Deep Factor Alpha

Guanhao Feng[1] Nicholas Polson [2] Jianeng Xu[2]

[1]City University of Hong Kong

[2]Chicago Booth

[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

Seeking Alpha

$$
\begin{aligned}
R_{i,t} &= \alpha_i + \beta_i^\intercal F_t + \epsilon_{i,t} \\
&= \beta_i^\intercal F_t + \alpha_{i,t}
\end{aligned}
$$

- $R_{i,t}$ 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.

Seeking Alpha

The forecasting machine is the portfolio

$$\widehat{R}_{i,t} = \beta_i^\intercal F_t.$$

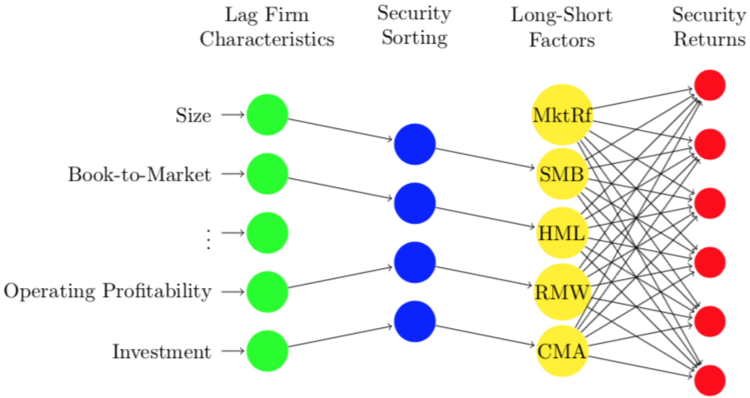
- We trade the portfolio $R_{i,t} - \widehat{R}_{i,t}$ $[= \alpha_{i,t}]$ for stocks with positive alphas, and $\widehat{R}_{i,t} - R_{i,t}$ for negatives.
- If market is efficient, $E(\alpha_{i,t}) = 0$ for all stock i (GRS test).
- However, the factor model is unknown (APT).

- 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



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.

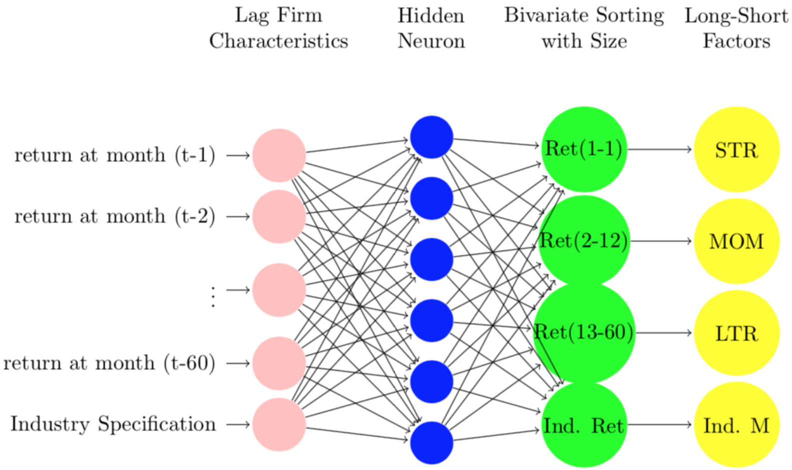
- Momentum (Carhart, 1997)
- Long-Term Reversal (De Bondt and Thaler, 1985)
- Industry Momentum (Moskowitz and Grinblatt, 1999)
- Short-Term Reversal (Jegadeesh and Titman, 1993)
- 6-month Momentum (Jegadeesh and Titman, 1993)
- 36-month Momentum (Jegadeesh and Titman, 1993)
- Change in 6-month Momentum (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



Deep Learning Objective

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

$$\alpha_{i,t} = R_{i,t} - \widehat{R}_{i,t}$$

- 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

$$R_{i,t} = \alpha_i + \beta_i^\intercal F_t + \gamma G_t + \epsilon_{i,t}$$

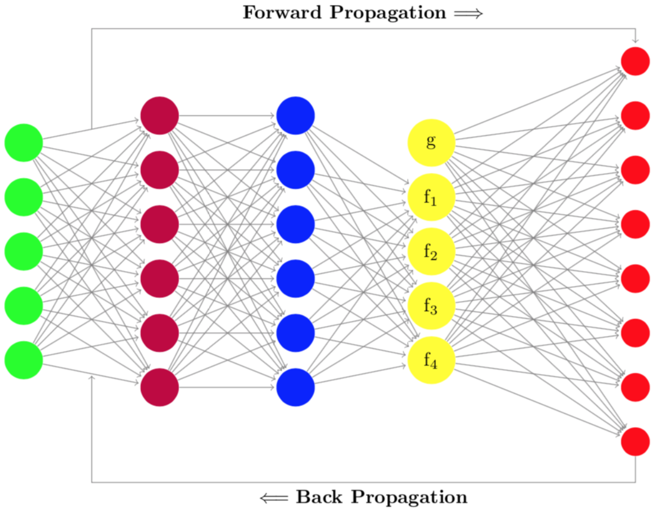
- F_t are latent factors generated from the underlying neural network
- G_t are other observable tradable factors to add (e.g. MktRf)

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

$Z(t-1)$ $\quad Z^{[1]}(t)=g(WZ^{[0]}(t)+b)$ $\quad S(t)$ $\quad F_t=S(t)W(t)r(t)$ $\quad \widehat{R}(t)=\beta F_t+\gamma G_t$

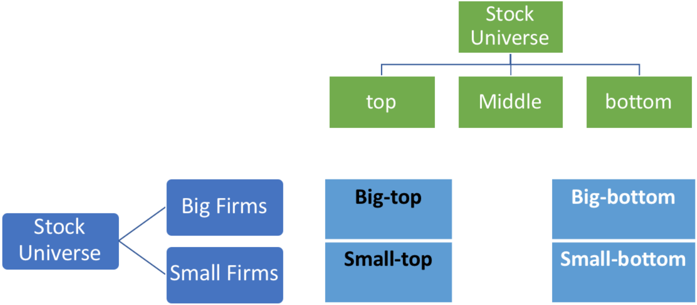


How TensorFlow helps?

- 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.

Bivariate Sorted Portfolios

Decomposing the stock universe into six subsets, we use the top and bottom 20% firms to create the factors.



Bivariate Sorted Portfolios

$$F_t = \Big(\text{P(B-top)} + \text{P(S-top)}\Big)/2 - \Big(\text{P(B-bottom)} + \text{P(S-bottom)}\Big)/2$$

- 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

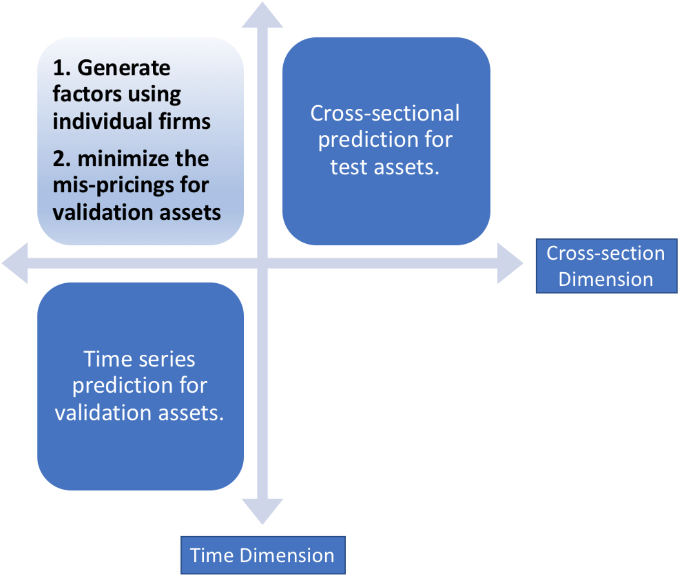
Monthly data from 1975 to 2017 for U.S. equities.

- We have 62 characteristics used in Feng, Giglio, and Xiu (2017).
- We add 11 industry specifications as dummy characteristics.
- We use use train data 1975-2009 and predict 2010-2017.

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



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.
- $R^2 = 1 - \text{RMSE}^2_{M_1}/\text{RMSE}^2_{M_2}$ is the relative performance of Model 1 over Model 2.
- Appraisal ratio measures the model-adjusted return of a portfolio. For $\alpha_t = R_t - \widehat{R_t}$, the annualized appraisal ratio is

$$\sqrt{12} * E(\alpha_t)/sd(\alpha_t).$$

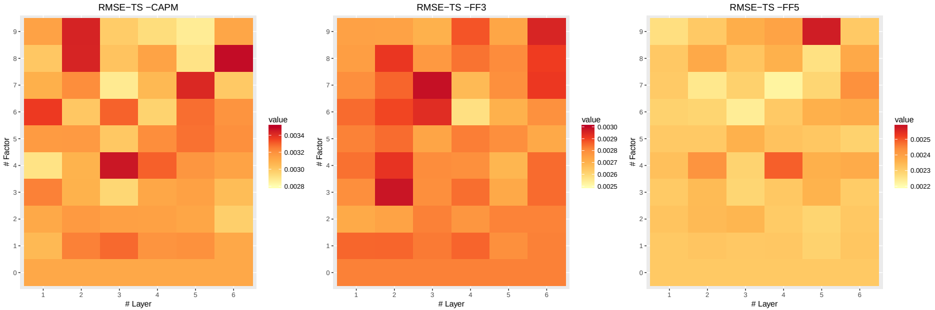
Out-of-Sample Prediction

- 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.

	RMSE-TS	R2-TS	RMSE-CS	R2-CS
CAPM+DL	0.305%	4.8%	0.752%	2.6%
FF3+DL	0.285%	16.6%	0.504%	56.2%
FF5+DL	0.230%	45.2%	0.406%	71.5%
CAPM	0.311%	0.0%	0.762%	0.0%
FF3	0.281%	18.3%	0.513%	54.7%
FF5	0.232%	44.9%	0.411%	71.0%

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 $\min \frac{1}{N}\sum_{i=1}^{N}\bar{\alpha}_i^2$, 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.

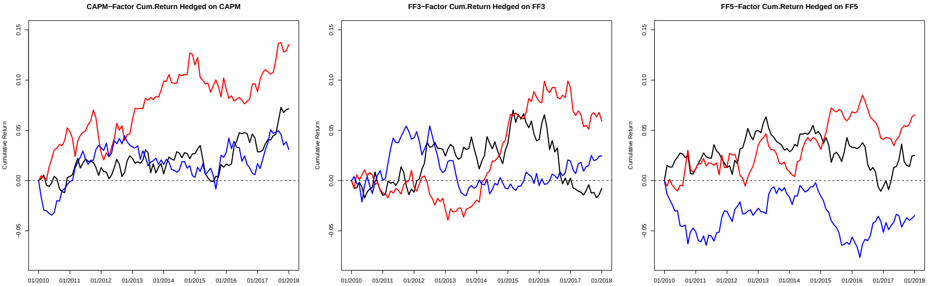
Deep Alpha Factor

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

	RMSE-TS	R2-TS	RMSE-CS	R2-CS
FF3+DL2	0.287%	-3.78%	0.510%	1.26%
FF3+DL4	0.281%	0.63%	0.506%	2.62%
FF3+DL6	0.274%	5.11%	0.504%	3.38%
FF5+DL2	0.233%	-1.46%	0.401%	4.45%
FF5+DL4	0.232%	-0.76%	0.409%	0.53%
FF5+DL6	0.232%	-0.78%	0.398%	5.80%

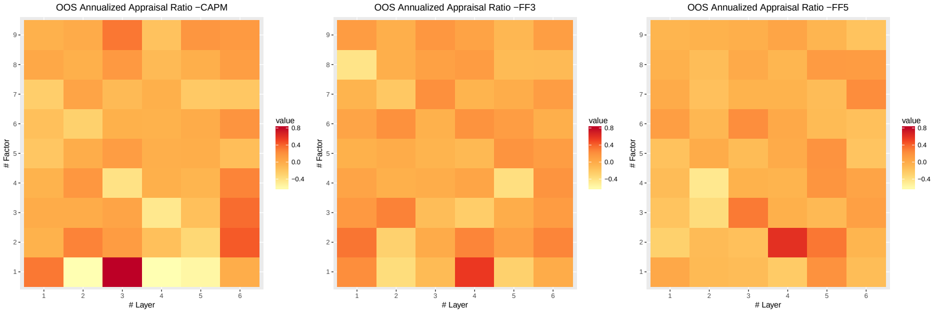
Deep Alpha Factor

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.



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.