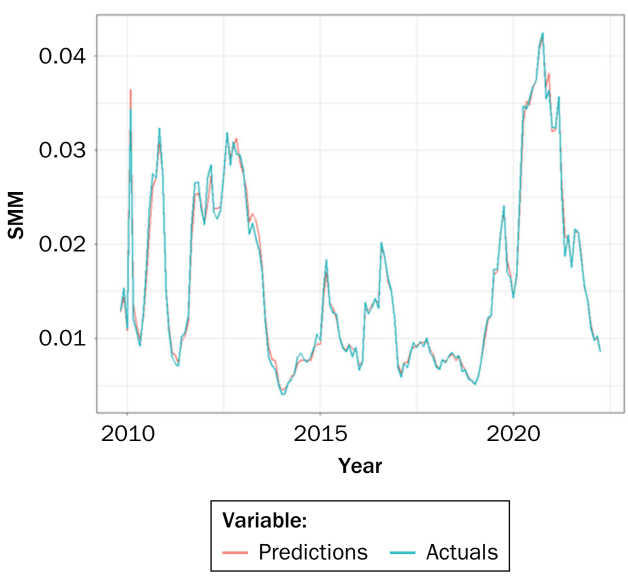
Feature Importance Ranking for Pool-Level Model



SOURCE: Franklin Templeton Investments.

EXHIBIT 11

Model-Predicted SMM vs. Actual SMM Over the Years



SOURCE: Franklin Templeton Investments.

ML models like RF have a different model structure, so they might be hard to interpret from a coefficient or weight perspective. Nevertheless, the idea is to show the marginal effect an independent variable has on the predicted values. Partial dependent plot (PDP) can present similar explainability. PDP is model agnostic, meaning it can be applied to not only the RF model but also to literally any model including ML and traditional models. Equation 7 shows the mathematical definition of partial dependence.

$$Partial\ Dependence_{X_s}(x_S) \stackrel{\text{def}}{=} \mathbb{E}_{X_c}[f(x_S, X_c)]$$
$$= \int f(x_S, x_c)p(x_c)dx_c$$

where X_S = set of input features
x_S = features in X_S
X_C = complement of X_S
x_C = features in X_C
$f(x_S, X_c)$ = model predicting function (7)

PDP is calculated by fixing a specified range for the variable of interest, and then for each value in the range, predicting based on that value and all other feature values. All predictions generated for each value in the range are averaged to form a curve. Exhibit 13 shows the PDP for the engineered factor. The tick marks on the *x*-axis indicate the minimum, maximum, and deciles of the independent variable's distribution, so the majority of the data actually lies in the bottom left corner. When the engineered factor increases, the general trend for the predicted SMM is to increase; however, it decreases on the lower end until a local inflection point is reached. We can understand clearly how the dependent variable is impacted by the independent variable, and it is not linear to say the least.

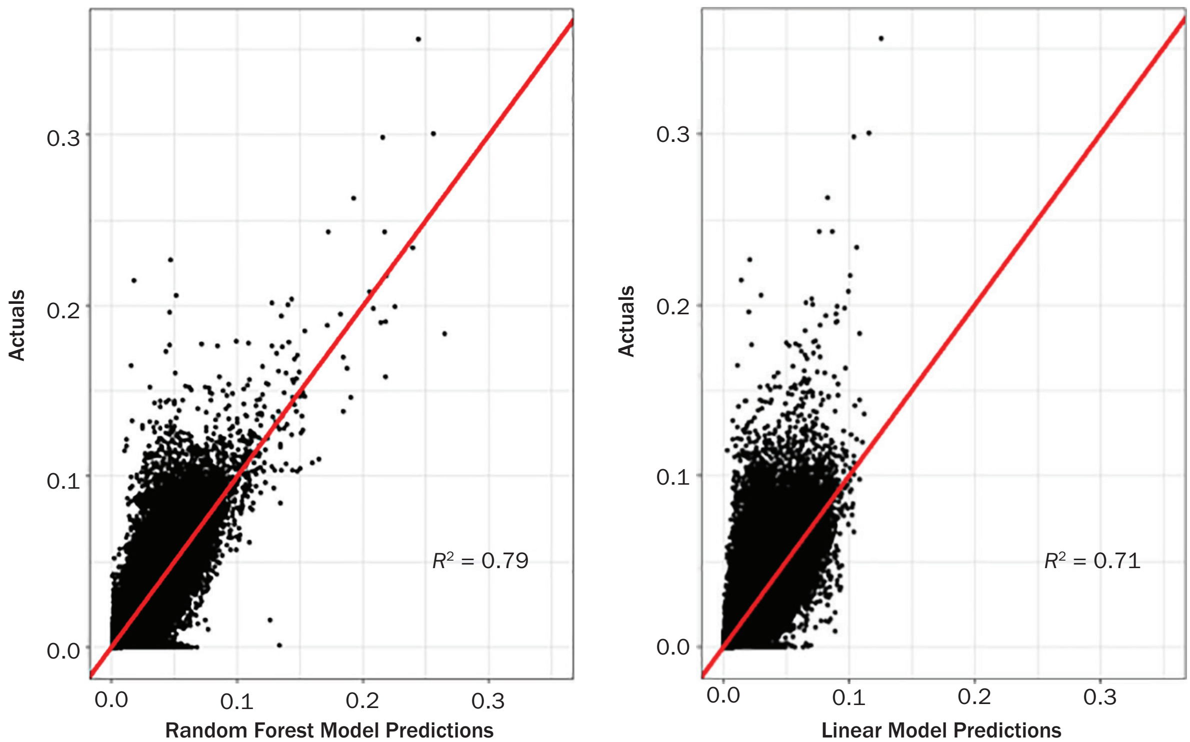
MODEL PERFORMANCE AND COMPARISON VS. STANDARD INDUSTRY MODEL

In terms of accuracy, the ML prepayment model has done well versus third-party models. In the following, we give a comparison of one-month (live/out-of-sample) accuracy versus forecasts from a vendor model, on a large sample of liquid FNMA 30-year MBS pools from July 2021 to March 2022.

We report accuracies on a sample of specified pools rather than for the entire universe because we are interested in evaluating pools that are actionable—ones that can be bought and sold in the current market. The pools evaluated in Exhibit 14 are representative of liquid MBS pass-throughs. Grid columns indicate coupon (2.0% through 4.5% coupon FNMA MBS in 50-bp increments), and N indicates the number of pool observations evaluated in comparison. In the exhibit, the ML model CPR projection is represented by the red line, the third-party model is the green line, and the actual realized CPR is the blue line. As we can see from Exhibit 14, the vendor model consistently underpredicted prepayment speeds versus the actual over the

EXHIBIT 12

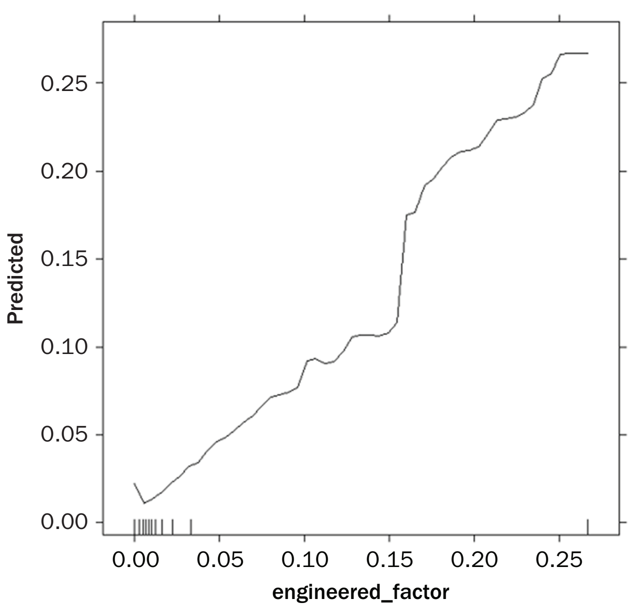
Predicted SMMs vs. Actuals for both RF and Linear Models



SOURCE: Franklin Templeton Investments.

EXHIBIT 13

Partial Dependence Plot for Engineered Factor from Loan-Level Model



SOURCE: Franklin Templeton Investments.

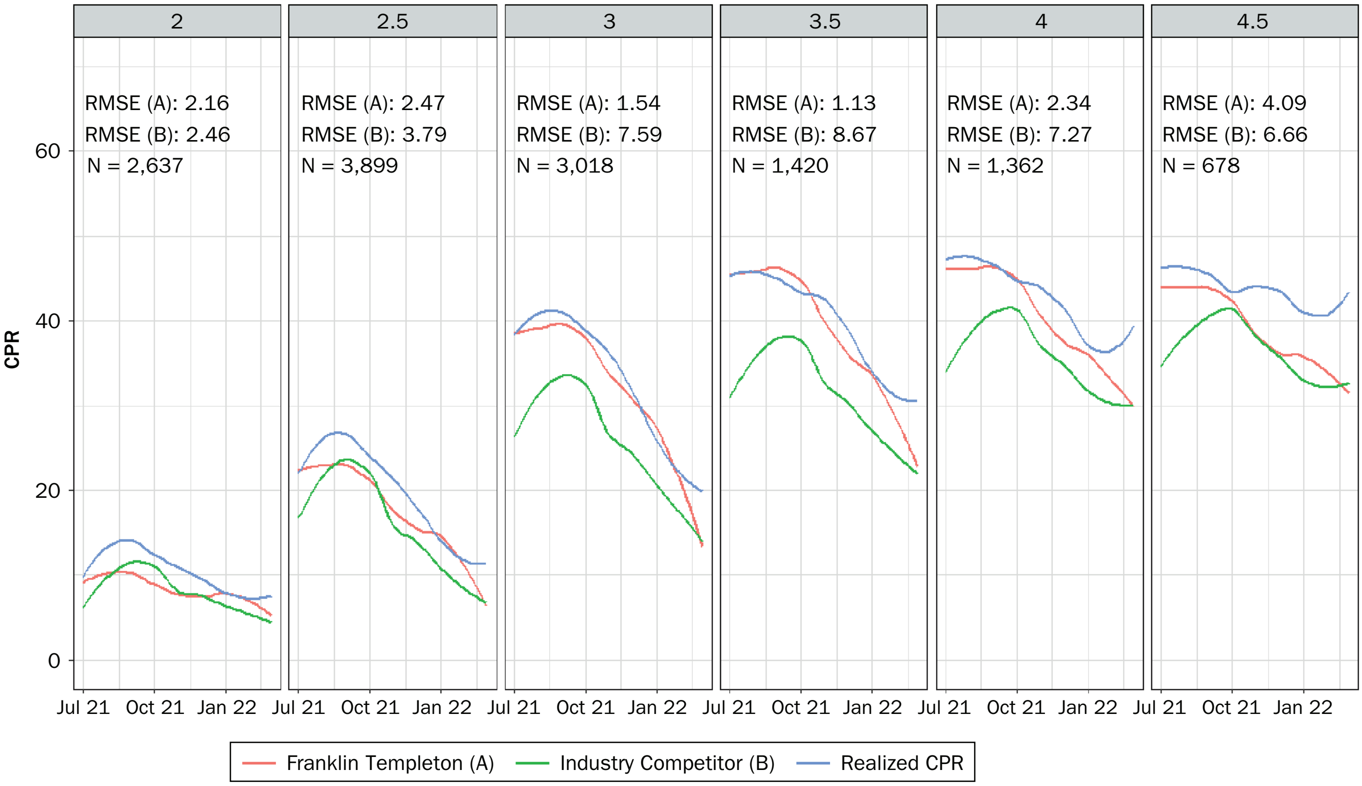
time horizon. The comparison highlights a few startling divergences. Large differences between the models were visible in the fall of 2021, when the third-party model predicted steeper one-month declines than the ML model as well as the observed CPRs in 3.0%, 3.5%, and 4% coupon prepayment speeds.

RMSE is one of the most common metrics for evaluating model performance. The basic idea is to measure the accuracy of the model's predictions compared with actual observed values, in this case realized CPR. Because the lower RMSE, the better (prediction closer to the actual), the ML prepayment model outperforms the vendor model. Over the period of July 2021–March 2022, the ML prepayment model achieved an RMSE of 2.2% CPR on 13,014 pool observations, and the vendor model achieved an RMSE of 4.0% CPR on the same set of pool observations.

The actual prepayment speeds decreased in the first quarter of 2022. It is important to note that in faster prepayment regimes, the differentials in RMSE can increase to an even greater extent. Over the period of August 2020–April 2021, the ML prepayment model achieved an RMSE of 2.8% CPR on 12,058 pool observations, and the vendor model achieved an RMSE of 7.4% CPR on the same set of pool observations.

EXHIBIT 14

Accuracy of 1M Franklin Templeton Model vs. Vendor Model



SOURCE: Franklin Templeton Investments.

EXHIBIT 15

Excess Performance from December 2020 to March 2022 (bps)

Time Period	ML-Based Specified Pool Optimizer
1 Year	19
Since December 10, 2020	30

SOURCE: Franklin Templeton Investments.

SPECIFIED POOL OPTIMIZER AND INVESTMENT STRATEGY PERFORMANCE

Based on CPRs generated by the ML prepayment model, we generate expected returns over a one-month horizon at a constant spread. The specified pool optimizer (SPO) suggests an optimal portfolio while maintaining a similar duration and convexity to the US MBS Index. The SPO can be configured to include, but is not limited to, aggressive/conservative CPR outlook, convexity constraint, outstanding face value lower bound, bid-ask spread, and a number of factors. The portfolio is rebalanced monthly based on the optimal portfolio suggested by SPO.

Since December 2020, the ML-based systematic investment strategy has been used to update a $100 million portfolio. Over this period of little more than one year, the ML-based strategy has yielded excess returns over the index of 30 bps exclusively through specified pools (Exhibit 15).

LOOKING AHEAD

Given the rate sensitivity of prepayments, MBS excess returns over Treasury are negatively correlated with interest rate volatility. The systematic strategy is using a short-term prepayment model assuming muted interest rate volatility impact, so it tends to underperform under extremely volatile market conditions. To further improve the return from the systematic strategy, we could take various volatility-related factors and regimes into consideration while configuring the optimal portfolio. Fundamentally, a separate model for predicting interest rates would benefit the ML prepayment model.

CONCLUSION

Agency MBS is a data-rich sector with a long history of quantitative modeling and is one of the most well understood, which makes it one of the best suited fixed-income sectors for ML-based systematic investing.

By combining loan-level and pool-level models, the ML prepayment model appears to provide more accurate and less biased short-term forecasts than standard industry models. Equipped with the ability to capture the nuances of pool characteristics, it provides the most value in the specified pool market. The addition of SHAP values and partial dependence plots aids us in understanding relationships between mortgage prepayment predictions and related factors. An ML prepayment model with more transparency and interpretability creates a feedback loop to constantly improve the model accuracy and investment performance.

Although model refinements are underway, so far, the ML systematic strategy has outperformed the Bloomberg US MBS Index and has pioneered a new way to invest in the MBS market.

ACKNOWLEDGMENT

We thank Theodore Vadpey, formerly a research analyst/data scientist at Franklin Templeton in San Mateo, for his exceptional work on ML prepayment modeling.

We thank Paul Varunok, a Senior Vice President, Director of Securitized Research and Trading/Portfolio Manager at Franklin Templeton in New York, for his continuous support, market expertise and making this endeavor possible.

REFERENCES

Killian, C. "TBA Market Fact Sheet." The Securities Industry and Financial Markets Association (SIFMA), 2015.

Schultz, G. M., and F. J. Fabozzi. "Rise of the Machines: Application of Machine Learning to Mortgage Prepayment Modeling." New York, N.Y.: Pageant Media Ltd, 2021.

Waldman, M. *The Salomon Brothers Prepayment Model: Impact of the Market Rally on Mortgage Prepayments and Yields.* New York, N.Y.: Salomon Brothers, Inc., 1985.

Disclaimer

All investments involve risk, including loss of principal.

Any information, statement or opinion set forth herein is general in nature, is not directed to or based on the financial situation or needs of any particular investor, and does not constitute, and should not be construed as, investment advice, forecast of future events, a guarantee of future results, or a recommendation with respect to any particular security or investment strategy. Investors seeking financial advice regarding the appropriateness of investing in any securities or investment strategies should consult their financial professional.

Franklin Templeton, Inc., its affiliates, and its employees are not in the business of providing tax or legal advice to taxpayers. These materials and any tax-related statements are not intended or written to be used, and cannot be used or relied upon by any such taxpayer for the purpose of avoiding tax penalties or complying with any applicable tax laws or regulations. Tax-related statements, if any, may have been written in connection with the "promotion or marketing" of the transaction(s) or matter(s) addressed by these materials to the extent allowed by applicable law. Any such taxpayer should seek advice based on the taxpayer's particular circumstances from an independent tax professional.