Benchmark Bond Trade Price Challenge

https://www.kaggle.com/c/benchmark-bond-trade-price-challenge

Background & Objective

- Bond prices not available real-time
- No public quotes available
- Lack of liquidity, data may be obsolete
- Difficulty to get accurate bond pricing

➔ Predict the next price that a US corporate bond might trade at

Explanatory Variables (1/3)

Name	Type	Missing values?	Description
Id	Integer	No	Identifier for the row
bond_id	Integer	No	unique identifier of the bond to aid in time-series reconstruction
weight	Double	No	Weight of the row for evaluation purposes
current_coupon	Double	No	Coupon of the bond at the time of the trade
time_to_maturity	Double	No	Number of years until the bond matures
is_callable	Integer (Cat.)	No	Binary value indicating whether or not the bond is callable by the issuer
reporting_delay	Double	No	Time the trade was reported by Trace – Time the trade occurred in seconds
trade_size	Integer	No	Notional amount of the trade (un-signed)
trade_type	Integer	No	2=customer sell, 3=customer buy, 4=trade between dealers
curve_based_price	Double	No	Fair (medium term) price estimate based on implied hazard and funding curves of the issuer of the bond

Explanatory Variables (2/3)

Name	Type	Missing values?	Description
received_time_diff_last{1..10}	Integer	Yes	Cumulative time difference in seconds between the time the trade was reported and the time the previous {1..10} were reported
trade_price_last{1..10}	Double	Yes	Trade price of the last {1..10} trades
trade_size_last{1..10}	Integer	Yes	Notional amount of the last {1..10} trades
trade_type_last{1..10}	Integer	Yes	Trade type of the last {1..10} trades
curve_based_price_last{1..10}	Double	Yes	Curve based price of the last {1..10} trades

Explanatory Variables (3/3)

- Reporting delay:
  - Negative values
  - Extremely large values (months, years)
- Historical columns:
  - NaN values when dataset has no previous trades to report – but does not mean the bond has never traded before
- Curve based price:
  - Factored in parameters in the computation of bond price (reference data, interest rate data, trades/quotes of related bonds/CDS)
- Weight:
  - Used for evaluation purposes only

Response Variable

Name	Type	Missing values?	Description
trade_price	Double		Price at which the trade occurred

Train Data Set

- 762,678 rows and 61 columns
- Excluding id, bond_id, weight

	current_coupon	time_to_maturity	is_callable	reporting_delay	trade_size	trade_type	curve_based_price
count	762678.000000	762678.000000	762678.000000	762678.000000	762678.000000	762678.000000	762678.000000
mean	5.588142	6.564305	0.111405	28237.606291	267207.791545	3.231778	103.219652
std	1.724374	6.267664	0.314633	1064998.166263	772585.858316	0.764844	9.932693
min	0.000000	0.082471	0.000000	-58.581000	1.000000	2.000000	1.746127
25%	4.750000	2.493513	0.000000	2.491000	10000.000000	3.000000	98.899486
50%	5.625000	4.775924	0.000000	9.438000	25000.000000	3.000000	102.121523
75%	6.500000	8.258087	0.000000	34.967750	100000.000000	4.000000	106.988161
max	13.500000	84.159498	1.000000	99879143.063000	5000001.000000	4.000000	172.956162

Test Data Set

- 61,146 rows and 59 columns
- Does not include bond_id and trade_price

	current_coupon	time_to_maturity	is_callable	reporting_delay	trade_size	trade_type	curve_based_price
count	61146.000000	61146.000000	61146.000000	61146.000000	61146.000000	61146.000000	61146.000000
mean	5.642645	6.798083	0.136902	27284.158880	291197.037451	3.215075	103.291058
std	1.865132	6.191860	0.343747	1003977.848805	808272.376423	0.772668	9.566156
min	0.000000	0.091015	0.000000	-45.508000	1000.000000	2.000000	36.301502
25%	4.700000	3.005876	0.000000	2.469000	10000.000000	3.000000	98.740655
50%	5.550000	5.236462	0.000000	9.733000	25000.000000	3.000000	102.383006
75%	6.625000	8.367985	0.000000	36.947500	110000.000000	4.000000	107.326145
max	13.500000	85.370731	1.000000	73960945.012000	5000001.000000	4.000000	159.714789