Submission Number: 3

Group Number: 15

Group Members:

Full Legal Name	Location (Country)	E-Mail Address	Non-Contributing Member (X)
Ewurama Fordjour	Ghana	ewurama.fordjour@gmail.com	
Deven N. Valecha	India	devenvalecha@gmail.com	
Ishaan Narula	India	ishaan.narula@outlook.com	

Statement of integrity: By typing the names of all group members in the text box below, you confirm that the assignment submitted is original work produced by the group (*excluding any non-contributing members identified with an "X" above*).

Ewurama Fordjour, Deven N. Valecha, Ishaan Narula

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed.

N/A

** Note, you may be required to provide proof of your outreach to non-contributing members upon request.*

Answer 9: Model Fit

VARMA provides the best fit to the data as it factors the previous exogenous values. The measure of the p values for the variances of both the response variable and the dependent variable were zero as well as their covariance, quite indicative of some stationarity amongst the data.

Regarding Neural Networks, we evaluate the performance of 2 classification nets and 1 regression net on the test data. This has been summarized in the table below.

	Classification NN		Regression NN
	Model 2	Model 3	Model 1
Activations	ReLU in Hidden Layers Sigmoid in Final Layer	ReLU in Hidden Layers Sigmoid in Final Layer	ReLU in Hidden Layers Linear in Final Layer
Cost Function	Cross-entropy	Cross-entropy	Mean Squared Error
Hyperparameters			
Learning Rate	0.006	0.006	0.04
# Iterations	150,000	150,000	30,000
Layers and Nodes	3 Layers 10 Nodes in Hidden Layers	4 Layers 10 Nodes in Hidden Layers	3 Layers 5 Nodes in Hidden Layers
Lambda	0.00	0.25	0.00
Performance			
Evaluation Criteria	% of Correct Predictions	% of Correct Predictions	Root Mean Squared Error
Training Set	99.99%	99.99%	2.15%
CV Set	52.00%	48.00%	2.15%
Test Set	50.00%	53.85%	2.81%

More details regarding how we arrived at the models with these hyperparameters has been provided in the technical report and the Jupyter notebook.

Answer 10: Model Interpretability

The VARMA provides the best interpretation of the results as it provides better forecasting accuracy as opposed to deep neural networks. Stationarity plot, that is the quantile quantile plot for both the response and explanatory variable showed a clear linear plot with a few outliers to both sides of the line.

Deep Neural Networks are typically quite difficult to interpret, which is why they are often referred to as black box algorithms. This is because it is difficult to comprehend why the neural network learns the parameters it does in its hidden layers. This makes it difficult to understand how the model makes a prediction and how this can be influenced by changing input features.

Answer 11: Work Split Report

- Deven worked on questions 1-7
- Ishaan worked on questions 8, 9-12 (Neural Networks part)
- Ewurama worked on 9-13 (VARMA part)
- Together we all had a look at each others’ parts and made necessary suggestions for the completion of the assignment

Answer 12: Technical Report

Introduction

In a fast-growing global market that we find ourselves in, there has been the need for high-speed technology in everyday transactions so has given birth to many machine learning algorithms. In this module, we sought to understand how and why some machine learning algorithms fail. Many were the factor analysis modules we implemented. We used the dataset provided in the file MScFE 650 MLF GWP Data.csv.

Dataset Characteristics – Skewness, Structural Breaks, Kurtosis, Distributions and Outliers

We run a skewness text to measure the distributions lack of symmetry with respect to its mean and realized that the distribution was normal as all the values were within the significance range of +/- 1.The kurtosis text to also was to quantify the distributions tendency to produce extreme values far from the mean which was also a normality test as all of the values were between +/- 1 of the skewness. The most values from the difference from the mean and median were negative with a few strong positive values.

Our comparison with the skewness however, we realized that most series with a negative skewness had a negative mean median difference as well. There were a few outliers however where the signs tend to be opposite each other.A threshold regression model was equally run to help us identify at least 1 regime shift.

Upon running the quantile-quantile plot (QQ plot), that is super-imposing the points on points we observed that our data followed a normal distribution, there was a clear horizontal line with few outliers. Hence indicative that the data comes from a common distribution.

VARMA

The comment on the QQ-plot does justice to the VARMA. the QQ-plot for the response variable (LUXXX) and the explanatory variable (MSCI ITALY) both showed a very clear linear relationship in the data, thereby signifying stationarity.

Neural Networks

We then train classification and regression neural networks on the dataset, assuming LUXXX to be the independent variable. For the former, the dependent variable captures LUXXX returns’ direction relative to the previous week, taking a value of 1 if the return is higher or 0 if it is lower relative to that in the previous week. For the latter, the dependent variable is simply the weekly return on the LUXXX fund. Both neural nets use Rectified Linear Unit (ReLU) activations in the hidden layers. In the final layer, a sigmoid activation is used for classification and linear activation for regression.

Approach
We start by training a 3-layer network with 5 units in each hidden layer using gradient descent with 30,000 iterations for various learning rates. We then choose a learning rate such that the cost declines for every iteration of gradient descent. After deciding upon the learning rate and number of iterations for which gradient descent seems to work best, we evaluate the models first for bias and then for variance. Bias is fixed by increasing the number of iterations. Then, if the model has high variance, this is corrected by introducing L2 regularization.

Observations (read in conjunction with NN Table in Question 9)
In case of the classification NN, the resulting cross-entropy cost seems to be high indicating high bias within the model, which is reduced by increasing the number of iterations to 150,000. But, because the number of iterations are so high and gradient descent becomes a bit erratic at further iterations, we choose a very small learning rate of 0.006.

This helps us bring down bias but introduces high variance since the resulting models fit the CV set poorly. We, therefore, train the neural nets with L2 regularization. However, that only improves the performance on the CV set marginally (see below).

In regression, however, the MSE cost declines smoothly and rapidly with relatively few updates of gradient descent (30,000 iterations). The resulting model performs well on the training set (implying low bias) and, when tested on the CV set, has low RMSE as well. This indicates that the model generalizes well to unseen data. So, we do not attempt to change the number of layers or other hyperparameters.

Conclusion
The final classification NN does not seem to do a good job in learning from the given dataset. In contrast, the chosen regression NN seems to do a fair job in predicting weekly returns. It does well when weekly return changes on the test set are small but doesn't return a satisfactory prediction when the weekly return changes by a large magnitude. Overall VARMA seems to have modelled the data well with reference to stationarity being evident in the QQ-plot of both the response and the explanatory variables.

Answer 13: Non-technical Email

Dear Management,

Kindly note that in our bid to understand how the various performance metrics for classification models affect trading strategies we put to practice models such as the measures of skewness, kurtosis, the VARMA and Neural Networks. We also performed a regime shift on the distribution and concluded that the series is stationary as the p-values are zero for both regimes. We observe that the classification NN seems to do a poor job in learning from the given dataset. This is because it is accurate only 50% of the time. In contrast, regression NN seems to do a fair job in predicting weekly returns. It does well when weekly return changes on the test set are small but doesn't return a satisfactory prediction when the weekly return changes by a large magnitude. Over all VARMA seemed to have done a better interpretation of our data set.