Statistics Citadel
Jeffrey R. Russell

Logistic and Multinomial Project

The file detailed_trades.csv contains information on tick by tick price changes. PCHANGE is an indicator variable taking the value 1 if the price change is up and 0 if the price change is down. PCHANGE0 is a categorical variable that takes the value 2 if the price change is up, 1 if the price change is zero and 0 if the price change is negative (but we won't use this variable in this homework). The following variables contain information regarding the previous trade and can be used as predictors for PCHANGE and PCHANGE0. RETURN is the one-period lagged return. INDICATOR is a variable that takes the value +1 if the previous trade was buyer initiated and -1 if the previous trade was seller initiated. VOLUME is the lagged volume. SIGN_VOL is the interaction of the log of VOLUME and INIDIATOR (just to give you an idea of things to think about when building your model). LASK is the log of the depth on the ask side. LBID is the log of the volume on the bid side NEWDAY is an indicator variable for a new day. There are 2955 observations.

1) Fit a logistic model for the direction of the next price change (PCHANGE) as a function of LASK LBID, RETURN, and SIGN_VOL. Report the model and the AIC and BIC for this model.

2) Create some new variables that you think will be helpful in forecasting the next price change. You should be clever here and look for variables that reflect any imbalance in order flow. Report the variables that you construct.

3) Build the best model you can using any of the methodologies we have discussed. Explain your procedure and report your final model.

4) Use your final model to evaluate the error rate in your sample. The error rate is the fraction of times that the probability is above .5 buy the outcome is a 0 or below .5 and the outcome is a 1.

5) Come to class ready to predict your model using a new data set consisting of 1000 observations. We will evaluate the model by the error rate. The smallest error rate will be deemed the best model.