AMCIS 2022 Proceedings

SIG ODIS - Artificial Intelligence and Semantic Technologies for Intelligent Systems

Aug 10th, 12:00 AM

Using Explainable AI to Understand Bond Excess Returns

Lars Beckmann
University of Muenster, l_beck51@uni-muenster.de

Jörn Debener
University of Muenster, joern.debener@wiwi.uni-muenster.de

Johannes Kriebel
University of Muenster, johannes.kriebel@wiwi.uni-muenster.de

Using Explainable AI to Understand Bond Excess Returns

Emergent Research Forum (ERF)

Lars Beckmann
University of Muenster
lars.beckmann@wiwi.uni-muenster.de

Jörn Debener
University of Muenster
joern.debener@wiwi.uni-muenster.de

Johannes Kriebel
University of Muenster
johannes.kriebel@wiwi.uni-muenster.de

Abstract

Recent empirical evidence indicates that bond excess returns can be predicted using machine learning models. While the predictive power of machine learning models is intriguing, they typically lack transparency. We introduce SHapley Additive exPlanations (SHAP), a state-of-the-art explainable artificial technique, to open the black box of these models. Our analysis identifies the key determinants that drive the predictions of bond excess returns in machine learning models and how these determinants are related to bond excess returns. Thereby, our approach facilitates an in-depth interpretation of the predictions of bond excess returns made by machine learning models.

Keywords

Bond excess returns, machine learning, explainable artificial intelligence.

Introduction

There is a recent large increase in interest in using machine learning to predict returns on assets in financial markets. Important recent contributions show that the returns on assets can be predicted much better with machine learning methods than with traditionally used linear methods (Bianchi et al. 2021; Gu et al. 2020). This study addresses the use of machine learning to predict bond excess returns on financial markets. When using such methods in decision making, there is a need for transparency and a need for decision makers to understand and interpret the results. Therefore, we study the use of explainable artificial intelligence to investigate what information is most important for predictability.

Bond excess returns are defined as the return from buying a long-term bond today and selling it after a certain holding period less the return from investing into a short-term bond today and holding it for the same period until maturity (Cochrane and Piazzesi 2005). Early empirical studies have been able to show that bond excess returns vary considerably over time, and, in this way, predictability would be very beneficial. There are important studies arguing that bond excess returns can be predicted using information on the structure of interest rates over different maturities (e.g., Campbell and Shiller 1991; Cochrane and Piazzesi 2005). Furthermore, several studies argue that different macroeconomic indicators such as inflation, labor market conditions, and economic growth are predictive beyond information on the structure of interest rates over maturities (e.g., Huang and Shi 2021; Joslin et al. 2014; Ludvigson and Ng 2009; Wright 2011). Besides investors, it is important for central banks to know and understand the determinants of bond excess returns as precisely as possible as it provides central banks with continuous information on market expectations and serves as direct feedback for monetary policy actions (Bauer and Hamilton 2018; Kim and Orphanides 2007).

The following section provides a brief overview of the literature on bond excess returns and machine learning in finance. The third and fourth sections introduce the data we use and our research methodology. The fifth section presents preliminary results. The sixth section concludes the paper.

Literature Review

From a theoretical perspective on the predictability of bond excess returns, the so-called expectation hypothesis posits that investors expect the same return on long-term bonds as on short-term bonds when held for the same period (McCallum 1975). However, this assumption has long been a focal point of discussion. There are theoretical arguments that bond excess returns can be predicted using information on the structure of interest rates over different maturities which some empirical studies could find evidence for (e.g., Campbell and Shiller 1991; Cochrane and Piazzesi 2005). This has been extended by several recent empirical studies showing that different macroeconomic indicators such as inflation, labor market conditions, and economic growth have predictive power beyond the information in the structure of interest rates and should therefore be included in predictive models (e.g., Bianchi et al. 2021; Huang and Shi 2021; Joslin et al. 2014; Ludvigson and Ng 2009; Wright 2011).

From a methodological perspective, most studies until recently relied on linear methods to predict bond excess returns. However, advantages of using machine learning methods have been widely documented in many different research areas, including finance, for example in predicting asset prices (Ndikum 2021). Important recent contributions show that stock returns can be predicted to some extent with machine learning methods such as neural networks, random forests, or gradient boosting while there is only low predictability with linear methods (Gu et al. 2020; Gu et al. 2021). While these studies highlight the usefulness of machine learning methods to predict bond excess returns, an important shortcoming of these methods is their lack of interpretation of the contribution of individual features and the related difficulty for decision makers to act based on their outcomes. This black-box characteristic of machine learning methods is a focal point of interest in the information systems literature and have led to the development of several explainable artificial intelligence techniques (Adadi and Berrada 2018; Barredo Arrieta et al. 2020). We draw on this literature and introduce state-of-the-art explainable artificial intelligence to the prediction of bond excess returns. Most important for our study, Lundberg and Lee (2017) suggest the use of SHapley Additive Explanations (SHAP), a technique that uses Shapley values from game theory to identify the contribution of model features for the prediction.

Data

In line with the previous literature, we focus on the excess returns on US government bonds. We use the monthly dataset of the zero-coupon US-Treasury bond yield curve constructed by Liu and Wu (2021) that is available online.[1] In line with Bianchi et al. (2021), the sample period is from August 1971 to December 2018. From this dataset on the structure of interest rates over different maturities, we calculate bond excess returns. Based on the literature as described in the previous section, we further use two types of information to predict bond excess returns. The first type of information is the structure of interest rates. This information is directly available from the Liu and Wang (2021) dataset. The second type of information is a large panel of macroeconomic variables, as proposed by Ludvigson and Ng (2009). McCracken and Ng (2016) describe these variables in detail. The time series in the panel were selected to reflect broad categories of macroeconomic information: real output and income, employment, retail, manufacturing and sales, international trade, consumer spending, housing, inventories, orders, labor costs, capacity utilization measures, price indexes, interest rate spreads, stock market indicators, and foreign exchange rates (Ludvigson and Ng 2009). This dataset has been widely used in previous studies (e.g., Stock and Watson 2002; Ludvigson and Ng 2009).

Research Methodology

This section describes how we sample the data, the prediction method, and the methods for explainability. For the described prediction task, it is crucial that we have a temporal structure of how information

[1] https://sites.google.com/view/jingcynthiawu/yield-data.

becomes available to the decision makers. In this way, it is important that we choose a design for hyperparameter tuning and for making predictions that considers this temporal structure. We, therefore, use a rolling train, validation, and test split that adapts to newly acquired information. This is in line with recent literature on machine learning on financial market data such as Gu et al. (2020) or Bianchi et al. (2021). In line with the investment situation of a decision maker, we aim to predict the one-year bond excess return in every month based on the two information sets as described above. We start the rolling out-of-sample prediction in January 1989 and predict the bond excess return between January 1989 and January 1990. In this step, it is important to be very careful with the information that the decision maker could actually have in this situation. As the most recent bond excess return the decision maker observes initially is the one between January 1988 and January 1989, the data available for the training and validation sample reaches from August 1971 to January 1988. We use 85% of this data as the training sample and 15% as the validation sample. For each of the machine learning methods described below, we fit models based on 100 hyperparameter constellations from which we determine the best model on the validation set. The 100 hyperparameter constellations are a random sample from a larger hyperparameter space grid each. The prediction is then made on the respective hold-out month. The train-test split then moves further in a rolling approach increasing the data available for the training and validation sample in each step. The training and validation split remains constant during this process.

As common machine learning methods, we use random forests, extreme gradient boosting, and neural networks as alternative approaches. We further benchmark these results with a linear regression. The random forest is a tree-based prediction method that was suggested by Breiman (2001) and builds an ensemble of classification or regression trees. The ensemble is created by sampling parts of the data for each tree and randomly restricting the number of features that are available in each split of the tree. Extreme gradient boosting as suggested by Chen and Guestrin (2016) is an ensemble of trees as well. The algorithm subsequently builds trees that particularly address errors of trees earlier in line. The algorithm further offers regularization and delivers high performance. We then also use neural networks for the prediction task. Neural networks process inputs through subsequent layers of neurons where each neuron contains a non-linear activation function. We use a fully connected neural network, where each neuron of a layer is connected to neurons of the previous layer.

Our study suggests the use of methods from artificial explainable intelligence to derive interpretable results on what drives bond excess returns and to provide information on this to decision makers. We use SHAP values by Lundberg and Lee (2017) for this purpose. SHAP values answer the question of how a certain feature or combination of features has contributed to the prediction that a machine learning model made. To derive this contribution SHAP calculates how much a certain feature or a combination of features contributed to the specific prediction for a certain prediction target. This is done by varying the feature values while holding the other features constant. SHAP values have a nice characteristic in so far as they add up to form the overall deviation from mean cases.

Preliminary Findings

This section presents preliminary results for our study. We show exemplary analyses comparing extreme gradient boosting as one of the machine learning methods and linear regression as a linear method that was quite common before the recently increased use of machine learning in finance. These results are presented in Table 1. We further show mean absolute Shapley values for the individual prediction features in Figure 1. When assessing the results in Table 1, there is a clear distinction between the linear approach and the machine learning approach. The linear method tends to have difficulties retrieving the underlying mechanisms from the training sample. In this way, the method overfits on linear projections on the training data and produces quite unfavorable results on the test data. In fact, the R^2-values are even below zero for two-year bonds. This indicates that even using historical means as predictions is favorable compared to using predictions from linear methods. Only looking at the linear results indicates that the hypothesis that bond excess returns cannot be predicted is correct. Most interestingly, this is not the case with the extreme gradient boosting results. In this case, the R^2-values are above zero and indicate that recent machine learning methods can predict bond excess returns. We use the test statistic from Clark and West (2007) to assess statistical significance and find the positive R^2-value to be significantly larger than zero. This indicates that non-linear relationships or interactions are relevant and important to take into account. It is further worth noting that the performance of the linear method is quite far-off compared to machine learning, which indicates a great advantage of using machine learning methods.

Model	**Two-year bond excess return [R^2]**
Linear regression	-0.2738
Extreme gradient boosting	0.1057**
p-value for the null hypothesis $R^2 \leq 0$ calculated as in Clark and West (2007). We report a p-value only when the R^2 is positive (p<0.1, ** p<0.05, *** p<0.01).*	

Table 1. Benchmark between machine learning and linear methods

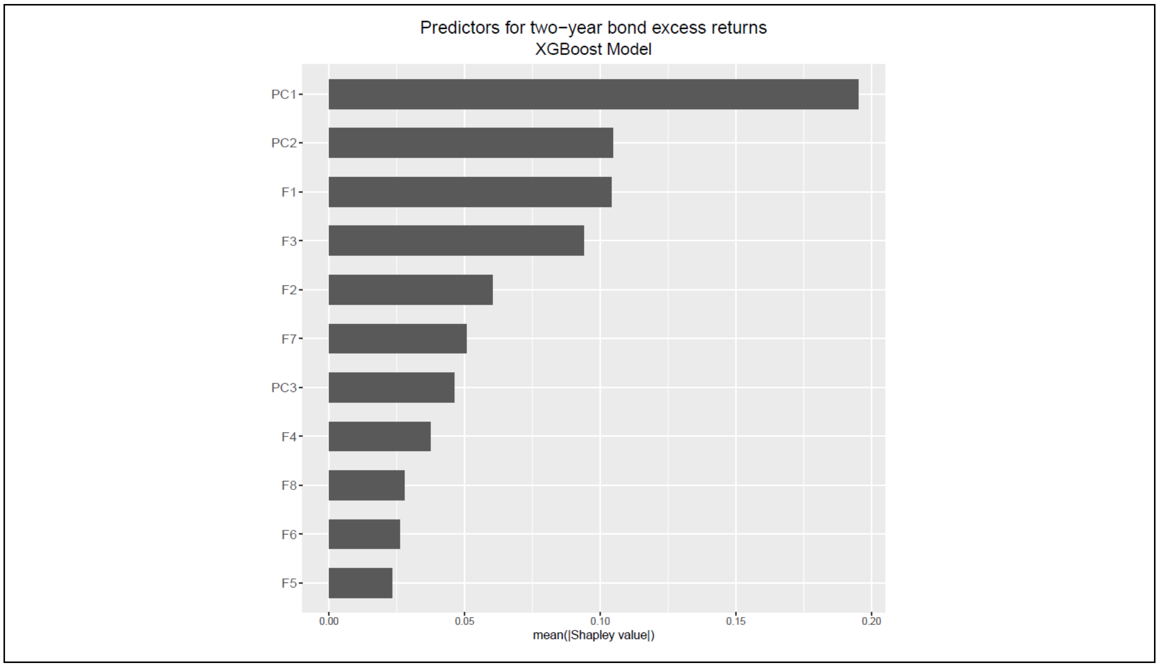

Figure 1. Mean absolute Shapley values

Figure 1 displays the mean absolute Shapley values over the features that are included in the extreme gradient boosting model. When assessing the SHAP values, it is first noteworthy that the two strongest variables to predict the two-year bond excess returns are principal components of the interest rates over different maturities. This is well in line with arguments that the bond excess return is determined by the level of interest rates (PC1) and the difference between short- and long-term interest rates (PC2) (e.g., Bauer and Hamilton 2018). Interestingly, some of the macroeconomic factors follow immediately afterward. This particularly applies to macroeconomic factor F1, which is associated with employment and production, and macroeconomic factor F3, which is associated with inflation (Ludvigson and Ng 2009). We can see that other macroeconomic factors have some relevance in decreasing order as well. This supports arguments regarding the predictability of bond excess returns based on the structure of interest rates as well as on macroeconomic information in addition to interest rate data.

Conclusion

There is a recent interest in solving prediction tasks in financial markets using machine learning methods. From a machine learning perspective, this is a problem with immense practical implications. From a finance perspective, some long-held beliefs about how markets process information and what information markets fail to process receive new perspectives in this way. More specifically, we address the problem of predicting bond excess returns which are the differences between returns on long-term bonds and short-term bonds over a common holding period. We particularly suggest applying recent advances in explainable artificial intelligence to identify which characteristics are important in predicting bond excess returns. Our results indicate that modern machine learning methods allow predictability of bond excess returns that was not present using only linear methods. Most importantly, we show that explainable artificial intelligence methods such as SHAP allow the identification of important predictors for bond excess returns in machine learning models and help to link the increase in predictability due to the use of machine learning methods to interest rate and macroeconomic characteristics.

REFERENCES

Adadi, A., and Berrada, M. 2018. "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE access* (6), pp. 52138–52160.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., and Herrera, F. 2020. "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," *Information Fusion* (58), pp. 82–115.

Bauer, M. D., and Hamilton, J. D. 2018. "Robust Bond Risk Premia," *Review of Financial Studies* (31:2), pp. 399–448.

Bianchi, D., Büchner, M., and Tamoni, A. 2021. "Bond Risk Premiums with Machine Learning," *Review of Financial Studies* (34:2), pp. 1046–1089.

Breiman, L. 2001. "Random Forests," *Machine Learning* (45:1), pp. 5–32.

Campbell, J., and Shiller, R. J. 1991. "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies* (58:3), pp. 495–514.

Chen, T., and Guestrin, C. 2016. "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.

Clark, T. E., and West, K. D. 2007. "Approximately Normal Tests for Equal Predictive Accuracy on Nested Models," *Journal of Econometrics* (138), pp. 291–311.

Cochrane, J., and Piazzesi, M. 2005. "Bond Risk Premia," *American Economic Review* (95:1), pp. 495–514.

Gu, S., Kelly, B., and Xiu, D. 2020. "Empirical Asset Pricing via Machine Learning," *Review of Financial Studies* (33:5), pp. 2223–2273.

Gu, S., Kelly, B., and Xiu, D. 2021. "Autoencoder Asset Pricing Models," *Journal of Econometrics* (222:1), pp. 429–450.

Huang, J.-Z., and Shi, Z. 2021. "Machine-Learning-Based Return Predictors and the Spanning Controversy in Macro-Finance," *Management Science*, forthcoming.

Kim, D. H., and Orphanides, A. 2007. "The Bond Market Term Premium: What is it, and How Can We Measure it?," *BIS Quarterly Review* (June), pp. 27–40.

Joslin, S., Priebsch, M., and Singleton, K. J. 2014. "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Journal of Finance* (69:3), pp. 1197–1233.

Liu, Y., and Wu, J. 2021. "Reconstructing the Yield Curve," *Journal of Financial Economics* (142:3), pp. 1395–1425.

Lundberg, S. M., and Lee, S. I. 2017. "A Unified Approach to Interpreting Model Predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 4768-4777.

Ludvigson, S., and Ng, S. 2009. "Macro Factors in Bond Risk Premia," *Review of Financial Studies* (22:12), pp. 5027–5067.

McCallum, J. S. 1975. "The Expected Holding Period Return, Uncertainty and the Term Structure of Interest Rates," *Journal of Finance* (30:2), pp. 307–323.

McCracken, M. W., and Ng, S. 2016. "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business & Economic Statistics* (34:4), pp. 584–589.

Ndikum, P. 2021. "Machine Learning Algorithms for Financial Asset Price Forecasting," *arXiv preprint*, arXiv:2004.01504.

Stock, J. H., and Watson, M. 2002. "Macroeconomic Forecasting using Diffusion Indexes," *Journal of Business & Economic Statistics* (20:2), pp. 147–162.

Wright, J. H. 2011. "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset," *American Economic Review* (101:4), pp. 1514–1534.