Course Project Proposal

MF796

Team Memeber:

Yuxuan Liu, Yanbin Liu, Haoyu Guan, Chang Deng

Introduction:

The most fundamental question in asset pricing is to understand why different assets have different average returns. No-arbitrage pricing theory provides a clear answer - expected returns differ because assets have different exposure to systematic risk.

All pricing information is summarized in the stochastic discount factor (SDF) or pricing kernel. The empirical quest in asset pricing for the last 40 years was to estimate a stochastic discount factor that can explain expected returns of all assets. There are four major challenges that the literature so far has struggled to overcome in a single model:

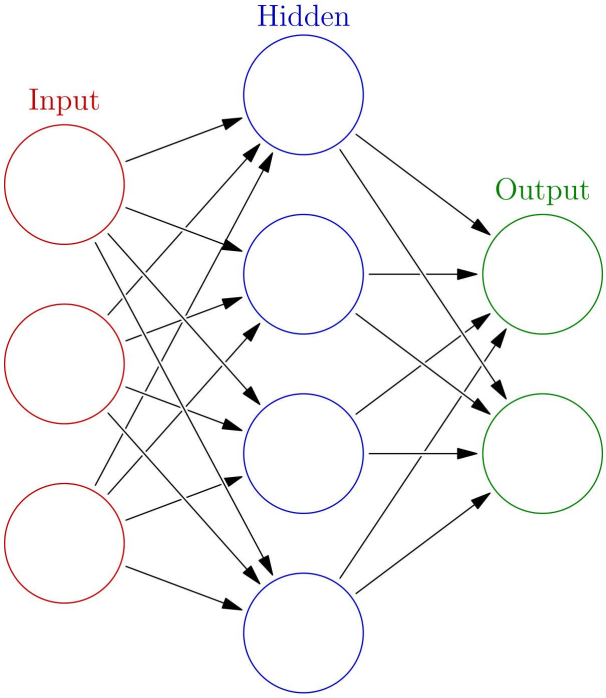
· SDF is function of all available variables

· The exact functional form of SDF is unknown and likely complex

· SDF has a dynamic structure

· risk premium of assets has low signal-to-noise ratio

The main idea to overcome those challenge is to use neural network to estimate SDF.



Neural Network

· component: input, hidden, output

· type: DNN, CNN, RNN, GAN, etc

Related to Provious Challenge

· non linearity - NN

· time-variation - RNN/LSTM

· seperate signal from noise - no-arbitrage constraint GAN

Model:

Close-form solution for price kernel portfolio weights needs to inverse the covariance matrix of all the existing assets, which is unreliable due to low signal-to-noise ratio.

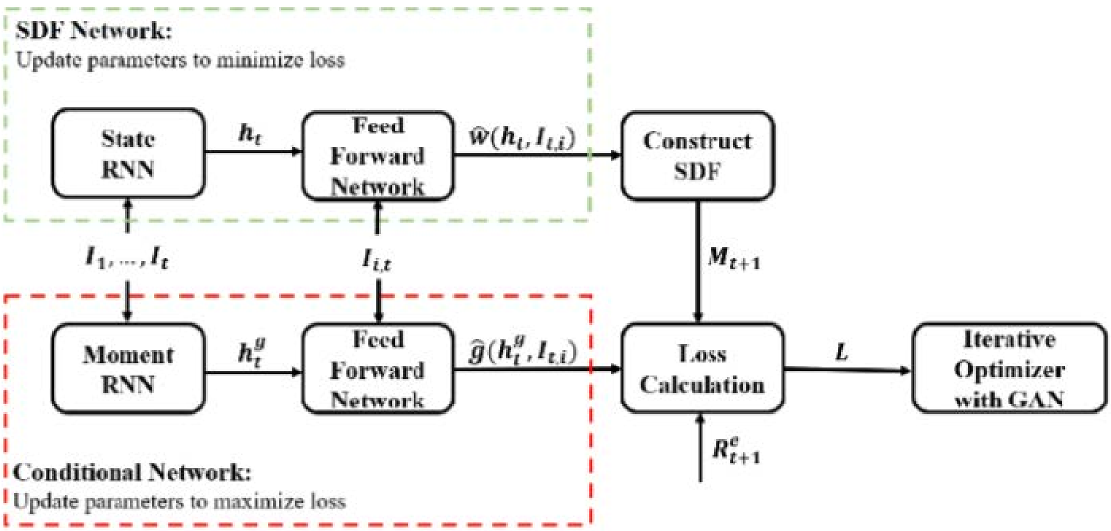
Think about conditional no-arbitrage moment condition:

$$E\left[M_{t+1} R_{t+1,i}^{e} g\left(I_{t}, I_{t,i}\right)\right]=0$$

So the approach that minimizes the violation of no-arbitrage is:

$$\min_{\omega} \max_{g} \frac{1}{N} \sum_{j=1}^{N}\left\|E\left[\left(1-\sum_{t=1}^{N} \omega\left(I_{t}, l_{t,i}\right) R_{t+1,i}^{e}\right) R_{t+1,j}^{e} g\left(l_{t}, l_{t,j}\right)\right]\right\|^{2}$$

To solve the problem, the following architecture is used:



we have the SDF portfolio,

$$F=w^{T} R_{t+1}^{e}$$

and

$$E_{t}\left[R_{t+1,i}^{e}\right]=\beta_{t,i} F_{t+1}+\epsilon_{t+1,i}$$

Evaluation:

Sharpe Ratio of SDF Portfolio:

$$S R=\frac{E[F]}{\sqrt{\operatorname{Var}[F]}}$$

Explained individual Variance:

$$EV = 1 - \frac{\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\epsilon_i^2/N}{\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(R_{t+1,i}^e\right)^2/N}$$

Explained cross-section mean:

$$XS - R^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\epsilon_i/T\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}R_{t+1,i}^e/T\right)^2}$$

Data source:

Returns and Firm Specic Characteristic Variables

We will collect monthly equity return data for all securities from Bloomberg Terminal. The sample period spans January 1970 to January 2020, totaling 50 years. We divide the full data into 20 years of training sample (1970 - 1990), 5 years of validation sample (1990 - 1995), and 25 years of out-of-sample testing sample (1995 - 2020). We use the one-month Treasury bill rates as the risk-free rate to calculate excess returns.

We will also collect the 46 firm-specific characteristics listed on Kenneth French Data Library.

Macroeconomic Variables

We collect 178 macroeconomic time series from three sources. We take 124 macroeconomic predictors from the FRED-MD database as detailed in McCracken and Ng (2016). Next, we add the cross-sectional median time series for each of the 46 rm characteristics. The quantile distribution combined with the median level for each characteristics is close to representing the same information as the raw characteristic information but in a normalized form. Third, we supplement the time series with the 8 macroeconomic predictors fromWelch and Goyal (2007) which have been suggested as predictors for the equity premium.

Reference:

[1] Chen, Luyang and Pelger, Markus and Zhu, Jason, Deep Learning in Asset Pricing (December 4, 2019). Available at SSRN: https://ssrn.com/abstract=3350138