Deep Learning for Event Betting

Srinivasan Sathiamurthy **Rebecca Derham**

1 Introduction

In this research project, we explore mid-frequency event betting on currencies using deep learning techniques. Specifically, we will train a model that has good predictive power over the returns of our chosen securities, and ultimately use this predictive power to create a trading strategy. Mid-frequency trading refers to making at most one trade per day, and event betting refers to trading on high-confidence beliefs in an attempt to ensure that we always make positive returns, rather than only maximizing the expected value of profits.

We hope that this process will yield several results useful to portfolio managers as well as to financial researchers. First, portfolio managers who trade on a daily scale are sensitive to large losses, as their clients then would want to take their money out the next day. In addition, it is harder to make positive returns consistently when only trading once per day as opposed to when trading at a higher frequency. As a result, a methodology for generating reliable mid-frequency strategies would prove highly valuable.

In addition, deep learning methods will be able to capture more relationships between features without having to hand craft them as in traditional financial models. While the value of this has already been seen in many other fields such as computer vision, deep learning has not been as thoroughly explored in mid-frequency trading, mainly due to lack of data. However, we will overcome this limitation in three main ways. First, we will group data from several individual securities into one dataset which can be used to train a single model, rather than limiting ourselves to training on only one security's data. Second, we will discretize our outcome variable by predicting whether the returns will be positive, negative, or neutral, as either positive or negative returns are profitable while neutral ones are not; whether a trade will be profitable is what determines a trading strategy, and the magnitude of the returns is less important, and so this simplification will reduce overfitting while preserving the practicality of our results. Finally, we will bet on only high-confidence beliefs, which will further reduce overfitting and thus provide an advantage over traditional regression-based models.

2 Background Research & Existing Literature

To first address our data selection, it is well known that availability of information has a large effect on markets. When important indicators are released, investors with vastly different requirements, interpretations of the data, and opinions about the market immediately begin trading (Basdekidou, 2017). Ultimately, this results in overreactions to market information, which leads to the existence of profitable strategies trading at this medium frequency (Park, 2017). As such, we use publicly available domestic macroeconomic indicators as part of our feature set.

Next, we must transform this information into actual predictions of future asset prices. Researchers have achieved success with various deep learning techniques across multiple time horizons. The main reason for deep learning's success is that it learns suitable representations directly from raw data, as opposed to conventional methods whose features are engineered manually and combined with domain expertise (Kolm, Turiel, and Westray 2021). However, a lack of data, especially at the mid-frequency level, pervades this space, which has led to mixed results (Li et al. 2010). In our case, since the scope of our research considers trading on a daily scale, we are limited to considering events that occur roughly once per trading day, which is about 252 data points per security per year.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

Macroeconomic events tend to occur less frequently than this, and don't necessarily yield useable information at every occurrence. It then follows that draw-downs pose a much higher risk in medium frequency trading than in high frequency trading, since our events do not occur often enough for us to have the luxury of relying on a positive expected value of our strategy's returns to offset bad draw-downs in the short to medium term. To circumvent this problem effectively, many researchers have discretized the trading process by labeling each time stamp as likely to increase, decrease, or remain relatively stable over some fixed period, and attempting to accurately predict these three classes (Passalis et al. 2020).

The most common models used are combinations of CNNs, LSTMs, and MLPs. Zhang et al, 2018, explored the architecture of a CNN followed by an LSTM for better memory retention and obtained good results. However, Briola et al, 2020, discovered that a large enough MLP performs just as well due to MLPs being universal function approximators. However, one gap we found in our research is that most of these existing approaches use high frequency data in the minute to nanosecond range (Huang et al. 2011), while our investment time horizons are much longer than that.

For our model, we are most interested in the work of Passalis et al., since it incorporates the probability classification structure we wish to ultimately trade off of, and also considers various look-back periods for each feature.

3 Methods & Model

A large portion of our work thus far was focused on gathering data and selecting features to be used in our models. We chose to trade the G10 currencies, as currencies are highly correlated with global economic events and the G10 currencies are less subject to idiosyncrasies than those of emerging markets. We use the past four daily prices as well as the first three moments of each asset's historical price information as features, as well as a variety of macroeconomic indicators. Based on our aforementioned background research, we decided on the following seven macroeconomic features: the Case-Schiller Index, the average monthly manufacturing employee hours, the unemployment rate, the median CPI, the VIX index, the US trade balance, and the spread between 2 year and 10 year US treasury yields. We decided to use data from January 2008 to December 2020 as training data, and from January 2021 to September 2023 as validation data. This gives us approximately 30,000 training datapoints and 4,000 validation datapoints.

After gathering data, we conducted some exploratory data analysis. Based on this, we concluded that predicting whether the security rises above the 66% quantile or below the 33% quantile of returns would be a good goal. We then created a label for each observation based on whether this was the case, and built a simple MLP architecture to predict the label based on our chosen features.

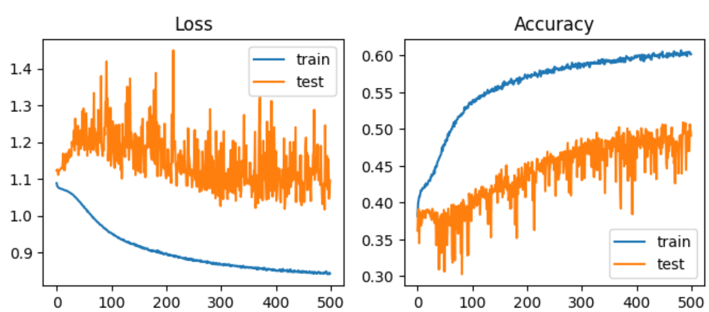

Figure 1: Loss and Accuracy Curves of Baseline Model

4 Preliminary Results

After experimenting with several architectures, we found a single hidden layer MLP with 30 hidden nodes, a ReLU activation, and then a fully connected layer to 3 output nodes with a softmax activation to have the best performance. We trained the model for 500 epochs with cross entropy loss, and got the loss and accuracy curves shown in Figure 1. The final validation accuracy was 49.12%.

```
Training confusion matrix:
[[0.66809221 0.23031188 0.10159591]
 [0.24350808 0.50034297 0.25614895]
 [0.08254105 0.23044512 0.68701383]]
```

```
Test confusion matrix:
[[0.82621951 0.1554878  0.01829268]
 [0.32743078 0.44397939 0.22858983]
 [0.17268623 0.29458239 0.53273138]]
```

Figure 2: Confusion Matrices

To ensure that our model was learning, we also investigated the confusion matrices for both the training and validation predictions (Figure 2). We note here that the classes are nearly perfectly balanced in both the training and validation datasets. Each row corresponds to a true class (-1, 0, or 1), and the columns correspond to the predictions. For instance, the first row of the validation confusion matrix indicates that of all the datapoints which are truly -1 (negative returns), 83% were correctly classified while 16% were classified as 0 (neutral) and only 1% were predicted to be 1 (positive returns). This pattern applies to the other classes as well, for both training and validation; more points are incorrectly classified as a class closer to the true one than as the opposite class. This is a good sign which suggests that our model is truly learning a relationship between the provided features and the returns of the asset, as we did not give our model any ordering of the classes.

5 Evaluation of preliminary work

Overall, our baseline model gave promising results. A testing accuracy of 49% is significantly above the random-guessing baseline accuracy of 33%. In addition, the confusion matrix is encouraging as it indicates that the model is learning the ordering of classes. This means that when we get a prediction that an asset's returns will be, for example, positive, we can be more certain that even if the model may be incorrect, the returns are unlikely to be negative. This will go a long way in helping us achieve our goal of consistently having positive returns on our strategy.

However, there are two potential ways to improve our model. First, based on the loss and accuracy curves, it appears that the model is overfitting significantly; while we were not able to reduce this much by choosing different parameters for our MLP, we suspect that another model could reduce the overfitting gap and increase out-of-sample accuracy. In addition, while the confusion matrix is mostly encouraging, we do see that there is limited accuracy at predicting neutral events. Hopefully, a more advanced model will be able to predict neutral outcomes more reliably, instead of choosing either positive or negative.

6 Future Work

In addition to adding additional features, there are two main ways in which we desire to expand upon this work. First, we will train a more advanced model to give us better accuracy, as well as better interpretability. In particular, we are interested in exploring graphical models such as Markov random fields or potentially Bayesian neural networks. We believe that these models will allow us to encode our domain knowledge more accurately, as well as provide more interpretability as to learned relationships between the features and the outcome. Through these models, we will build off of our textbook intuition as well as our analysis of our baseline model. In addition, we would like to create a trading strategy based off of our predictions, as time permits.

7 Teammates and work division

As Srinivasan has more knowledge of finance, he will focus on finding additional features, investigating the interpretability of the models, and creating and backtesting trading strategies. Rebecca will focus on creating the additional models, and making the driver signal as powerful as possible without overfitting. We will aim to spend 2-3 weeks designing and training models (by November 18th), and 1 week creating a strategy (by November 25th). We will then have 1 remaining week to write the final paper, as well as a few days of buffer in case either of the earlier stages takes longer than anticipated.

References

[1] Park, Seoungbyung (2017) *Factor Based Statistical Arbitrage in the U.S. Equity Market with a Model Breakdown Detection Process*, pp. 419. Master's Theses (2009-)

[2] Passalis, N., Tefas, A., Kannaiainen, J., Gabbouj, M. & Iosifidis, A. (2020) *Temporal Logistic Neural Bag-of-Features for Financial Time Series Forecasting Leveraging Limit Order Book Data*, pp. 183–189. Elsevier

[3] Kolm, P. & Westray, N. (2021) *Information Content Of Cross- Sectional Multilevel Order Flow Imbalance*

[4] Yuhong, L. & Ma, W. (2010) *Applications Of Artificial Neural Networks In Financial Economics: A Survey*, pp. 211-214 Vol 1. 2010 International Symposium On Computational Intelligence And Design. IEEE

[5] Briola, A., Turiel, J. & Aste, T. (2020) *Deep Learning Modeling of Limit Order Book: A Comparative Perspective* arXiv

[6] Huang et al. (2011) *LOBSTER: Limit Order Book Reconstruction System* SSRN

[7] Zhang et al. (2018) *BDLOB: Bayesian Deep Convolutional Neural Networks For Limit Order Books* arXiv

[8] Basdekidou, V. (2017) *Nonfarm Employment Report Trading with Binary Options and Temporal Functionalities*, pp. 15–24. Annales Universitatis Apulensis Series Oeconomica