Machine Learning - Bond Liquidity Prediction

Introduction

A corporate bond is a debt instrument issued by companies to raise money for business operations. During issuance, investors lend money to the issuing company in exchange of a legal document promising fixed future payments in the form of a coupon on a notional (where coupon is the annual interest rate paid on the bond). After a bond has been issued, it can be traded amongst investors, often via banks with dealers acting as an intermediary.

Corporate bonds are not traded on a public exchange unlike stocks. This means its hard to gain insight into how much they are going to be traded or in other words its liquidity. In simple terms, liquidity is the relative ease with which a bond can be traded in the market. A liquid bond is one which trades more frequently and in higher volumes. Knowing how liquid a particular entity would be on the market is very important. It drives multiple things not limited to but including the price of the entity, the risk associated with holding on to it etc. Gaining this insight is precisely the objective of this challenge. In this challenge given bond characteristics like coupon rate etc. and the historical prices and volumes for various bonds you must compute how much each of these bonds would be bought or sold in the future.

You would be given historical prices/volumes for a 3 month period, and your objective would be to predict the total buy and sell volume for each bond over the 3 days immediately after the aforementioned 3 month period.

Input data

There would be two input datasets for this challenge:

Bond characteristics data

This dataset contains static characteristics for the bonds being traded in the market. This would be in the following format:

```
isin,issuer,issue
Date,market,amtIssued,amtOutstanding,collateralType,coupon,couponFrequency,couponType,industryGroup,industrySector,indust
rySubgroup,maturity,maturityType,securityType,paymentRank,144aFlag,ratingAgency1Rating,ratingAgency1Watch,ratingAgency1
EffectiveDate,ratingAgency2Rating,ratingAgency2Watch,ratingAgency2EffectiveDate
isin1,issuer1,issue
Date1,market1,amtIssued1,amtOutstanding1,collateralType1,coupon1,couponFrequency1,couponType1,industryGroup1,industryS
ector1,industrySubgroup1,maturity1,maturityType1,securityType1,paymentRank1,144aFlag1,ratingAgency1Rating1,ratingAgency
1Watch1,ratingAgency1EffectiveDate1,ratingAgency2Rating1,ratingAgency2Watch1,ratingAgency2EffectiveDate1
```

The various properties of the bond that are being described are:

- isin - Unique Identification number to identify the bond.
- amtIssued - Total notional of bond when it was first issued
- amtOutstanding - Total tradable notional of bond remaining in market
- industrySector - High level Industry Sector classification of the issuer of the bond. Eg. Communications, Financials, Consumer Non-Cyclical
- industryGroup - Intermediate level Industry Group classification of the issuer of the bond. Eg. Telecommunication, Banks, Health Industries
- industrySubGroup - Lower level Industry Group classification of the issuer of the bond. Eg. Telephone-Integrated, Diversified Banking Inst, HealthCare Services

- issuer - Name of the Entity Issuing the Bond
- ticker - Symbol to identify the Entity issuing the bond
- country - Country where the issuing entity has its headquarters
- collateralType - Collateral (if any) posted against the issued debt.
- coupon - Coupon rate of bond
- couponType - Type of Coupon Rate eg. Fixed, floating, variable
- couponFrequency - Frequency with which coupons are paid in a year (annually, semi-annually, Quarterly, No Coupon)
- issueDate - Date when the bond was issued
- maturityType - Type of Maturity of bond e.g. callable, convertible, perpetual, sinkable
- maturityDate - When the bond gets paid out in full.
- market - Broad categorisation of bonds based on ratings and scope of trading
- securityType - Defines the scope of trading of bond e.g. Euro-Dollar, US Domestic etc
- paymentRank - The order of repayment in event of bankruptcy
- 144aFlag - Indicates if the bond was a 144a offering which places some restrictions on trade ability of bonds.
- side - Indicates if the trade was a buy or sell
- ratingAgency1Rating - Rating issued by Credit Rating Agency 1 to assert the “credit quality” of bond
- ratingAgency1Watch - Indication if Rating agency 1 is considering taking any action (e.g. upgrade, downgrade. no-action) regarding the rating of product
- ratingAgency1EffectiveDate - Date when Rating agency 1 last rated the product.
- ratingAgency2Rating - Rating issued by Credit Rating Agency 2 to assert the “credit quality” of bond
- ratingAgency2Watch - Indication if Rating agency 2 is considering taking any action (e.g. upgrade, downgrade. no-action) regarding the rating of product
- ratingAgency2EffectiveDate - Date when Rating agency 2 last rated the product.

The bond characteristics data can be downloaded from here

We noticed some issues with the date formats in the file posted previously. A standardized version can be downloded from here. All dates in this updated file are in DD-MM-YYYY format. Missing dates are marked NA

Historical price and volume data

This dataset contains the price and volume at which bonds were traded at various points of time. This would be in the following format:

```
isin,time,price,side,volume,timeofday,date
isin1,time1,price1,side1,volume1,timeofday1,date1
isin2,time2,price2,side2,volume2,timeofday2,date2
```

The values being mentioned above are:

- isin - Unique identification number to identify the bond
- time - Timestamp at which the trade was executed. Format: dd-mm-yyyyhh:mm:ss
- price - Price at which the bond was traded. Format: No decimal values

- side - Indicates if the trade was a buy or sell trade (Buy/Sell represented by B or S respectively)
- volume - The notional traded in the trade for the isin at a particular time
- timeofday - Represents if the trade was executed in morning/afternoon session. Possible values: morning, afternoon
- date - Date when the trade was executed

The historical price and volume data can be downloaded from here

Ouput data format

For each bond mentioned in the bond characteristics dataset, you need to predict the total buy and sell volume for the 3 days following immediately after end of the training dataset

```
isin,buyvolume,sellvolume
isin1,buyvolume1,sellvolume1
isin2,buyvolume2,sellvolume2
```

Note: DO NOT quote your csv

Evaluation criteria

Your solutions will be evaluated in two ways:

1. Objective evaluation of the volume predictions

   This has two parts to it

   - Numerical accuracy of your predictions
   - Relative ordering of your predictions

     You can find more details here

2. Subjective evaluation of your models based on your documentation, code etc.

The leaderboard would comprise purely of the objective part of the evaluation strategy

Sample input

Historical price and volume data

```
isin,time,price,side,volume,timeofday,date
isin10033,Mon 16May16 02:51:27.000 pm,117.36361835280748,S,730000,Afternoon,16May2016
isin10033,Mon 16May16 03:45:52.000 pm,115.71532940722403,B,730000,Afternoon,16May2016
isin100,Fri 25Mar16 11:46:33.000 am,104.83997339853714,B,210000,Morning,25Mar2016
isin10272,Wed 23Mar16 11:43:48.000 am,109.6939434878436,S,3900000,Morning,23Mar2016
isin10272,Wed 23Mar16 11:45:15.000 am,109.6939434878436,S,3900000,Morning,23Mar2016
isin10272,Wed 23Mar16 12:02:36.000 pm,109.58922373272395,B,3900000,Afternoon,23Mar2016
```

Bond characteristics data

```
isin,issuer,issue
Date,market,amtIssued,amtOutstanding,collateralType,coupon,couponFrequency,couponType,industryGroup,industrySector,indust
rySubgroup,maturity,maturityType,securityType,paymentRank,144aFlag,ratingAgency1Rating,ratingAgency1Watch,ratingAgency1
EffectiveDate,ratingAgency2Rating,ratingAgency2Watch,ratingAgency2EffectiveDate
isin167,issuer350,11-02-
2015,IG,5235990000,5235990000,collateralType26,3.4,2,couponType1,industryGroup70,industrySector1,industrySubgroup337,25
-02-2025,maturityType5,securityType4,paymentRank8,N,rating18,,02-02-2015,rating14,,02-02-2015
isin219,issuer16,14-05-
2013,IG,4184850000,4184850000,collateralType26,1.75,2,couponType1,industryGroup57,industrySector3,industrySubgroup190,1
5-08-2017,maturityType0,securityType4,paymentRank8,N,rating6,,08-02-2016,rating15,,11-03-2016
isin225,issuer16,24-02-
2015,IG,3926990000,3926990000,collateralType26,3.6,2,couponType1,industryGroup57,industrySector3,industrySubgroup190,24
-02-2025,maturityType5,securityType4,paymentRank8,N,rating6,,08-02-2016,rating15,,11-03-2016
```

Sample output

```
isin,buyvolume,sellvolume
isin1,1000000,1000000
isin2,1000000,1000000
isin3,1000000,1000000
```

NOTE: Again note that there should be no quoting in the output's you submit

Model Doc & Source Code Submission

Before the competition ends, please upload the following in the corresponding sections:

1. All source code files (quoting any references you may have used)
2. A documentation of your solution. This can be in PDF, PPT, PPTX, DOC or DOCX formats. Do include the following aspects in your modeldoc:
    - Assumptions
    - Any mathematical simplifications/approximations
    - Modelling choices and comments on appropriateness of answers
    - Any plots/graphs which explain the process of your model selection

Note:
Please ensure consistency between output and the methodology you explain in the model document If there are any additional ideas that you wish to include in the documentation which could not be implemented, please clearly mark them so.

Hints

1. Certain events (like new issuance, rating upgrade / downgrades to name a few) can cause a "spike" in liquidity for bonds.
2. Liquidity of some bonds can affect liquidity of other similar bonds.