Lecture 6b

Transformers in Portfolio Choice?

Lars A. Lochstoer

UCLA Anderson School of Management

AI ideas in investment

AI-based asset management?

- Applying "transformer"-type techniques to asset allocation questions
- Expected return estimation and real-time trading strategies
- Today, transformer and attention ideas for maximal Sharpe ratio portfolio estimation

This short note is inspired by the paper "AI Asset Pricing Models" posted on BruinLearn

Maximal Sharpe ratio portfolio

We know the maximal Sharpe ratio portfolio weights are:

$$\omega_t = k\Sigma_t^{-1}\mu_t$$

However, estimating Σ_t and μ_t incredibly hard

One approach is to use the condition that $\alpha_i = 0$ for all trading strategies i

$$R_{it} = \alpha_i + \beta_i R_t^{MVE} + \varepsilon_{it}$$

if and only if R_t^{MVE} is the mean variance efficient portfolio (MVE)

- That is, let's find the weights ω_{t-1} such that $R_t^{MVE} = \omega_{t-1}' R_t$
- The latter is verified by finding α_i for all trading strategies

Standard approach in literature

Simply assume

$$\omega_t = X_t \lambda.$$

Here, X_t is an $N_t \times K$ matrix, where each row corresponds to the K characteristics of each of the N_t stock we are considering at time t

λ is the $K \times 1$ vector we are going to estimate

- Thus:

  $$R_{it+1} = \alpha_i + \beta_i \lambda X_t' R_{t+1} + \varepsilon_{it+1},$$

  where we find λ by setting α_i's as close to zero as possible across all stocks

In this standard approach, a stock's portfolio weight is a linear function of its own characteristics

Attention

The attention mechanism of transformers state that one should pay attention to the current context

In our case, that means both macro-economic variables, other stocks' characteristics, and possibly lagged values of the same variables

Let's define an attention matrix A, such that

$$\omega_t = AX_t\lambda,$$

where A is an $N \times N$ matrix of parameters to be estimated.

Note that now, each portfolio weight is a function of other stocks's characteristics as well as its own characteristics

- However, we should have A_t as the number of stocks N_t could change each period

Attention (cont'd)

To accommodate a time-varying number of stocks, consider:

$$A_t = X_t W X_t',$$

where W is a $K \times K$ matrix of parameters to be estimated.

This conditional attention mechanism shares information based on the conditional similarity between the characteristics of asset i and asset j

- In fact, if the rows of X_t are variance-standardized, and W is the identity matrix, then $(i,j)^{th}$ element of A_t reports the correlation across characteristics for asset pair (i,j)
- If W is not identity but is instead a $K \times K$ matrix of free parameters, then the model has the flexibility to learn which conditioning characteristics represent the most fruitful pathways of cross-asset prediction.

Thus:

$$\omega_t = \left(X_t W X_t'\right) X_t \lambda$$

Multi-head attention

We can easily accommodate multiple heads:

$$\omega_t = \underbrace{\left(X_t W_1 X_t'\right) X_t \lambda_1}_{\text{head } \#1} + ... + \underbrace{\left(X_t W_H X_t'\right) X_t \lambda_H}_{\text{head } \#\text{H}}$$

Here, the W_h matrices can attend to different aspects of the drivers of the economy and investor behavior

- I.e., the drivers of the conditional risk-return trade-off

Good out-of-sample performance (allegedly)

For more details, see the paper on AI Asset Pricing posted on BruinLearn

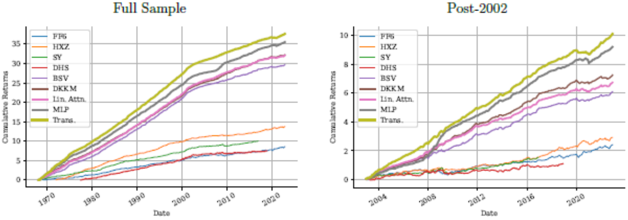

Figure 2: **Cumulative SDF Returns**

This figure shows the cumulative sum of SDF returns for the linear attention and nonlinear transformer models and benchmark models. The left panel shows the full out-of-sample period beginning in 1968, and the right panel begins in 2003.