Develop Modeling Report 3

The modeling approach employed in this report revolves around refining the predictive capability of logistic regression models for determining the direction of monetary policy. Building upon insights from the previous week's analysis, the focus lies on enhancing the accuracy of predicting whether monetary policy will increase or decrease. While the latter half of the previous analysis demonstrated promising results in forecasting the magnitude of monetary policy shifts, challenges were encountered in accurately predicting the direction. Therefore, this report emphasizes the use of more sophisticated models to accomplish a two-step process: first, predicting the direction of monetary policy, and subsequently predicting the degree of change. By enhancing the models, the aim is to achieve a comprehensive understanding of the factors influencing monetary policy.

1. Ordinal Ridge Regression

Ordinal Ridge Regression presents significant benefits in predicting the direction of monetary policy. The approach refines logistic regression models, enhancing the accuracy of determining whether monetary policy will increase or decrease. Firstly, it maintains the ordinal direction of increasing, decreasing, or holding, since they are not just words, but have information encoded. Ordinal Ridge Regression takes this into consideration when building the model. The model also benefits from estimating probabilities for each category rather than just class predictions. Thus, it provides more information about the predicted values for each observation.

Ordinal Ridge Regression	
Training	
Mean Squared Error (MSE)	0.178
Training Root Mean Squared Error (RMSE)	0.422
R^2	0.399
Adjustsed R^2	-0.189
Accuracy	0.822
Validation	
Mean Squared Error (MSE)	0.500
Training Root Mean Squared Error (RMSE)	0.707
R^2	-0.220
Adjustsed R^2	1.395
Accuracy	0.500

This model does not perform as well enough to be considered usable in the test set. An R^2 of -0.220 in the validation set do not meet the requirements for use beyond model selection, though it will provide a healthy baseline accuracy metric on which to evaluate the following models. The accuracy of 0.5 in the validation set is also well below the requirements for model selection.

2. XGBoost Model

XGBoost Models also provide benefits when predicting multi-level ordinal logistic variables like Direction variable. Firstly, boosting is a helpful method of combining predictions of multiple weak learners to create a strong predictive model. It also allows for feature importance, showing which variables are the most integral to predicting the outcome variable. Finally, it also performs regularization, which helps avoid overfitting.

XGBoost Model	
Training	
Mean Squared Error (MSE)	3.29983E-07
Training Root Mean Squared Error (RMSE)	0.001
R^2	1.000
Adjustsed R^2	1.000
Accuracy	1.000
Validation	
Mean Squared Error (MSE)	0.239
Training Root Mean Squared Error (RMSE)	0.489
R^2	0.417
Adjustsed R^2	1.189
Accuracy	0.750

This model performs much better in the training set, but is weaker when generalized to the validation set. This shows that XGBoost Models do add some predictive power over Ordinal Ridge Regression thanks to its many features. The accuracy score also greatly increases confidence in this model.

3. RNN Model

Finally, RNN Models were used to leverage the strengths of neural networks into this analysis. One benefit of RNN Models is that they are well-suited for sequential data, or time series data such as this. For this model's purpose, one iteration of the model conducted some feature engineering to convert the Date variable to the number of days since the first meeting. This converts the dataset into panel data for which time series analysis can be done. RNN Models also iterate a few times to identify the best model in terms of fit, which provides strength over the previous models that only analyze one model and one set of hyperparameters. In this analysis, hyperparameter optimization was leveraged to choose the best hyperparameters for the model.

RNN Model	
Training	
Mean Squared Error (MSE)	0.000
Training Root Mean Squared Error (RMSE)	0.000
R^2	1.000
Adjustsed R^2	1.000
Accuracy	1.000
Validation	
Mean Squared Error (MSE)	0.000
Training Root Mean Squared Error (RMSE)	0.000
R^2	1.000
Adjustsed R^2	1.000

Accuracy	1.000

This model performed suspiciously well, to the point that some more analysis should be done into the selected features. Unfortunately, this was unable to be performed, as it is still somewhat of a black box. Future weeks will no doubt see an investigation into this model's suspiciously high accuracy.