Welcome
We will begin promptly at 10 AM ET.
If you are unable to hear the speakers, please let us know in the chat box. You may enter your questions in the Q&A, we will address them at the end of the presentation. You can find a copy of the slide deck and recording of this webinar: www.fdpinstitute.org/webinars
Financial Data Professional Institute

FDP Institute provides world class training and education to financial professionals to meet the accelerating needs of digital transformation in the industry.
Introductions

Today’s Topic:

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

August, 2023

Presented by: Nikhil Jagannathan
Portfolio Manager
Franklin Templeton Fixed Income

Leo Bao
Research Analyst
Franklin Templeton Fixed Income
Agenda

• Introduction and Machine Learning Overview

• Fixed Income Market and Agency Mortgage-Backed Securities (MBS)

• Why Machine Learning (ML) in Agency MBS

• Tree-Based Algorithms

• Model Construction

• Model Performance

• Systematic Investment Strategy Application

• Q&A
Traditionally computer programs combined with human created rules can **produce answers to a problem**. Instead, machine learning uses data and answers to **discover the rules behind a problem**.

**Terminology**
- Dataset
- Features
- Model

**Process**
- Data Collection
- Data Preparation
- Training
- Evaluation
- Tuning

Source: Abdul Rahid (WordStream)
Introduction and Machine Learning Overview
Supervised Learning vs Unsupervised Learning

Based on the nature of the learning signal available ML can be roughly categorized as Supervised Learning, Unsupervised Learning and Reinforcement Learning. If based on the output, ML can also be categorized as Classification, Regression and Clustering.

**Supervised Learning**
- need labeled data.

**Minimization**: minimize the fitting error of a model to the presented data

**Unsupervised Learning**
- do not need labeled data.

**Maximization**: identifies the suitable decisions to be made to maximize a well-defined objective

**Reinforcement Learning**
- Trial and error based learning most similar to human learning process.

Source: Franklin Templeton Research
### Introduction and Machine Learning Overview

#### Traditional Linear Model vs ML Model

<table>
<thead>
<tr>
<th></th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Linear Model</td>
<td>• Transparent</td>
<td>• Need to define function form specifically to capture non-linearity</td>
</tr>
<tr>
<td></td>
<td>• Easily interpretable</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Computationally tractable</td>
<td></td>
</tr>
<tr>
<td>Machine Learning Model</td>
<td>• Higher accuracy</td>
<td>• Need assistance to visualize and interpret</td>
</tr>
<tr>
<td></td>
<td>• Better capability on learning complex relationships</td>
<td>• Computational intensive</td>
</tr>
</tbody>
</table>

#### Graphical Representation

```
<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Explainability</th>
</tr>
</thead>
<tbody>
<tr>
<td>★ Deep Learning</td>
<td>★ Linear Models</td>
</tr>
<tr>
<td>★ Random Forests</td>
<td>★ Generalized Additive Models</td>
</tr>
<tr>
<td>★ Decision Trees</td>
<td></td>
</tr>
</tbody>
</table>
```

**Source:** Franklin Templeton Research
Introduction and Machine Learning Overview
Traditional Linear Model vs ML Model

Neural Network

- Best for situations where the data is high-dimensional.

Source: Junaid Qadir

Source: Investopedia
Mortgage-Backed Securities constitute the second largest fixed-income sector (behind Treasuries) in the US market.

Source: Securities Industry and Financial Markets Association (SIFMA) Research as of 12/31/2022
What are Mortgage-Backed Securities (MBS)?

- A mortgage-backed security (MBS) is a fixed income security representing an ownership interest in a pool of residential mortgage loans.

- A residential mortgage is a loan issued by an originator, such as a bank, to a borrower for the purpose of purchasing a residential property.

- Residential homeowners make mortgage payments which are pooled each month, and “passed through” to MBS holders in the form of principal and interest cash flows.
How are agency MBS created?

- To create MBS, a lending bank or originator pools together a group of mortgage loans that it has issued.
- The originator will then present the mortgage loans to a government-sponsored enterprise (GSE) designated to issue and guarantee the MBS.
- The enterprise securitizes pools across various issuers or originators into a pass-through MBS.
- MBS are sold in the global capital market to investors worldwide.

Source: Franklin Templeton Research
Fixed Income Market and Agency Mortgage-Backed Securities

Agency Mortgage-Backed Securities

Prepayment Risk

• The primary risk in the agency MBS market is prepayment risk, the risk of uncertain cash flows caused by prepayments in the pools' underlying mortgages as most residential mortgages originated in the U.S. provide the borrower with the option to prepay some or all of their outstanding principal balance on their loans at any time.

Conditional Prepayment Rate (CPR)

• A CPR is an estimate of the percentage of a loan pool's principal that is likely to be paid off prematurely.

Source: eMBS as of 12/31/2022
Fixed Income Market and Agency Mortgage-Backed Securities

Agency Mortgage-Backed Securities

Drivers of prepayments

- Mortgage Rates
- Collateral Characteristics
- Turnover
- Seasonality
- Catastrophes / Natural Disasters
- Industry Exposure
- Servicers
- Government / Legal Reforms

Primary Mortgage Market Survey Rate

Source: Freddie Mac as of 6/30/2023
Why ML in Agency MBS
Data Availability and Scale

Long history and large amount of data
• 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.

Magnitude of Data
• 20 million records for Conventional
• 10 million records for Ginnies
• Over 1 billion data points.
• Long history that goes back to late 2000.

Source: eMBS from 9/2014 – 3/2023
Complex and dynamic relationship
• 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.

Source: Franklin Templeton Research
Cloud computing and parallel Computing

- 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.

Source: Databricks as of 4/8/2022
Tree-Based Algorithms

Decision Tree

Decision tree Demo – Do we want to play baseball outside?

Attribute Selection Measures

• Entropy
• Information Gain

Source: Franklin Templeton Research
**Tree-Based Algorithms**

**Attribute Selection Measures**

**Entropy** – a measure of the randomness in the information being processed. The higher the entropy, the harder it is to draw any conclusions from that information.

\[ E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i \]

<table>
<thead>
<tr>
<th>Play Baseball</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Play Baseball</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td></td>
<td>8</td>
</tr>
</tbody>
</table>

\[
E(\text{Play Baseball}) = -\left(\frac{10}{10 + 6}\right) \log_2 \left(\frac{10}{10 + 6}\right) - \left(\frac{6}{10 + 6}\right) \log_2 \left(\frac{6}{10 + 6}\right)
\]

\[
= -0.625 \log_2 0.625 - 0.375 \log_2 0.375
\]

\[
\approx 0.954
\]

Source: Franklin Templeton Research
Tree-Based Algorithms
Attribute Selection Measures

**Information Gain** – a measure on how well a given feature separates the data according to the target classification.

\[ E(S, X) = \sum_{i \in X} P(i)E(i) \]

<table>
<thead>
<tr>
<th>Weather</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Cloudy</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Rainy</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>6</td>
<td>16</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
E(\text{Play Baseball}, \text{Weather}) &= P(\text{Sunny})E(4,2) + P(\text{Cloudy})E(5,0) + P(\text{Rainy})E(1,4) \\
&\approx \frac{6}{16} \times 0.918 + \frac{5}{16} \times 0 + \frac{5}{16} \times 0.722 \\
&= 0.569
\end{align*}
\]

\[
\begin{align*}
\text{Information Gain}(S, X) &= \text{Entropy}(S) - \text{Entropy}(S, X) \\
&= \text{Entropy}(\text{Before}) - \text{Entropy}(\text{After}) \\
&= E(\text{Play Baseball}) - E(\text{Play Baseball}, \text{Weather}) \\
&= 0.954 - 0.569 \\
&= 0.385
\end{align*}
\]

Source: Franklin Templeton Research
Ensemble Learning Methods: combines multiple algorithms to obtain better predictive performance than the one from a single model.

Source: Scott Fortmann-Roe/ BiasVariance
Model Construction
Ensemble Learning Methods

Ensemble Learning Methods: combines multiple algorithms to obtain better predictive performance than the one from a single model.

- 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.
- Bagging adopts a sequential approach, where the prediction of the current model is transferred to the next one. Each model iteratively focuses attention on the observations that are misclassified by its predecessors.

<table>
<thead>
<tr>
<th></th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
</table>
| Bagging| • Significant effect on reducing variance, especially high-dimensional data  
         • Unbiased estimate of the out-of-bag error | • Computationally intensive               |
| Boosting| • Efficiently reduces bias  
          • Prioritize features that increase overall accuracy, hence reducing computation time | • Not scalable  
          • Computationally intensive |

Source: Franklin Templeton Research
Model Construction

Roadmap

Data

- Macroeconomic: unemployment
- Microeconomic: application level, house value, affordability, permits issued for construction
- Loan level: remaining balance, credit score, loan to value ratio
- Unconventional data like (hit trend for certain key words)

Source: Franklin Templeton Research
Model Construction
Loan Level Modeling

Loan Level Modeling
• 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.
• LightGBM uses gradient-boosting algorithms, which increases its prediction speed and accuracy, particularly with large and complex datasets.

Source: Franklin Templeton Research
Model Construction
Pool Level Modeling

Pool Level Modeling
• Random Forest 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.

Source: Franklin Templeton Research
Model Construction
Feature Importance

Feature-Importance Ranking for Loan-Level Model

Features
- incentive_1
- payment indicator_1
- incentive_2
- loan_age
- incentive_3
- incentive_4
- red_indicator_1
- red_indicator_2
- house price index
- loan balance

Variable Importance
0e+00 5e+07 1e+08

Source: Franklin Templeton Research

Feature Importance Ranking for Pool-Level Model

Features
- engineered factor
- engineered smm_1
- age
- engineered smm_2
- incentive

Variable Importance
0 20 40 60

Source: Franklin Templeton Research
Model Performance
Model Performance Metrics

Single Monthly Mortality (SMM)

\[
\text{SMM} = \frac{\text{Actual principal payments} - \text{Scheduled principal payments}}{\text{Beginning mortgage balance} - \text{Scheduled principal payment}}
\]

\[
\text{Prepayment} = \frac{\text{Prepayment}}{\text{Outstanding balance}}
\]

\[
= 1 - (1 - \text{CPR})^{\frac{1}{12}}
\]

Source: Franklin Templeton Research
Model Performance

Model Performance Metrics

True Positive Rate (TPR) = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}

False Positive Rate (FPR) = \frac{\text{False Positive}}{\text{True Negative} + \text{False Positive}}

Simple Example of the Calculation of TPR and FPR

<table>
<thead>
<tr>
<th>Probability</th>
<th>Actual</th>
<th>Prediction (threshold = 0.5)</th>
<th>Prediction (threshold = 0.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Threshold = 0.5

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

Threshold = 0.8

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

Source: Franklin Templeton Research
Model Performance
Model Performance Metrics

Model Performance for Loan-Level UMBS 30-Year Model (Receiver Operating Characteristic (ROC) on train data)

ROC Curve (AUC: 0.76)

Model Performance for Loan-Level UMBS 30-Year Model (Receiver Operating Characteristic (ROC) on test data)

ROC Curve (AUC: 0.76)

Source: Franklin Templeton Research
Make sure the model has the right amount of complexity, so it generalizes well on unseen data.

Source: Medium/Bias-Variance Tradeoff Explained as of 6/25/2020
Model Performance
Model Performance Metrics

Coefficient of determination ($R^2$) – measures proportion of variance explained by features.
Root Mean Square Error (RMSE) – measures average deviation between the predicted and actual

Source: Franklin Templeton Research
Partial Dependence (PD) 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.

\[
\text{Partial Dependence}_{x_S}(x_S) \overset{\text{def}}{=} \mathbb{E}_{X_C}[f(x_S, X_C)] \\
= \int f(x_S, x_C)p(x_C)dx_C
\]

where \(X_S = \text{set of input features}\)
\(x_S = \text{features in } X_S\)
\(X_C = \text{complement of } X_S\)
\(x_C = \text{features in } X_C\)
\(f(x_S, X_C) = \text{model predicting function}\)

Visualizing the average effect of a particular feature by marginalizing all other features.

Source: Franklin Templeton Research
## Model Root Mean Square Error (RMSE)

### Comparison with Other Models

#### Model RMSE Overview

<table>
<thead>
<tr>
<th>Source: Franklin Templeton Research as of 6/30/2023</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th>FNCL</th>
<th>FNCL</th>
<th>G2SF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RMSE (A):</strong></td>
<td>0.28</td>
<td>0.62</td>
<td>1.04</td>
</tr>
<tr>
<td><strong>RMSE (B):</strong></td>
<td>0.65</td>
<td>2.44</td>
<td>2.27</td>
</tr>
<tr>
<td><strong>N:</strong></td>
<td>12,485</td>
<td>63,109</td>
<td>6,528</td>
</tr>
</tbody>
</table>

### Conditional Prepayment Rate

- **1m**
- **3m**
- **6m**

**Franklin Templeton (A)**
- **Industry competitor (B)**
- **Realized CPR**
Model Root Mean Square Error (RMSE)
Comparison with Other Models

Closer look on 1M horizon

Liquid FNCL pools, 1M forecast, July 2021- April 2023

RMSE (A): 0.62
RMSE (B): 2.44
N=63,109

Source: Franklin Templeton Research
Model Root Mean Square Error (RMSE)
Comparison with Other Models

RMSE by Coupon

Source: Franklin Templeton Research June 2021 to June 2023
Model Performance
Model Surveillance & Enhancements

Current Production Model

Latest Model Update
- Infrastructure update to reduce training time
- Methodology update to reflect regime changes

Model Update

Model Surveillance

Source: Franklin Templeton Research
Systematic Investment Strategy Application

Specified Pool Optimizer

Based on CPRs generated by the ML prepayment model, we generate expected returns over a one-month horizon at a constant spread. The systematic Investment Strategy

• adopts CPRs from the ML prepayment model
• maintains same duration and convexity to the US MBS Index
• rebalances monthly
• can be configured to incorporate/adjust
  – aggressive/conservative CPR outlook
  – bid-ask spread
  – duration and convexity
  – inclusion/exclusion of certain pool characteristics

For illustrative purposes only, data does not represent actual values for any investment.

Source: Franklin Templeton Research
Important information

This presentation reflects the analysis and opinions of the speaker as of August 2023 and may differ from the opinions of other portfolio managers, investment teams or platforms at Franklin Templeton. Because market and economic conditions are subject to rapid change, the analysis and opinions provided may change without notice. The commentary does not provide a complete analysis of every material fact regarding any country, market, industry or security. An assessment of a particular country, market, security, investment or strategy is not intended as an investment recommendation, nor does it constitute investment advice. Statements of fact are from sources considered reliable, but no representation or warranty is made as to their completeness or accuracy.

The holdings and characteristics of an actual account managed by Franklin Templeton may vary materially from a prospective client’s current holdings. This information is provided for discussion purposes only and does not represent investment advice of any kind or securities recommendations and is subject to change without notice. It should not be assumed that investments in the securities identified were or will be profitable. Certain information has been provided by third party sources which we believe to be reliable, but not guaranteed for accuracy or completeness.

The information contained in this piece is not a complete analysis of every material fact regarding the market and any industry sector, a security or a portfolio. Statements of fact cited by the manager have been obtained from sources considered reliable, but no representation is made as to the completeness or accuracy. Because market and economic conditions are subject to rapid change, opinions provided are valid as of the date of the material.

Portfolio holdings and the manager’s analysis of these issuers, market sectors, and of the economic environment may have changed since the date of the material. The manager’s opinions are intended solely to provide insight into how the manager analyzes securities and are not a recommendation or individual investment advice for any particular security, strategy, or investment product.

This presentation contains certain performance and statistical information. There is no assurance that employment of any of the strategies will result in the intended targets being achieved. The results do not represent actual results and actual results may significantly differ from actual client accounts. Investors should review their investment objectives risk tolerance and liquidity needs before choosing a manager. There is no guarantee that investment strategies will work under all market conditions and investors should evaluate their ability to invest for the long-term, especially during periods of market downturns.

What are the risks?
All investments are subject to certain risks, including possible loss of principal. Generally, investments offering the potential for higher returns are accompanied by a higher degree of risk. Bond prices generally move in the opposite direction of interest rates. Thus, as prices of bonds in a portfolio adjust to a rise in interest rates, the portfolio’s value may decline. Changes in the credit rating of a bond, or in the credit rating or financial strength of a bond’s issuer, insurer or guarantor, may affect the bond’s value. High-yield, lower-rated (junk) bonds generally have greater price swings and higher default risks. Capital gains, if any, are taxable.

Data from third party sources may have been used in the preparation of this material and Franklin Templeton has not independently verified, validated or audited such data. Franklin Templeton accepts no liability whatsoever for any loss arising from use of this information and reliance upon the comment's opinions and analyses in the material is at the sole discretion of the user. Important data provider notices and terms available at www.franklintempletondatasources.com.

Indexes are unmanaged and one cannot invest directly in an index. They do not reflect any fees, expenses or sales charges.

© 2023 Franklin Templeton. All rights reserved.
Q&A

Contact Us:

🌐 fdpinstitute.org
✉️ info@fdpinstitute.org
🐦 @FDPbyCAIA
LinkedIn: linkedin.com/company/FDP Institute