Develop Modeling Report 2

The modeling approach employed in this analysis involves a two-level logistic regression. Firstly, the focus is on predicting the direction of monetary policy, which can be categorized as Increase, Decrease, or Hold. Subsequently, a second logistic regression is performed to predict the degree of policy change, considering that the Federal Reserve typically adjusts the Federal Funds Target Rate in increments of 0.25 (0.25, 0.50, 0.75, 1.00). The aim is to first anticipate the direction of monetary policy action and then estimate the magnitude of that action using subsets of instances characterized by either an Increase or Decrease in policy.

Before beginning, there is some data augmentation to perform. A "Direction" dummy variable was included, showing "1" when the Fed policy was to Increase the Federal Funds Rate, "-1" for a Decrease, and "0" for a Hold. This will allow not only multi-level logistic regression, but also allow for the model to leverage the ordinal direction of monetary policy. For these models, accuracy will be a key metric, as the likelihood of predicting the variable correctly is integral to the thought process behind choosing these models.

1. Multinomial Logistic Regression

For this model, the exogenous independent variables are leveraged to produce a model that will predict the direction of monetary policy. The utilization of a multinomial logistic regression model provides several advantages in predictive modeling. Firstly, it enables the analysis of categorical variables with multiple classes, extending the capabilities beyond binary logistic regression. The model's simplicity and interpretability facilitate the understanding of relationships between independent variables and the categorical outcome. Moreover, it offers the estimation of probabilities for each class, aiding decision-making processes. Additionally, the model's interpretability and diagnostics allow for evaluation, including assessing model fit, variable importance, and significance.

The model achieves an accuracy of 0.932, making it very accurate against the training data. For the validation test, it achieves metrics as shown below:

Multinomial Logistic Regression	
Accuracy	0.5
Cross-entropy loss	1.145
AICc	-118.502

These accuracy metrics will provide the baseline to which the next models will be compared. To begin, the accuracy of 0.5 leaves something to be desired. The cross-entropy loss quantifies the dissimilarity between the predicted distribution and the actual distribution. This will also be compared against the other models, as there is no standardized metric for these.

2. Random Forest Classifier

A Random Forest Classifier is another method of multi-level logistic regression. The Random Forest Classifier offers several benefits in the context of machine learning. It is an ensemble learning method that combines multiple decision trees to make predictions, resulting in a robust and accurate model. One key advantage is its ability to handle high-dimensional data with a large number of features, as it selects a subset of features at each split, reducing the risk of overfitting. Random Forests are also resistant to outliers and can handle missing data without requiring imputation. Random Forests are relatively more complex than multinomial logistic regression, as each tree has its own subset of features and samples, and involves the combining of multiple decision trees.

There are some hyperparameters that were chosen using example templates for writing this code. To test them, Bayesian hyperparameter selection was performed to calculate the best parameters for running a Random Forest Classifier model, and did not perform as well as the template Random Forest Classifier. The accuracy of the Random Forest Classifier model was 1.0, indicating a perfect prediction for the training set. Running a validation test, there are the following metrics:

Random Forest Classifier	
Accuracy	0.625
Cross-entropy loss	0.998
AICc	-111.469

All accuracy metrics are better than the previous model. The Accuracy shows that the model predicted 12.5% more of the variation in the data. The cross-entropy loss is smaller, indicating that the model was more accurate in predicting the direction of monetary policy than the multinomial logistic regression model.

3. Support Vector Machine

Support Vector Machines (SVM) are a powerful class of supervised machine learning algorithms used for classification and regression tasks. The SVM model is trained with a linear kernel, which allows it to learn linear decision boundaries between different classes. The accuracy of the trained SVM model on the training data is calculated and reported. SVMs have the advantage of being effective in high-dimensional spaces and are known for their ability to handle complex datasets. They aim to find an optimal hyperplane that maximally separates different classes, thus making them a valuable tool for various classification problems.

This model achieves an accuracy of 0.942 in the training set, and the following accuracy measures for the validation set:

Random Forest Classifier	
Accuracy	0.5
Cross-entropy loss	1.056
AICc	-114.221

This model does not perform as well as the others, since it may not have the flexibility of the Random Forest Classifier model. As such, the Random Forest Classifier model still performs the best across all three models.

Magnitude Modeling

Here, the degree to which monetary policy will increase or decrease is evaluated. The data is separated into subsets of when the monetary policy action was an Increase or a Decrease, and the model is fit. This results in a fairly small subset, as it is a subset of a particular datapoint within a training subset. The validation set is even smaller than that. As such, these results should be taken with a grain of salt.

The two models chosen were the multinomial logistic regression model and the Random Forest Classification model. The accuracy against the training sets are as follows:

Increase	
Multinomial Logistic Regression	
Accuracy	1.0
Random Forest Classification	
Accuracy	1.0
Decrease	
Multinomial Logistic Regression	
Accuracy	0.958
Random Forest Classification	
Accuracy	1.0

Their accuracies against the validation sets are as follows:

Increase	
Multinomial Logistic Regression	
Accuracy	1.0
Random Forest Classification	
Accuracy	1.0
Decrease	
Multinomial Logistic Regression	
Accuracy	0.0
Random Forest Classification	
Accuracy	0.75

The Random Forest Classification seems to be more robust to the Decrease levels. As such, the Random Forest Classification is chosen across the board.