Deep Factor Alpha

Guanhao Feng\textsuperscript{1}  Nicholas Polson \textsuperscript{2}  Jianeng Xu\textsuperscript{2}

\textsuperscript{1}City University of Hong Kong
\textsuperscript{2}Chicago Booth
\textsuperscript{3}Chicago Booth

R/Finance 2018
Research Question

Whether there exists alpha in the cross-section of U.S. equities?

- Linear Factor Model / Factor Investing
- Firm Characteristics (Fundamentals) / Risk Anomalies
- Machine Learning / Neural Network / Artificial Intelligence
\[ R_{i,t} = \alpha_i + \beta_i^T F_t + \epsilon_{i,t} \]
\[ = \beta_i^T F_t + \alpha_i, \]

- \( R_{i,t} \) are excess returns of stock \( i \) at time \( t \)
- \( F_t \) are tradable factors (long-short zero-investment portfolios)
  - For CAPM, \( F_t \) is the excess market return
  - For Fama-French 3 factors, \( F_t \) are MktRf, SMB, and HML
- Our alpha defined in the model needs to be tradable.
The forecasting machine is the portfolio

\[ \hat{R}_{i,t} = \beta_i^T F_t. \]

- We trade the portfolio \( R_{i,t} - \hat{R}_{i,t} \) \( [= \alpha_{i,t}] \) for stocks with positive alphas, and \( \hat{R}_{i,t} - R_{i,t} \) for negatives.
- If market is efficient, \( E(\alpha_{i,t}) = 0 \) for all stock i (GRS test).
- However, the factor model is unknown (APT).

- Our goal is to use deep learning to create the factor model.
Fama-French Factor Models

- Any firm pattern to seek alphas is called an anomaly (size, value, momentum, ...).
- The asset pricing literature keeps discovering risk factors to eliminate these alphas.
  - CAPM
  - Fama-French 3/5 Factor Models
  - Hundreds of factors published (Harvey, Liu, and Zhu, 2016)
- However, still find significant alphas (small caps).
Fama-French five-factor model

- Lag Firm Characteristics
  - Size
  - Book-to-Market
  - Operating Profitability
  - Investment

- Security Sorting
  - Security Returns
  - Long-Short Factors
  - MktRf
  - SMB
  - HML
  - RMW
  - CMA
Tradable Factor

The advantage of factor model is to reduce the dimension of thousands of stocks into a few portfolios, tradable factors.

How does the literature create risk factors?

- Sort Securities on firm characteristics (size, value, momentum, …)
- Create long-short portfolios as tradable factors
- These factors have risk premium itself and help to explain the cross-section.
Why machine learning?

- Can’t apply Machine learning directly like using firm characteristics to forecast cross-sectional stock returns.
  - The imbalanced data structure.
  - Missing value issues.
- However, machine learning is useful for
  - Selecting and test factors (Feng, Giglio, and Xiu, 2017).
  - Generating latent factors (Kelly, Pruitt, and Su, 2018)
  - Extracting nonlinear signals (Freyberger, Neuhierl, and Weber, 2017)
- We build a forecasting machine to implement Fama-French type factor models from square zero - the firm characteristics.
How does the literature create firm characteristics?

The picture from Chicago Booth Review “The 300 secrets to high stock returns” provides an answer.
Example: Which momentum?

There are many momentum factors discovered in the literature.

- Momentum (Carhart, 1997)
- Long-Term Reversal (De Bondt and Thaler, 1985)
- Industry Momentum (Moskowitz and Grinblatt, 1999)
- Short-Term Reversal (Jegadeesh and Titman, 1993)
- 6-month Momentum (Jegadeesh and Titman, 1993)
- 36-month Momentum (Jegadeesh and Titman, 1993)
- Change in 6-month Momentum (Gettleman and Marks, 2006)
- ...
Why deep learning?

- Neural network is helpful to explore the multi-layer nonlinear space.
  - Extract signals using firm trading and accounting information (ME, BE, ROE, dividend, earning, historical returns, ...).
  - Forward propagation and multi-layer transformations.
  - Perform the security sorting within training the neural network.
  - Augmented linear factor model and backward propagation.
- We still rely on academic discovered characteristics, but let the machine to explore all possibilities.
- Our goal is to push the tradable factor generation from human fundamental research to AI.
Deep Generated Momentum

<table>
<thead>
<tr>
<th>Lag Firm Characteristics</th>
<th>Hidden Neuron</th>
<th>Bivariate Sorting with Size</th>
<th>Long-Short Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>return at month (t-1)</td>
<td></td>
<td>Ret(1-1)</td>
<td>STR</td>
</tr>
<tr>
<td>return at month (t-2)</td>
<td></td>
<td>Ret(2-12)</td>
<td>MOM</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>Ret(13-60)</td>
<td>LTR</td>
</tr>
<tr>
<td>return at month (t-60)</td>
<td></td>
<td>Ind. Ret</td>
<td>Ind. M</td>
</tr>
<tr>
<td>Industry Specification</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Deep Learning Objective

The objective is to eliminate the mis-pricings using generated risk factors. Define the tradable alphas as

\[ \alpha_{i,t} = R_{i,t} - \hat{R}_{i,t} \]

The objective is min \( \frac{1}{N} \sum_{i=1}^{N} \bar{\alpha}_i^2 \), where \( \bar{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} \alpha_{i,t} \).

We build a unified deep learning factor model from the firm characteristics to minimizing the mis-pricings.
Augmented Linear Factor Model

\[ R_{i,t} = \alpha_i + \beta_i^T F_t + \gamma G_t + \epsilon_{i,t} \]

- \( F_t \) are latent factors generated from the underlying neural network
- \( G_t \) are other observable tradable factors to add (e.g. MktRf)

The unified deep learning loss function is

\[ \min_{F_t} L = \frac{1}{N} \sum_{i=1}^{N} \bar{\alpha}_i^2 + \text{penalty}(\beta, \gamma). \]
Augmented Linear Factor Model

\[
Z(t-1) \quad Z^{[1]}(t) = g(WZ^{[0]}(t) + b) \quad S(t) \quad F_t = S(t)W(t)r(t) \quad \hat{R}(t) = \beta F_t + \gamma G_t
\]
How TensorFlow helps?

- Security Sorting is an nonlinear activation function in our model.
- TensorFlow is extremely power to let us provide a joint estimation from latent factor generation to minimizing the mis-pricings.
- It takes 10 min to finish training a model (N 3000, T 500, P 70).
- In forecasting, we refit the model using $G_t$ and the generated $F_t$.
- R users can use the R interface to TensorFlow with the high-level Keras and Estimator APIs.
Decomposing the stock universe into six subsets, we use the top and bottom 20% firms to create the factors.
Bivariate Sorted Portfolios

\[ F_t = \left( \frac{P(\text{B-top}) + P(\text{S-top})}{2} \right) - \left( \frac{P(\text{B-bottom}) + P(\text{S-bottom})}{2} \right) \]

- \( F_t \) is a long-short zero-investment portfolio.
- Sort the firms every month using lag month characteristics.
  - FF sort firms annually on each June and fix the portfolio for the incoming year.
- Deal with the micro-cap issue.
- Fine with the imbalanced data structure.
- Solve the missing value issue.
Empirical Study


- We have 62 characteristics used in Feng, Giglio, and Xiu (2017).
- We add 11 industry specifications as dummy characteristics.

Train, validation and test asset design.

- The monthly largest 3000 firms to generate the factors (train).
- 202 sorted portfolios used in Giglio, and Xiu (2016) (validation).
- FF 96 portfolios (test).
Out-of-Sample Prediction

1. Generate factors using individual firms
2. Minimize the mis-pricings for validation assets

Cross-sectional prediction for test assets.
Out-of-Sample Prediction

We use a few measures to report the empirical results.

- **RMSE** = $\sqrt{\frac{1}{N} \sum_{i=1}^{N} \bar{\alpha}_i^2}$ is the standard deviation for the pricing error.

- $R^2 = 1 - \frac{\text{RMSE}_{M_1}^2}{\text{RMSE}_{M_2}^2}$ is the relative performance of Model 1 over Model 2.

- Appraisal ratio measures the model-adjusted return of a portfolio. For $\alpha_t = R_t - \hat{R}_t$, the annualized appraisal ratio is $\sqrt{12} \cdot E(\alpha_t)/sd(\alpha_t)$. 
Out-of-Sample Prediction

- We build a 6-layer 3-factor neural network on the benchmark model.
- For TS prediction, the improvement is marginal over FF models.
- For CS prediction, we see significant improvement over all models.

<table>
<thead>
<tr>
<th></th>
<th>RMSE-TS</th>
<th>R2-TS</th>
<th>RMSE-CS</th>
<th>R2-CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPM+DL</td>
<td>0.305%</td>
<td>4.8%</td>
<td>0.752%</td>
<td>2.6%</td>
</tr>
<tr>
<td>FF3+DL</td>
<td>0.285%</td>
<td>16.6%</td>
<td>0.504%</td>
<td>56.2%</td>
</tr>
<tr>
<td>FF5+DL</td>
<td>0.230%</td>
<td>45.2%</td>
<td>0.406%</td>
<td>71.5%</td>
</tr>
<tr>
<td>CAPM</td>
<td>0.311%</td>
<td>0.0%</td>
<td>0.762%</td>
<td>0.0%</td>
</tr>
<tr>
<td>FF3</td>
<td>0.281%</td>
<td>18.3%</td>
<td>0.513%</td>
<td>54.7%</td>
</tr>
<tr>
<td>FF5</td>
<td>0.232%</td>
<td>44.9%</td>
<td>0.411%</td>
<td>71.0%</td>
</tr>
</tbody>
</table>
Out-of-Sample Prediction

For each benchmark $G_t$, we plot the mis-pricing RMSE after adding $F_t$. The first row 0-factor corresponds to RMSE of the benchmark.
Beyond Fama-French Models

- We provide a nested model comparison to evaluate the deep predictability beyond FF models.
  - We use all FF factors as $G_t$.
  - We only use the corresponding FF characteristics to generate $F_t$.

- The objective is the mis-pricings $\min \frac{1}{N} \sum_{i=1}^{N} \tilde{\alpha}_i^2$, and adding characteristics does not necessarily improve the goodness-of-fit.

- This is a useful framework to evaluate the importance of a new characteristics beyond a benchmark models.
Deep Alpha Factor

- DAF are useful complement to FF3 to forecast future returns.
- DAF are useful to price other assets out of the cross-section.

<table>
<thead>
<tr>
<th></th>
<th>RMSE-TS</th>
<th>R2-TS</th>
<th>RMSE-CS</th>
<th>R2-CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF3+DL2</td>
<td>0.287%</td>
<td>-3.78%</td>
<td>0.510%</td>
<td>1.26%</td>
</tr>
<tr>
<td>FF3+DL4</td>
<td>0.281%</td>
<td>0.63%</td>
<td>0.506%</td>
<td>2.62%</td>
</tr>
<tr>
<td>FF3+DL6</td>
<td>0.274%</td>
<td>5.11%</td>
<td>0.504%</td>
<td>3.38%</td>
</tr>
<tr>
<td>FF5+DL2</td>
<td>0.233%</td>
<td>-1.46%</td>
<td>0.401%</td>
<td>4.45%</td>
</tr>
<tr>
<td>FF5+DL4</td>
<td>0.232%</td>
<td>-0.76%</td>
<td>0.409%</td>
<td>0.53%</td>
</tr>
<tr>
<td>FF5+DL6</td>
<td>0.232%</td>
<td>-0.78%</td>
<td>0.398%</td>
<td>5.80%</td>
</tr>
</tbody>
</table>
Deep Alpha Factor

For the generated $F_t$, we plot their out-of-sample alpha hedged on the corresponding benchmark $G_t$. 
Deep Alpha Factor

For the generated $F_t$, we plot their Appraisal ratios for out-of-sample alpha hedged on the benchmark corresponding $G_t$. 

OOS Annualized Appraisal Ratio –CAPM

OOS Annualized Appraisal Ratio –FF3

OOS Annualized Appraisal Ratio –FF5
Summary

▶ Our model combines deep neural network, characteristics security sorting, and linear factor model.

▶ TensorFlow is powerful to provide a joint estimation to both neural network and augmented linear model.

▶ Our model can be used as a framework to evaluate future new characteristics by controlling for a benchmark model.

▶ We still have a lot to finish the paper. If interested, please do not hesitate to check back with me in a few weeks.