Welcome
We will begin promptly at 11 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.
The FDP Institute

The FDP Charter

The FDP Institute offers a self-study program that provides financial data professionals an efficient path to learn the essential aspects of financial data science.

THE GLOBAL DESIGNATION FOR FINANCE PROFESSIONALS IN A DATA-DRIVEN INDUSTRY

Financial Data Professional Institute™ was created to

✓ Provide financial professionals with the knowledge necessary to succeed in an industry disrupted by the advent of big data and machine learning.
✓ Advocate for the highest levels of professional ethics and standards.
✓ Establish the FDP Charter as a global professional designation in the area of financial data science.

ONE COMPREHENSIVE EXAM
Offered twice per year. (Q2 & Q4)

TWO ONLINE CLASSES
Choose either Python or R, www.fdpinstitute.org/onlineclasses

EARN YOUR FDP DESIGNATION
A globally-recognized charter is awarded to FDP Charter holders.

VALUE ADD
Employers increasingly seek to find professionals who have the skills to apply data science tools to solve their most challenging problems.
Introductions

Kathryn Wilkens, Ph.D., CAIA
Founder, Pearl Quest LLC

Andrew Li
Vice President, Quantitative Researcher
State Street Associates

Today's Topic:
Beyond the Black Box:
Applying Machine Learning to Equity Investing
Beyond the Black Box
Applying Machine Learning to Equity Investing

December 2022

Andrew Li
State Street Associates
State Street Associates

Part of State Street Global Markets, State Street Associates is a partnership with leading academics, focused on bridging the worlds of financial theory and practice.

Big data
A focus on robust, differentiated data sets

Academic partnerships
Active partnerships with leading academics across the country

Research expertise
More than 20 years of investment research and thought leadership

SSA's office near the campus of Harvard University
Investable and Interpretable Machine Learning for Equities

Applying Machine Learning to Investing

Key Challenges: Trust and Transparency
Simple or Complex?

Human brain has 86 billion neurons, but a person can easily express a simple thought. Complicated models can find simple and interpretable relationships.
What Advantages Can Machine Learning Provide?

Information from alternative data
What advantages can machine learning provide?

Nonlinear relationships

Interaction effects
Research Motivations

**Investable**
Focus on the most liquid and accessible market

**Interesting**
Portfolios can outperform relevant benchmarks

**Interpretable**
Probe for intuition of machine learning model behavior
Investable
Interesting
Interpretable
Test Setup

Focus on the most liquid and accessible market

- Universe: S&P 500
- Prediction Target (Y): Stock total return
- Prediction Input (X): Stock and regime indicators
- Models: Neural Networks, Boosted Trees, Random Forest, Lasso, and OLS
- Rebalance frequency: Monthly
## Model Inputs (X)

<table>
<thead>
<tr>
<th>Stock Level Indicators</th>
<th>Regime Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Company Value</strong></td>
<td><strong>Turbulence</strong></td>
</tr>
<tr>
<td>• Size</td>
<td>• Market unusualness</td>
</tr>
<tr>
<td>• Value</td>
<td></td>
</tr>
<tr>
<td><strong>Past Price Trends</strong></td>
<td><strong>Recession Likelihood</strong></td>
</tr>
<tr>
<td>• Short term mean reversion</td>
<td>• Economic Indicator</td>
</tr>
<tr>
<td>• Momentum</td>
<td>• Change in level</td>
</tr>
<tr>
<td>• Sector Momentum</td>
<td></td>
</tr>
<tr>
<td>• Long term mean reversion</td>
<td></td>
</tr>
<tr>
<td><strong>Riskiness</strong></td>
<td></td>
</tr>
<tr>
<td>• Volatility</td>
<td></td>
</tr>
<tr>
<td>• Beta</td>
<td></td>
</tr>
<tr>
<td><strong>Return on Equity</strong></td>
<td></td>
</tr>
<tr>
<td>• Leverage</td>
<td></td>
</tr>
<tr>
<td>• Profitability</td>
<td></td>
</tr>
<tr>
<td><strong>Operating Model</strong></td>
<td></td>
</tr>
<tr>
<td>• Investment</td>
<td></td>
</tr>
<tr>
<td>• Dividend yield</td>
<td></td>
</tr>
</tbody>
</table>
Investable
Interesting
Interpretable
Machine Learning Has More Predictive Power

Top – Bottom performance in the testing period (2015.1 – 2019.12)

Chart covers the period January 2015 through December 2019.
Source: State Street Global Markets, DataStream.
Performance of Different Models
Portfolios of top 100 predicted stocks of each model in the testing period (2015.1 – 2019.12)

<table>
<thead>
<tr>
<th>Model</th>
<th>Return</th>
<th>Risk</th>
<th>Return/Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P 500</td>
<td>11.9%</td>
<td>12.7%</td>
<td>0.94</td>
</tr>
<tr>
<td>OLS</td>
<td>10.5%</td>
<td>14.2%</td>
<td>0.74</td>
</tr>
<tr>
<td>Lasso</td>
<td>10.4%</td>
<td>13.4%</td>
<td>0.77</td>
</tr>
<tr>
<td>Random Forest</td>
<td>11.5%</td>
<td>9.8%</td>
<td>1.18</td>
</tr>
<tr>
<td>Boosted Trees</td>
<td>13.0%</td>
<td>11.8%</td>
<td>1.10</td>
</tr>
<tr>
<td>Neural Network</td>
<td>13.4%</td>
<td>11.8%</td>
<td>1.14</td>
</tr>
</tbody>
</table>

Chart covers the period January 2015 through December 2019.
Source: State Street Global Markets, DataStream.
Changing the Prediction Target (Y)

Instead of targeting total return, we can target excess returns over a benchmark

- Estimate CAPM-expected returns for each stock
- Performance evaluations change accordingly

<table>
<thead>
<tr>
<th>Model</th>
<th>Excess Return</th>
<th>Risk</th>
<th>Return/Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>1.9%</td>
<td>5.0%</td>
<td>0.39</td>
</tr>
<tr>
<td>Lasso</td>
<td>3.3%</td>
<td>5.8%</td>
<td>0.57</td>
</tr>
<tr>
<td>Random Forest</td>
<td>4.6%</td>
<td>7.4%</td>
<td>0.62</td>
</tr>
<tr>
<td>Boosted Trees</td>
<td>4.2%</td>
<td>5.7%</td>
<td>0.74</td>
</tr>
<tr>
<td>Neural Network</td>
<td>3.9%</td>
<td>6.1%</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Tailoring prediction goals to prior beliefs and preferences

Return benchmarks and time horizons


Covers the period January 2015 through December 2019.

Source: State Street Global Markets, DataStream.
Investable
Interesting
Interpretable
Coefficient-like Intuition for Any Model

Holding all else constant, how do model predictions change as we change variable $x_1$?
Model Fingerprint: Decomposition of Predictions

- For each value of a chosen variable, generate model predictions combining this value with all other historical inputs, and take the average. Repeat.
- Separate the resulting partial prediction curve into a linear and nonlinear effect.

Chart is illustrative; no live data.
Prediction Fingerprint – Decomposing Predictions

Sub-components Fingerprint
Boosted Trees (prediction target = total return)

Chart based on data that covers the period January 1992 through December 2014.
Source: State Street Global Markets.
### Prediction Fingerprint – Comparing Models

#### Most influential interaction effects

<table>
<thead>
<tr>
<th>Total return, 1-month</th>
<th>Random Forest</th>
<th>Boosted Trees</th>
<th>Neural Network</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top 3 Overall</strong></td>
<td>Beta, Recession (Shift)</td>
<td>Size, Turbulence</td>
<td>Volatility, Turbulence</td>
</tr>
<tr>
<td></td>
<td>Volatility, Recession (Shift)</td>
<td>Value, Turbulence</td>
<td>Size, Turbulence</td>
</tr>
<tr>
<td></td>
<td>Beta, Turbulence</td>
<td>Sector, Turbulence</td>
<td>Momentum, Turbulence</td>
</tr>
<tr>
<td><strong>Top 3 (Without Regime Variables)</strong></td>
<td>Beta, Volatility</td>
<td>Short Reversion, Sector</td>
<td>Momentum, Sector</td>
</tr>
<tr>
<td></td>
<td>Beta, Yield</td>
<td>Momentum, Sector</td>
<td>Long Reversion, Momentum</td>
</tr>
<tr>
<td></td>
<td>Beta, Sector</td>
<td>Value, Sector</td>
<td>Beta, Momentum</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total return, 12-month</th>
<th>Random Forest</th>
<th>Boosted Trees</th>
<th>Neural Network</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top 3 Overall</strong></td>
<td>Value, Sector</td>
<td>Leverage, Sector</td>
<td>Beta, Momentum</td>
</tr>
<tr>
<td></td>
<td>Investment, Value</td>
<td>Sector, Recession</td>
<td>Beta, Yield</td>
</tr>
<tr>
<td></td>
<td>Value, Volatility</td>
<td>Value, Sector</td>
<td>Size, Recession</td>
</tr>
<tr>
<td><strong>Top 3 (Without Regime Variables)</strong></td>
<td>Value, Sector</td>
<td>Leverage, Sector</td>
<td>Beta, Momentum</td>
</tr>
<tr>
<td></td>
<td>Investment, Value</td>
<td>Value, Sector</td>
<td>Beta, Yield</td>
</tr>
<tr>
<td></td>
<td>Value, Volatility</td>
<td>Yield, Sector</td>
<td>Momentum, Sector Momentum</td>
</tr>
</tbody>
</table>
Model Fingerprint: Decomposition of Performance

1. Compute returns of portfolios formed from the linear predictions in isolation.
2. Compute returns of portfolios formed from the linear and pairwise interaction predictions, minus those from step 1. This isolates pairwise interaction effects.
3. Compute returns of portfolios formed from the linear, nonlinear, and pairwise interaction predictions, minus those from steps 1 and 2. This isolates nonlinear (sizing) effects.
4. Compute returns of portfolios formed from the full model predictions, minus those from steps 1, 2, and 3. This isolates higher-order effects.

Cumulative returns

Chart is illustrative; no live data.
Performance Fingerprint – Comparing Models

Charts cover period January 2015 through December 2019.
Source: State Street Global Markets, DataStream.
Hedging by Interactions

Random forest as an example

Source: State Street Global Markets, DataStream.
Summary: Machine Learning for Equities

Predictive

• ML models are shown to outperform linear models, due to their ability to model interactions and nonlinear effects

Flexible

• ML models can be made to be goal oriented, hence helping predict most relevant measures of performance

Interpretable

• A helpful way to approach machine learning is to compare machines to people. Even complicated models can find simple and interpretable relationships.
• Model fingerprint is a model-agnostic statistical framework that decomposes ML predictions into linear, nonlinear, and interaction components.
• The “black box” can be cracked open to understand how a ML model thinks
Please join us for our upcoming webinar:

FDP Info Session
Join FDP Experts to learn about the FDP Charter, achieving exam success, and more.

January 11, 2023 @ 11 AM ET

Thank You

Contact Us:

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