

# Timely Bond Price Prediction: Machine Learning Approaches

Finance & Commerce

Mingze Liu, SUID: 06779954; Dongchen Li, SUID: 06780318

April, 2023

The fixed-income market is often considered one of the most essential components of the global financial system. Fixed income products, such as interest rate bonds, are constantly used for market risk-hedging purposes on trading floors.

Nevertheless, the fixed-income market is composed of a larger variety of bond products with less liquidity and more importantly, less transparency than normal equities. This often raises higher liquidity and potentially higher operational risks for both long and short parties. As a result of limited information available, bond prices are oftentimes days old, compared with equity prices that are refreshed within minutes with such abundant relevant information as bid-ask spread, offers and trade size, and don't reflect the true state of the fixed-income market.

This project aims to take a machine learning approach to predict bond prices in a timely manner in the interest to help mitigate inaccurate and outdated bond prices resulting from information asymmetry between required and available information for practical bond hedging activities.

In his study, Colin (2006) [1] applied an Artificial Neural Network (ANN) to predict 50-year US Treasury bond prices with 6 bond attributes from historical bond trades: coupon, maturity, bond yield, and settlement date with 1 hidden layer that outputs 500 predictions. In a similar fashion, Augusto dos Santos (2006) [2] attempted to predict 10-year US Treasury bond yield, on which bond prices are dependent using ANN but with key economic indicators instead, such as the Consumer Price Index (CPI).

There is seldom research on the combination of the 2 types of inputs, bond attributes from historical trading data, and key economic factors such as inflation. Our project goal is to tackle this practical bond price prediction problem by applying machine learning methods on both macroeconomic indicators and historical trading data. One of the datasets that might be suitable for our project is the Benchmark Bond Trade Price Challenge dataset posted by the bond trading firm Benchmark Solution [3] with historical data on 60 bond attributes.

This project will start with exploratory data analysis to perform data cleaning and pre-process data obtained. On top of which, PCA will be applied to select important well-balanced training and test datasets while conducting attribute selection among large datasets for dimensionality reduction. We decide to use GLM regression as our baseline model, the result of which will be obtained and compared to those of ANN, as we believe ANNs are well-suited for this task given its ability to fit highly non-linear data with enough neurons. Certain non-overlapping dataset will be selected and used for the validation process when error tracking is conducted for prediction power measuring. Lastly, we can easily compare results from our model to reality. The bond prices we are aiming to predict will be widely available by the time we finish.

## References

- [1] Colin, Andrew. (2006). Fixed Income Attribution with Minimum Raw Material.
- [2] Castellani, Marco & Santos, Emanuel. (2006). Forecasting Long-Term Government Bond Yields: An Application of Statistical and AI Models. Department of Economics at the School of Economics and Management (ISEG), Technical University of Lisbon., Working Papers