ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/366137544

Trend Forecast via LSTM & Transfer Learning to Capture Global Drift

Preprint · November 2022

CITATIONS
0

READS
81

5 authors, including:

Xue Wen Tan
National University of Singapore
7 PUBLICATIONS **0** CITATIONS

SEE PROFILE

All content following this page was uploaded by Xue Wen Tan on 09 December 2022.

The user has requested enhancement of the downloaded file.

Trend Forecast via LSTM & Transfer Learning to Capture Global Drift

Xue Wen Tan
Asian Institute of Digital Finance
National University of Singapore
xuewen@u.nus.edu

Chan Kern Chern
Department of Mathematics
National University of Singapore
e0072441@u.nus.edu

Jianyong Song
Asian Institute of Digital Finance
National University of Singapore
songjianyong@u.nus.edu

Zhou Zhou
Asian Institute of Digital Finance
National University of Singapore
zhou.zhou@u.nus.edu

Li Shuping
Asian Institute of Digital Finance
National University of Singapore
shuping.li@u.nus.edu

Abstract

In this project, we intend to develop a framework that can help make investment decisions by suggesting an optimal portfolio containing 10 stocks. Specifically, we would like to first build a model that can forecast stock's expected returns of S&P500 constituents. The global trends captured by the first model will be fine-tuned for every individual stocks to get a meta model. With the expected returns and covariance matrix of returns, we form the optimal portfolio that generates maximum Sharpe ratio.

1 Introduction

According to Barclays Hedge (2022), the Commodity Trend Advisor (CTA) industry manages about 500 billion AUM. These are hedge funds who employ trend following strategies based on prices. Typical strategy includes technical indicators such as moving average crossover and bollinger bands and could include more complicated techniques such as Kalman filtered trend following or wavelet transformation and trend following. These strategies try to capture momentum risk premium but they are sometimes susceptible to sudden change of trend and crash risk.

Classical efficient market hypothesis (EMH) disagrees that such strategies will generate alpha as they essentially trade on prices [Fama (1970)]. These strategies violate the weak-form market efficiency whereby it is assumed that all current information is reflected in stock prices and past information has no relationship with current market prices. However, it was repeatedly found that trend following strategies are profitable [Jegadeesh and Titman (1993)] in a direct contradiction to EMH. Lo (2004) proposed an alternative theory known as adaptive market hypothesis (AMH) that combines EMH with behavioral finance. The gist of the theory argues that the market is mostly efficient but at times, people are motivated by their self-interests, may make mistakes and create predictable price trends. The AMH seems to be more compatible to explain price and trend following strategies. Long Short Term Memory(LSTM) neural network was proposed in 1997 as a neural network that is capable of processing sequential data and remembering short and longer term data points [Hochreiter and

Submission for DSA5205: Data Science in Quantitative Finance ☺. Do not distribute.

Schmidhuber (1997)]. This is ideal for time series data which has been traditionally modeled by statistical methods such as ARIMA models. It is thus interesting to examine if LSTM can be used for trend following and replicate the return profile of LSTM funds. If it is possible, it may have interesting implications such as replacing hedge fund managers and fully allowing AI to automatically trade and follow trends.

Extant LSTM models that we found on Kaggle websites are able to predict stock prices. However, they are often for single period forecast and do not perform better than the traditional AR(1) model that uses last price as the prediction. In addition, they are always trained on a single stock data series. We propose a multi-period forecast whereby our LSTM forecast a stocks' next 21 days returns. The model is trained on a modified loss function whereby we weigh the accuracy of the furthest 21st day most and the 1st day the least. We also propose a two stage approach, whereby we first train the LSTM on 500 stocks (S&P 500 constituents) price series as a global model and apply transfer-learning and fine-tune it to individual stock as the local model. The local model is further augmented with forward-looking indicators such as Cleveland Inflation Nowcast, Financial stability index and S&P 500 sector returns performance time-series. Our predicted 21 days price series are then converted into a 1-month return forecast for all 500 stocks. We then apply a stock selection methodology that picks stocks which are more accurately predicted by the model and bring more profits during a one-year backtesting period. Finally, we apply portfolio optimization which finds weight allocation that maximizes the expected Sharpe ratio with the return forecasts. We found success in our methodology and that it was even robust against periods of great uncertainty such as Covid-19 crisis. Thus, we claim methodological contribution in the usage of neural networks to build investment portfolios.

2 Literature Review

Trend-following is not a new phenomenon. According to Jegadeesh and Titman (1993), buying stocks with high returns over previous 3 to 12 months and selling stocks with poor returns over the same time period earn profit of about one percent per month for the following year. This is further supported by Cliff Asness in a landmark paper titled "Value and Momentum Everywhere" who finds strong evidence of momentum across eight diverse markets and asset classes . ASNESS et al. (2013) then went on to establish AQR, which is now one of the largest hedge funds that trade on the momentum and trend following anomaly.

3 Stock Price Forecast

3.1 Methodology

3.1.1 Baseline LSTM

We started off with examining a plain vanilla LSTM and compared it against an autoregressive AR(1) model. We will use the LSTM model to predict the stock price. A natural question to ask is whether LSTM is powerful enough to capture the information using only the previous price. Hence, we run a simple vanilla LSTM model to predict the stock price of SPY for the following day.

It might seem that LSTM is capturing the trend correctly. However, when we compare it to a model that is using the lagged one-day stock price as the stock price prediction for the current day Figure 1b, the lagged-one-day predictor indeed outperforms the vanilla LSTM model.

Although we cannot claim that using LSTM will always underperform the lagged-one-day-stock-price predictor in predicting the stock price the next day after tuning all the hyperparameter and choosing all the important features, we think that we should fundamentally innovate on the way of how LSTM predicts the stock price.

3.1.2 LSTM and Transfer Learning

We separate our LSTMs into two stages as shown in Figure 2. First, we train a global LSTM model that trains on the price series of all 500 stocks and predicts the next 21 days price series. The objective is to capture global trend patterns that exhibit itself across all stocks. On top of the price series, we also did some feature engineering that created lag features and average stock prices across periods of

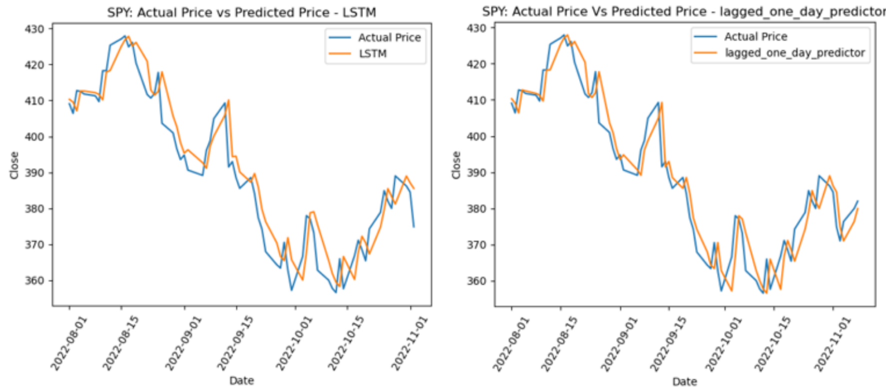

Figure 1: Actual Price vs Predicted Price, LSTM (a) vs AR-1 (b)

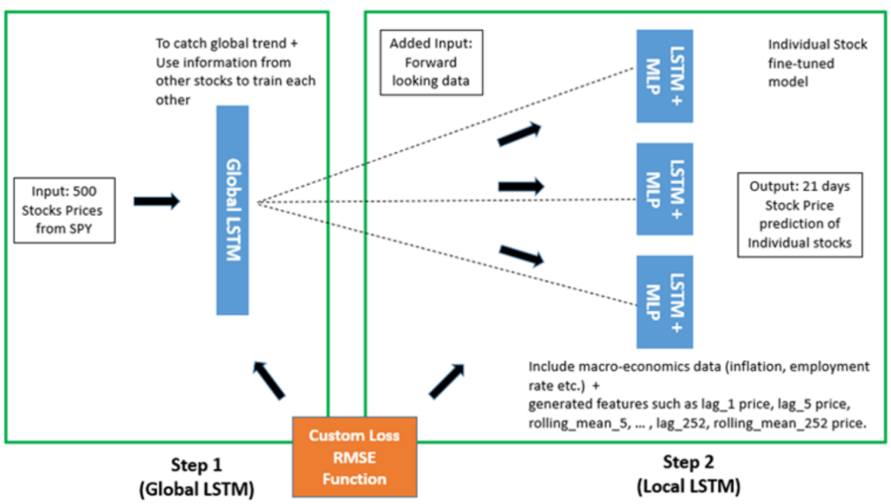

Figure 2: LSTM-Based Proposed Architecture

(1 week, 1 month,3 months,6 months, 1 year). We also calculated returns. In total we calculated 12 additional features and augment it to our price data. We implemented a custom RMSE loss function that weighs prices further out more as they are more difficult to forecast. The formula is as shown in equation 1:

$$rmse(t_{i1}...t_{t21}) = \sqrt{\left(i/\sum_{i}^{21} k\right) * (P_{forecast,i} - P_{actual,i})^2} \tag{1}$$

The network architecture for the global LSTM is as as shown in Figure 3:

Then, for each stock , we created a mixed-model that does transfer learning and fine-tunes the global LSTM and combines it with forward-looking macroeconomic indicators such as Cleveland inflation nowcast, financial stability index and S&P 500 sector performance time series. We also use the same loss function as above. The network architecture is as shown in Figure 4:

3.2 Performance

As mentioned, we trained the global LSTM model until the end of year 2019. Then for each month until September 2022, we will train a Local LSTM for each stock using data until day t-1 and predict the next 21 days stock price. To evaluate our model, we compute the arithmetic returns of our

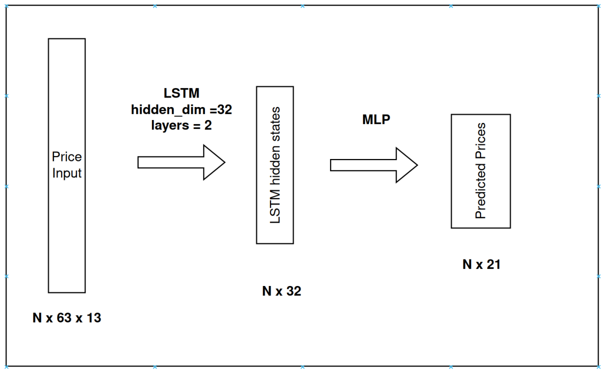

Figure 3: Global LSTM

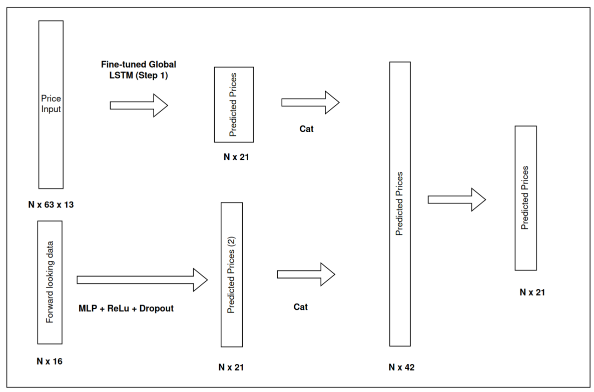

Figure 4: Local LSTM

predictor, and compare it with the actual returns observed. We will use Mean Absolute Error (MAE), confusion table and Spearman Correlation to evaluate the model. Also, we will back test the strategy by taking a long position on the top 50% of the stocks and short the bottom 50% of the stocks.

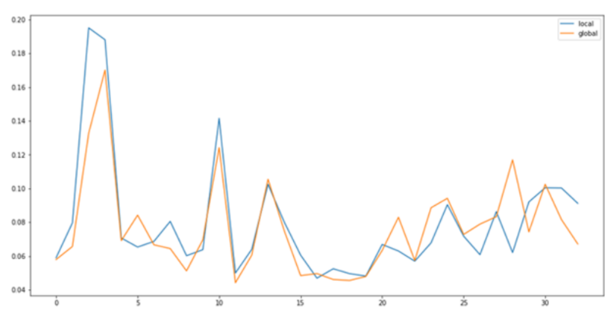

Figure 5: MAE of 1-month returns prediction

For the 33 months from year 2020 until Sep 2022, we showed the mean absolute error (MAE) of the local LSTM and the global LSTM. We can see that **local LSTM has a lower prediction error of 4.6%**, as compared to the MAE of global LSTM, which is around 7.7% as seen in Figure 5.

We also showed the confusion matrix of both global LSTM vs local LSTM in Table 1. Focusing on the diagonal, we can see that both global and local LSTM are doing better in terms of identifying the positive return, and this is mainly because the training period is bullish most of the time. However, the Local LSTM actually fair worse in identifying negative returns. This showed that local LSTM improved the precision of their forecast for positive returns at the expense of weakening the precision

Table 1: Confusion Matrix for Global and Local LSTM

		Ground Truth	
		Positive	Negative
Global Predicted	Positive	5521 (69%)	3658
	Negative	2412	2858 (43%)
Local Predicted	Positive	5593 (79%)	4474
	Negative	2340	2042 (31%)

of negative forecasts. This is not surprising given the lack of bearish data points during the bullish stock market times of our training sample

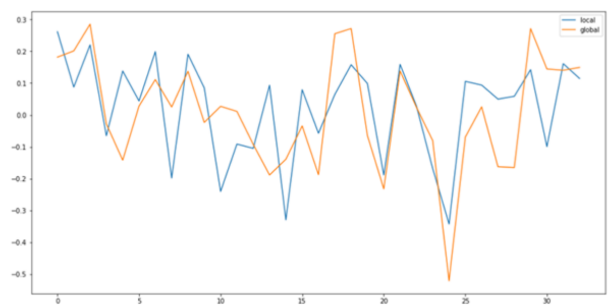

Figure 6: Cross-section Spearman Correlation 1-month price vs actual price

Next, we show the Cross-Section Spearman Correlation 1-month predicted price vs the actual price in Figure 6 for 500 stocks. We can also see that local LSTM indeed outperformed the global LSTM in terms of the spearman correlation (2.25% vs 0.8% on average).

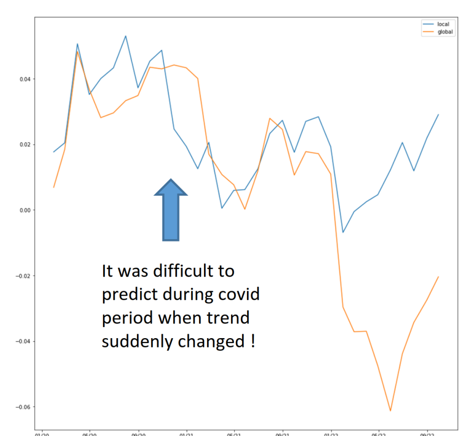

Figure 7: Back testing from 01/2020 – 09/2020 for 500 stocks

We also did back testing based on the strategy of taking a long position for the top 50% of the 500 stocks and short the bottom 50% of the 500 stocks. Our fine-tuned LSTM outperforms that global LSTM during this incredibly volatile times of Covid-19 crisis, an impressive feat for a trend following strategy.

4 Stock Selection

4.1 Trading Signal Generation

We run our model, trade, and rebalance our portfolio every 21 days. At day t, our model would predict future 21-day prices from $\hat{P}_{t+1}$ to $\hat{P}_{t+21}$. The predicted return is $\hat{R}_t = \frac{\hat{P}_{t+21} - P_t}{P_t}$. A positive predicted return indicates buying or longing a stock, while a negative predicted return indicates selling or shorting a stock. Considering the transaction cost, we set a threshold ϵ equal to 0.0005. Only when the absolute value of the predicted return is larger than the threshold, we would trade.

Table 2: Trading Signal Scenarios

Situation	Trading Signal
$\hat{R}_t > \epsilon > 0$	Buy (+1)
$\hat{R}_t \in [-\epsilon, +\epsilon]$	No Trade (0)
$\hat{R}_t < -\epsilon < 0$	Sell (-1)

4.2 Two Criteria

To select the optimal ten stocks, we conduct backtesting from July 2021 to July 2022 to generate two criteria, hit rate and cumulative return.

4.2.1 Hit Rate

After we generate the trading signal and trade correspondingly, when day $t + 21$ really achieves, we would get the real return rate $\hat{R}_t = \frac{\hat{P}_{t+21} - P_t}{P_t}$ for a stock from day t to day $t + 21$. The real return rate indicates a correct trading signal, either buy, sell, or no trade. If the correct trading signal is the same as our predicted trading signal 21 days ago, our trading strategy is correct and makes money. Through the one-year backtesting period, we calculate the hit rate as the percentage of our trading signals being correct. A higher hit rate indicates a higher accuracy of our model for a certain stock.

4.2.2 Cumulative Return

Based on the real price and our trading action, we can calculate the realized return of our trading strategy: $R_t^{'} = \frac{\hat{P}_{t+21} - P_t}{P_t} \times Trading\ Signal$. Through the one-year backtesting period, we calculate the annual cumulative return for each stock. A higher cumulative return indicates more money our model made for a certain stock during the backtesting period.

4.3 Filtering and Sorting

Since our training is based on S&P500 stocks, which include stocks from NYSE, NASDAQ, and AMEX, to fulfill the project requirement, we first filter out non-NYSE stocks. 128 NYSE stocks are left. Then we independently sort the 128 stocks into 5 hit rate portfolios based on their quintile breakpoints for hit rate and into 5 cumulative return portfolios based on their quintile breakpoints for cumulative return. In the intersection of the highest hit rate and highest cumulative return quintiles, there are 21 stocks. We select the top 10 stocks with the highest hit rate among these 21 stocks. Below is the final stock list (all stocks are traded on NASDAQ).

Using these 10 stocks to form an equal-weighted portfolio, from July 2021 to Oct 2022, the cumulative return of this equal-weighted portfolio is far beyond the S&P500 index as shown in Figure 8.

We plot the correlation heatmap. Our selected 10 stocks are all positively correlated. We can observe strong comovement among some stocks such as AMD and AVGO. On the other hand, BIIB has relatively weaker correlation with others. One potential reason attributing to this selection result is

Table 3: Final Stock Picks

Ticker	Company Name
ABMD	ABIOMED Inc
ADBE	Adobe Inc
AMD	Advanced Micro Devices
AVGO	Broadcom Inc
BIIB	Biogen Inc
CMCSA	Comcast Corporation
MKTX	Marketaxess Holdings Inc
MTCH	Match Group Inc
POOL	Pool Corporation
TROW	T Rowe Price Group Inc

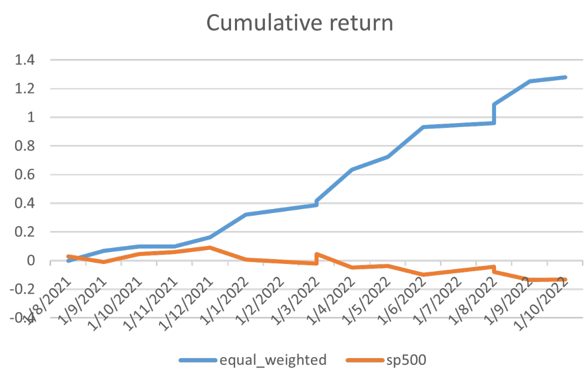

Figure 8: Cumulative return of the equal-weighted portfolio

that our price forecasting models tend to perform much better at predicting price movement at one direction. Since one of our selection criteria is hit rate, stocks with upward-moving price trends are more likely to be selected.

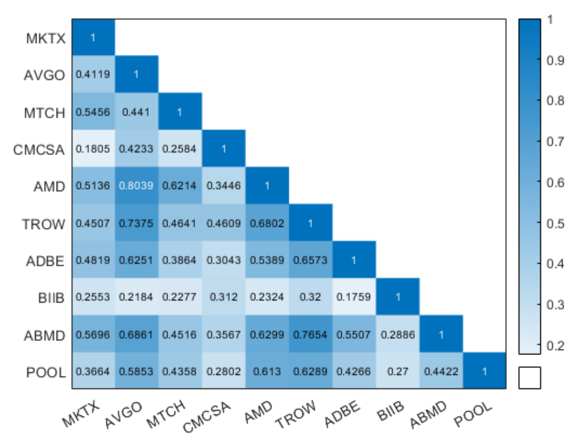

Figure 9: Correlation Heatmap of the selected stocks

5 Portfolio Optimization

5.1 Methodology

After selecting the 10 stocks, the next step is to determine the weight allocation for each stock within the portfolio. We rebalance the portfolio at the beginning of each month with the target of maximizing the expected Sharpe ratio. Short selling is allowed in our practice, i.e., the weight can be negative. To compute expected Sharpe ratio, two components are required. One is the expected portfolio return, the other is the portfolio volatility. The local LSTM model forecasts the daily prices of each stock for the coming month, from which we obtain the expected individual monthly returns. The portfolio covariance matrix is estimated using stocks' daily returns from the past 3 months. Mathematically, our objective function is defined as follows:

$$S_p(w) = \frac{E(R_p) - r_f}{\sigma_p} \tag{2}$$

$$E(R_p) = \sum_{i=1}^{10} w_i E(r_i) \tag{3}$$

$$\sigma_p = w^T \sum w \tag{4}$$

where $E(r_i)$ corresponds to the predicted return from the local LSTM model of stock i, Σ is estimated using historical daily returns as described above. Both $E(r_i)$ and Σ are annualized. The yield on U.S. treasury securities at 1-year constant maturity is used as the risk free rate in the calculation. One thing to note is that the 10 stocks' return time series can have various lengths and missing data could exist. When short time series occur, in order not to sacrifice the other time series, we adopt the following approach for a robust estimate of Σ. First of all, a symmetric matrix A containing pairwise correlations is computed. Then the nearest non-negative definite matrix $\hat{\Sigma}$ is found such that the Frobenius distance between $\hat{\Sigma}$ and A is minimized [Higham (2002), Qi and Sun (2011), Pang et al. (2003)]. $\hat{\Sigma}$ then serves as the robust estimate of Σ. Finally, the constraint portfolio optimization (weights must sum up to 1) is achieved using Python's optimization library scipy.optimize, which is a standard gradient-based optimizer.

5.2 Performance

To evaluate the performance of the portfolio, firstly we conduct backtest for the 1 year period from August 2021 to July 2022. Specifically, for at the beginning of month t, we train the forecast model for each stock using data from $t-3$ to $t-1$, which produces predicted daily price for month t. Portfolio volatility is estimated using daily return from $t-3$ to $t-1$ as well. With these, we determine the optimal stock weights. Then we compute the realized monthly portfolio return and volatility. We track and report the cumulative portfolio return as the first row in the following table, which is 47%. The Sharpe ratio is 0.82. We also report the performance over the 2-month period (from August 2022 to September 2022), and the 1-month period (October 2022)[1]. The returns reported below are annualized from the 2-month cumulative return and 1-month return respectively. Observing that the Sharpe ratio is less than 1 for the 1 year period. One reason is that we fixed the portfolio members due to project requirement. For long-term investment, it is of better practice to change members from time to time. And for the recent month, although the Sharpe ratio is high (5.27), it is calculated from one month observation only and doesn't guarantee future performance.

Apart from backtest to calculate Sharpe ratio, we examine the performance under CAPM model and Fama-French 3 factor model, to test whether our portfolio generates positive alpha. The regression results are tabulated below and confirm positive alphas, though they are not statistically significant. This is expected, as the U.S. market is relatively efficient, so any mispricing would be arbitraged away quickly.

[1]Note that this sample period split is due to project requirement. Under our design framework, our weight allocation is always out-sample (since we use forecast prices), though the stock selection uses information of period August 2021 to July 2022.

Table 4: Portfolio Performance from 2021M8-2022M10

	Annual return	Annual volatility	Sharpe ratio
2021M8 – 2022M7	0.47	0.57	0.82
2022M8 – 2022M9	1.42	0.88	1.58
2022M10	2.87	0.54	5.27

Table 5: Performance under CAPM & Fama-French

	Coefficients	Standard Errors	t stat	P-value
		CAPM		
Intercept	**4.13**	3.64	1.13	**0.27**
Mkt-RF	1.56	0.76	2.05	0.06
		Fama-French 3-Factor		
Intercept	**6.37**	3.95	1.61	**0.13**
Mkt-RF	1.20	0.80	1.50	0.15
SMB	-0.05	1.33	-0.04	0.97
HML	-1.09	0.78	-1.38	0.19

6 Conclusion

In this project, we proposed a framework that can help make investment decisions by suggesting an optimal portfolio containing 10 stocks. First of all, we develop stock price prediction models utilizing the state-of-the-art LSTM. Specifically, a global LSTM model is trained on all S&P500 stocks to capture the market trend. Then local LSTM models are fine-tuned on each stock to capture individual characteristics. The models give forecast daily prices for the subsequent month. Then we select the portfolio members by performing double-sorting on stock return performance and local model accuracy. Finally, with the fixed members, we optimize the weights with the objective of maximizing the Sharpe ratio. Overall, our pipeline enables automatic adjustment of trading strategy (either changing members or rebalancing portfolio with fixed members) every month. Performance-wise, the forecast models tend to be much more accurate in identifying positive jumps, which potentially could be improved by including more bearish period data. In terms of trading performance, we are able to generate positive returns and greater-than-one Sharpe ratios for some period, although when benchmarking against CAPM and FF 3-factor model we don't observe statistically significant alphas. For future work, we propose to introduce more forward-looking features in order to improve accuracy for future price prediction. Sentiment extracted from news or social media can be one promising candidate.

References

C. S. ASNESS, T. J. MOSKOWITZ, and L. H. PEDERSEN. Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985, 2013. doi: https://doi.org/10.1111/jofi.12021. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12021`.

E. F. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970. ISSN 00221082, 15406261. URL `http://www.jstor.org/stable/2325486`.

N. J. Higham. Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343, 2002. doi: 10.1093/imanum/22.3.329.

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

N. Jegadeesh and S. Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91, 1993. ISSN 00221082, 15406261. URL `http://www.jstor.org/stable/2328882`.

A. W. Lo. The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30(5):15–29, 2004.

J.-S. Pang, D. Sun, and J. Sun. Semismooth homeomorphisms and strong stability of semidefinite and lorentz complementarity problems. *Mathematics of Operations Research*, 28(1):39–63, 2003.

H. Qi and D. Sun. An augmented lagrangian dual approach for the h-weighted nearest correlation matrix problem. *IMA Journal of Numerical Analysis*, 31(2):491–511, 2011.

View publication stats