PREDICTING JUMP RISK IN HIGH YIELD SPREADS

December 2017

Sonja Tilly

Approach and method

- Project
- Dataset
- Feature engineering
- Data exploration
- Classification
  - Building basic logistic regression model
  - Optimising model parameters
- Results
- Conclusion

Project

- Predict jump risk in high yield spreads
- Using Bloomberg Barclays US Corporate High Yield Average OAS (LF98OAS Index)
- "Jump" is defined as increase in OAS of more than 50bps in 30 days
- This is a classification task

Dataset

```
data.columns
```

```
Index(['FDTR Index', 'VIX Index', 'USTW$ Index', 'USGG10YR Index',
       'USGG30YR Index', 'USGG5YR Index', 'USYC2Y10 Index', 'USYC2Y30 Index',
       'USYC5Y10 Index', 'USYC2Y5Y Index', 'CPI YOY Index', 'FDIDFDMO Index',
       'USURTOT Index', 'NFP TCH Index', 'USEMNCHG Index', 'IP CHNG Index',
       'MTIBCHNG Index', 'MGT2TB Index', 'PIDSPINX Index', 'NAPMPMI Index',
       'OUTFGAF Index', 'USGGBE10 Index', 'USHBMIDX Index', 'CONCCONF Index',
       'OEUSKLAR Index', 'RSTAMOM Index', 'USTBTOT Index', 'FDDSGDP Index',
       'M2% YOY Index', 'CLA Comdty', 'CONSSENT Index', 'PIDSDPS Index',
       'CICRTOT Index', 'ARDIMOYY Index', 'M1% YOY Index', 'SCGRRAI Index',
       'EMB US Equity', 'SPE AUTO Index', 'SPE CARD Index', 'BFCIUS Index',
       'INJCJMOM Index', 'LEI CHNG Index', 'GDP CYOY Index', 'SLDETIGT Index',
       'SLDETGTS Index', 'GFSIFFND Index', 'XLF US Equity', 'XLE US Equity',
       'XLK US Equity', 'XLV US Equity', 'XLI US Equity', 'XLY US Equity',
       'XLB US Equity', 'CDX HY CDSI GEN 5Y SPRD Corp', 'PCUSEQTR Index',
       'GFSIRLIQ Index', 'CIGMEMRA Index', 'CIGMGRAM Index', 'LF98OAS Index'],
      dtype='object')
```

- The data contains rows of economic indices in daily observations since 1997, showing 7,574 rows and 59 columns.

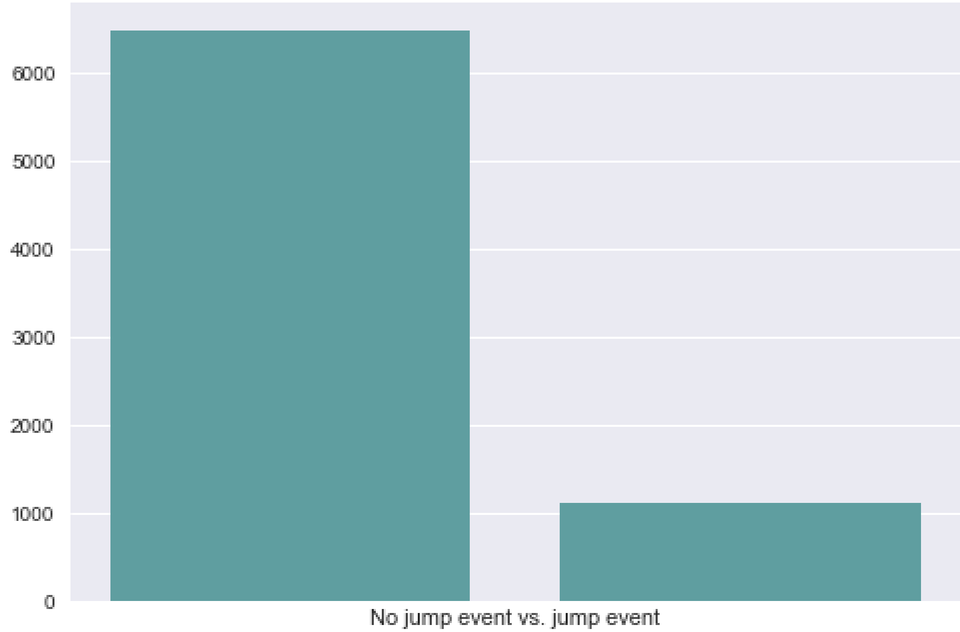
- The target column is the Bloomberg Barclays US Corporate HY Avg OAS Index (LF98OAS Index).

Feature engineering

- Address time lags in indicators
- Time windows of 30, 90 and 180 days have been defined
- The z-score has been calculated for all features for the above defined time windows
- Level change and percentage change have been calculated for all features for the above defined time windows

Feature engineering

- Transform target variable into binary classification:
  - LF98OAS Index change in 30ds larger or equal to 0.5 = 1
  - Else = 0
- The target variable is unbalanced, with “jumps” representing less than one sixth of the data.

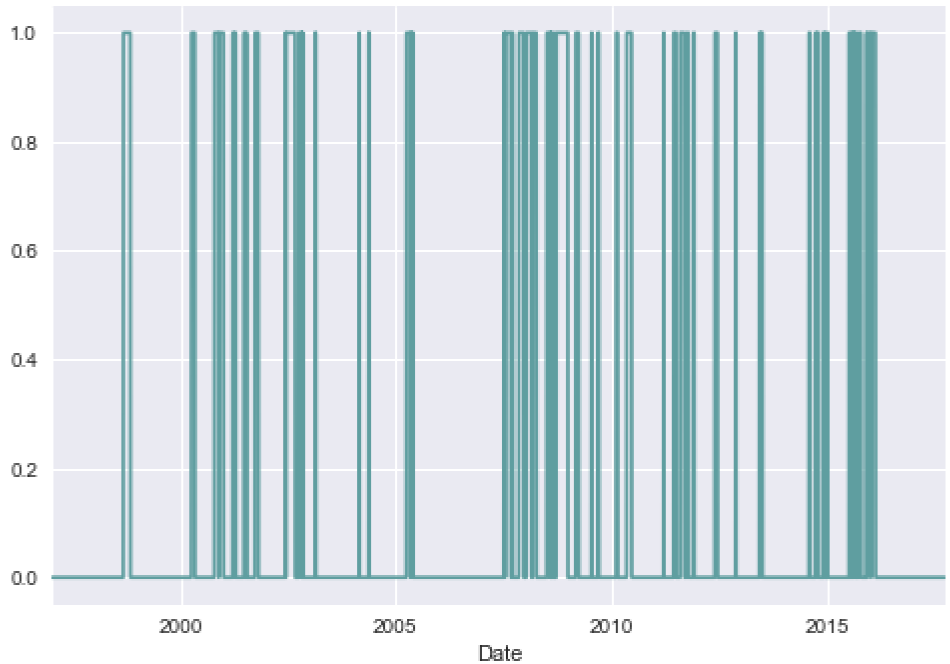


Feature engineering

- Drop the following indices as they are, or relate to the target variable and represent an information leak:
  - 'LF98OAS Index', 'LF98OAS Index change in 30ds','LF98OAS Index change in 90ds', 'LF98OAS Index change in 180ds', 'LF98OAS Index30zscore','LF98OAS Index90zscore', 'LF98OAS Index180zscore'
- Address infinite and nan values (replace with zero)

Data exploration

➢ The frequency of jumps in high yield spreads has markedly increased from the global financial crisis in 2008.



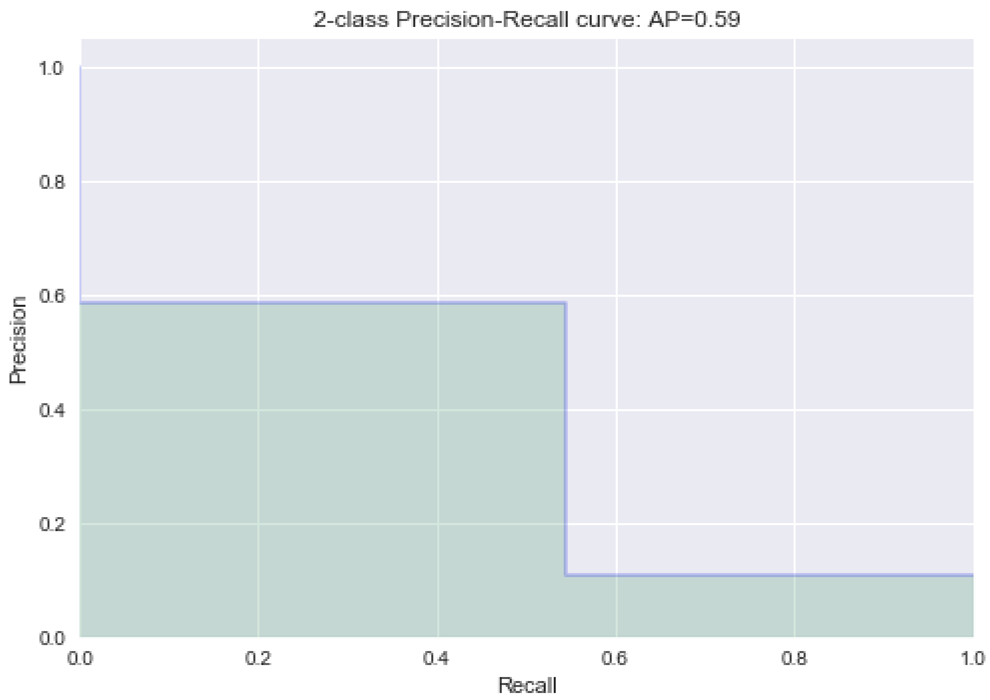
Classification

- Define predictor variables and target
- Split out training and testing sets
  - Training: first 6,015 observations
  - Testing: last 1,559 observations
- Build Gradient Boosted Tree model
  - Model evaluation metric: precision and recall metrics
  - Precision = True positives/(True positives/False positives)
  - Recall = True positives/(True positives/False negatives)

Classification

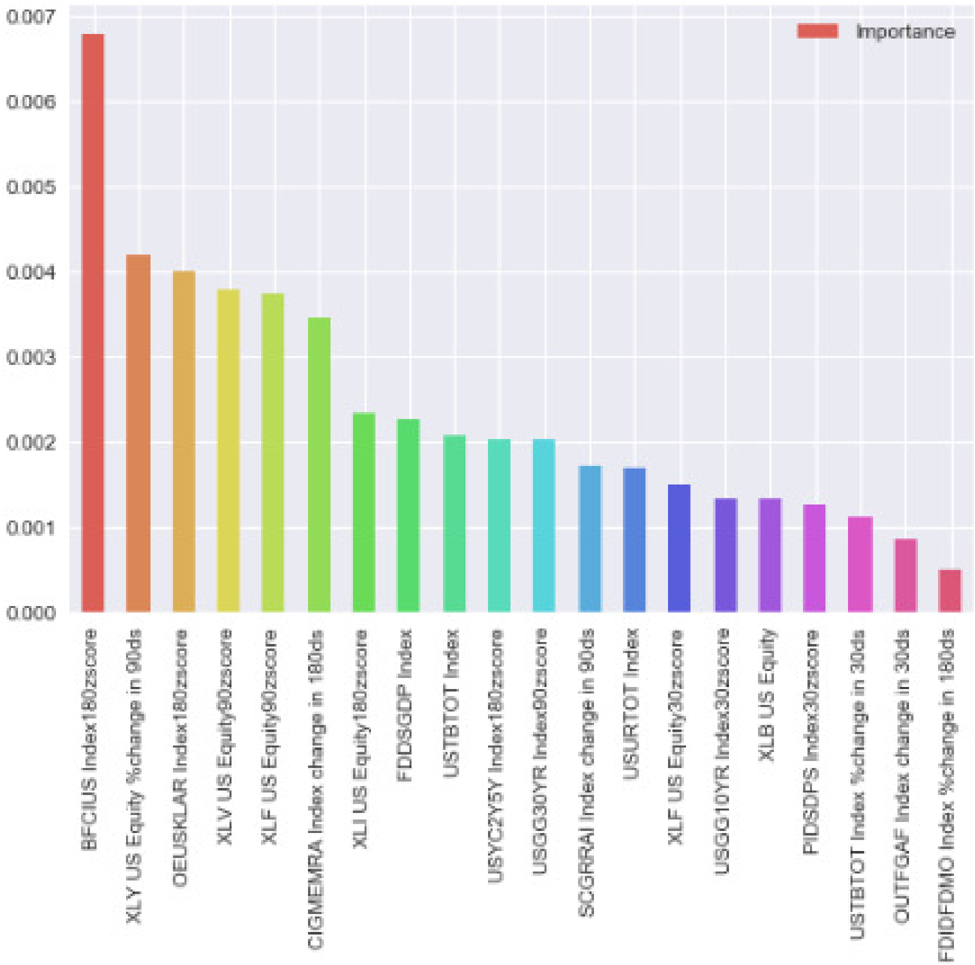
- Tuning the following parameters had positive impact:
  - Account for unbalanced classes (sample_weight);
  - Increase number of boosting stages to perform (n_estimators);
  - Increase the number of nodes in each individual tree (max_depth);
  - Reduce the number of features considered when looking for the best split (max_features).
- Evaluate model performance with cross-validation

Results: model performance



- The model achieves an average precision of 0.59 in predicting jump risk.

Results: feature importances



- Top six features:
  - BB US Financial Conditions Index;
  - Consumer Discretionary;
  - US OECD Lead Indicators;
  - Healthcare;
  - Financials;
  - Citi EM Risk Aversion index.

Conclusion

- The model delivers an average precision in predicting jump risk of 0.59.
- This demonstrates that the classifier performs better than random guessing.
- The dataset contains a lot of noise, which is likely to weigh on the score.

- The most important drivers of US high yield spreads are
  - Financial conditions;
  - Leading economic indicators;
  - Specific sectors (which may include a lot of high yield issuers);
  - Risk aversion.