Group 1
12/10/2021

The Influence of Oil And Gold Price On 10-year Treasury Bond

Abstract

This study explores the dynamic interplay between oil and gold prices and their impact on the yield of the 10-year Treasury bond. Grounded in economic theories and empirical evidence, we postulate that fluctuations in oil and gold prices are significant determinants of bond yields, primarily through their influence on inflation and as indicators of global economic health. We employ a multifaceted methodological approach, incorporating VAR, ARIMA, GARCH models, and advanced machine learning techniques to analyze and forecast these relationships comprehensively. Our study aims to elucidate the direct impacts and explore the potential spillover of volatility from oil and gold markets to the Treasury market, providing insights into the broader economic implications.

1. Introduction

This research aims to understand how oil and gold prices affect the yield of the 10-year Treasury bond, an important measure in global financial markets. The significance of this inquiry stems from established economic theories and empirical evidence suggesting a strong correlation between these commodities and bond yields. In particular, oil prices have been historically linked to inflationary trends, while gold is often regarded as a hedge against inflation. These factors are hypothesized to influence bond yields directly, reflecting in investors' expectations and market behaviors.

Our study aims to analyze these relationships using a comprehensive methodological framework rigorously. Initial efforts involved a detailed literature review, identifying gaps, and setting the stage for our empirical investigation. We employed a variety of models, ranging from traditional econometric approaches such as VAR, ARIMA, and GARCH, to advanced machine learning techniques, including Decision Trees, Lasso Regression, Random Forest, Gradient Boosting, and Neural Networks. This approach was designed to assess the direct impact of oil and gold prices on bond yields and examine the potential volatility spillover from these markets to the Treasury market.

However, our findings challenge some conventional assumptions. Though widely used, traditional models fell short in accurately predicting 10-year bond yields. In contrast, machine learning models demonstrated enhanced adaptability and effectiveness in our testing dataset. Additionally, our analysis revealed that relying solely on oil and gold prices as predictive variables is inadequate. To refine our models, we propose including broader economic indicators, such as inflation rates, market volatility indices, consumer confidence, yield curve dynamics, and the influence of monetary policies.

Furthermore, the timing of our dataset split, aligning with the tumultuous economic changes of 2020, brings considerations regarding the impact of extreme market conditions on our study. These conditions may have significantly altered the typical correlations among the variables under study, thereby influencing the predictive accuracy. This observation underscores the necessity for future research to adopt more flexible approaches in dataset segmentation, potentially employing multiple splits across different periods to capture the evolving nature of these market relationships.

II. Literature Review

The selection of suitable models for forecasting 10-year bond prices is pivotal in financial analysis. Within this domain, the random walk model stands out as advocated by Baghestan (2010). Its efficacy is underscored by its recognized optimal forecasting capabilities, notably outperforming the WSJ's US Treasury rate forecasts, as empirically evidenced by Brooks, K. et al. (2004). Additionally, the exploration of machine learning algorithms, as proposed by Chen (2009), may provide alternative models that can help us forecast bond yields as well. As noted by Chen, machine learning models, including neural networks, show promise in providing forecasting accuracy while offering an alternative approach from the random walk model in the realm of bond price prediction (Baghestan, 2010; Brooks et al., 2004; Chen, 2009).

Recent research has examined potential predictors influencing bond yields, revealing varied insights into their predictive power within forecasting models. Dai et al. (2021) argues for the integration of "realized volatility" in multivariate bond yield forecasting models, showing how its inclusion in a bond yield forecasting model significantly improves the models accuracy when compared to models that do not capture volatility. Additionally, Altavilla's (2017) includes survey expectations data in their models to show its potential in accurately predicting yields. However, It is important to note the limited availability of these forecasts across all maturities.

Gupta et al. (2021) delve deeper into quantifying relationships between fundamental factors and the US Treasury term structure. Their study explores nonparametric causality in quantiles test, revealing relationships between oil supply and demand shocks, as well as financial temporal risk shocks, and the US term structure. These findings support the predictive potential of these variables in shaping the term structure, offering insights into their role in forecast models.

Alternatively, Sumner et al. (2011) challenge traditional perceptions regarding gold's role as a reliable predictor of government bond yields. Their research revealed an inconsistent relationship between gold prices and government bonds, particularly during periods of volatility spillover from the stock market. This, in turn, casts doubt on gold's price's ability to be used as a feature variable for predicting 10 year bonds. Prompting a reevaluation of its role as a safe-haven asset in forecasting government bond yields. Nonetheless, we predict that the exclusion of gold from our forecast model, would lead to omitted variable bias; hence we opted for its inclusion.

III. Theoretical Analysis

We have four models to analyze to prove the hypothesis that fluctuations in oil and gold prices have a significant and measurable impact on the yield of the 10-year Treasury bond and that volatility in oil and gold markets may spill over into the treasury market, influencing price volatility.

A: ARIMA Model

The ARIMA model captures and forecasts the trends and cycles in the 10-year Treasury yields based on past values and establishes a baseline for comparison with other models. ARIMA model has three parts:

i). AR (Autoregressive) Part

This component of the ARIMA model captures the influence of previous values in the time series on current values. It's represented by 'p' in ARIMA(p, d, q). The function form for the AR part can be represented as

$$Yt = \varphi 1 Yt\text{-}1 + \varphi 2 Yt\text{-}2 + \ldots + \varphi p Yt\text{-}p + \varepsilon t$$

Y_t is the current value, φ_i are the parameters, and ε_t is the error term.

ii). I (Integrated) Part

This part involves differencing the time series to make it stationary, indicated by 'd' in ARIMA(p, d, q). The differencing operation is represented as

$$\nabla Y_t = Y_t - Y_{t-1}$$

∇Y_t is the differenced series. Repeated differencing may be applied depending on the degree 'd'.

iii). MA (Moving Average) Part

This component models the relationship between the time series and a moving average of past forecast errors. ARIMA represents it by 'q' in (p, d, q). The MA part is formulated as follows:

$$Y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \mu + \varepsilon_t$$

θ_i are the parameters, μ is the mean of the series, and ε_t is the error term.

ARIMA models do not include external explanatory variables or predictors in their traditional form (ARIMA(p, d, q)); however, an extension known as ARIMAX allows for including external variables such as oil price and gold price.

B. VAR & VCEM Model

These two models investigate the interdependencies and the dynamic relationship between oil prices, gold prices, and Treasury yields.

i). VAR (Vector Autoregression) Model

The VAR model is a system of equations where each variable is a linear function of past values of itself and past values of all the other variables in the system. A general VAR model with 'n' variables can be represented as:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_p Y_{t-p} + \varepsilon_t$$

Y_t is a vector of variables, A_i are coefficient matrices, p is the lag order, and ε_t is the error term vector. The variables include oil prices, gold prices, and 10-year Treasury yields. Each of these variables would be explained by their own lagged values and the lagged values of the other variables in the system.

ii). VECM (Vector Error Correction Model)

VECM is used when the variables in a VAR model are non-stationary but cointegrated. The VECM form introduces an error correction term to the VAR model to capture the long-term equilibrium relationship between variables. It is typically represented as:

$$\Delta Yt = \alpha(ECMt\text{-}1) + \Sigma^{p-1}_{i=1}\Gamma i\Delta Yt\text{-i} + \varepsilon t$$

ΔYt is the difference of the variables vector, and $ECMt$-1 is the error correction term. α and Γ are coefficient matrices.

C. Machine Learning Method

Machine learning methods leverage data-driven insights and capture non-linear relationships that may be present between the variables.

i). Lasso Regression

Lasso regression is a linear model that includes a regularization term. The function form is similar to linear regression but with a penalty on the absolute size of the coefficients.

ii). XGBoost

XGBoost (eXtreme Gradient Boosting) implements gradient-boosted decision trees. It optimizes a cost function that includes a regularization term to prevent overfitting. It optimizes a cost function that includes a regularization term to avoid overfitting. The function is more complex, involving an iterative process where each new tree is built to correct the errors made by the previous ones.

iii). Random Forest

A Random Forest is an ensemble of Decision Trees, typically trained with the “bagging” method. The general idea is to combine the predictions from multiple decision trees to make more accurate predictions than any individual tree.

iv). Decision Tree

A Decision Tree is a flowchart-like structure where each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label.

v. Neural Network

Neural Networks consist of layers of interconnected nodes (neurons). Each connection has a weight that gets adjusted during the learning process. The function form depends on the architecture of the network (number of layers, types of layers, etc.). A simple feedforward neural network function can represent a series of matrix multiplications and non-linear activations.

D: GARCH Model

The GARCH model determines if shocks to oil and gold prices persistently affect the volatility of 10-year Treasury prices. The general form of a GARCH(p, q) model for a time series Yt can be written as:

$$Yt = \mu + \varepsilon t$$

$$\varepsilon t = \sigma t z t$$

$$\sigma t^2 = \alpha 0 + \Sigma_{i=1}^{q} \alpha i \varepsilon_{t-i}^{2} + \Sigma_{j=1}^{p} \beta j \sigma_{t-j}^{2}$$

Yt is the observed variable at time t, σt^2 is the conditional variance, zt is the standard normal error term, αi are the coefficients for the lagged squared residuals, and βj are the coefficients for the lagged conditional variances. We chose the logarithm of the gold and oil price instead of the raw data because it helps stabilize the variance across time, which benefits models that focus on volatility. We also choose bond returns to represent the changes in bond prices from one period to the next. These returns capture the movement and variability in bond prices over time, directly measuring price volatility.

IV. Empirical Analysis

A: Data

The study chooses the daily data from 2008 to 2023 to analyze. Market yield on U.S. treasury securities at 10-year constant maturity, quoted on an investment basis (Bond/DSG), and crude oil prices are from Federal Reserve Economic Data (FRED). The gold price is from Yahoo Finance.

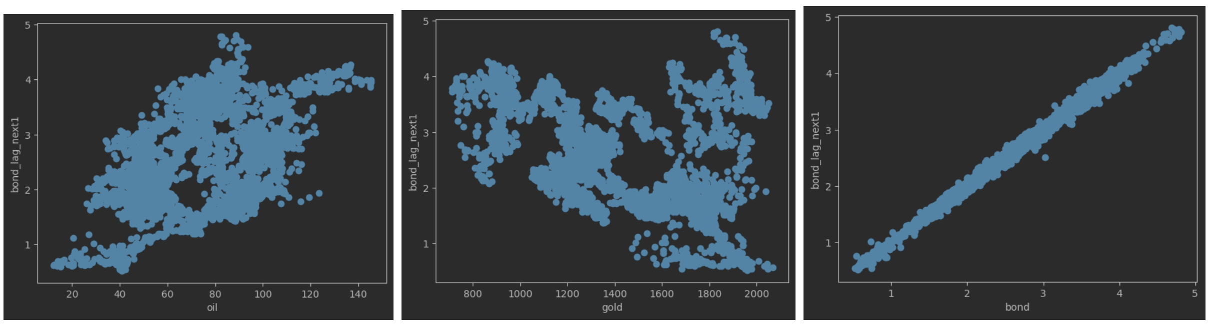
i). Summary Statistics

Variable	Mean	Standard Deviation	Kurtosis	Skew
Bond	2.480739	0.850786	-0.50	0.04
Oil	72.68687	23.430779	-0.67	0.17

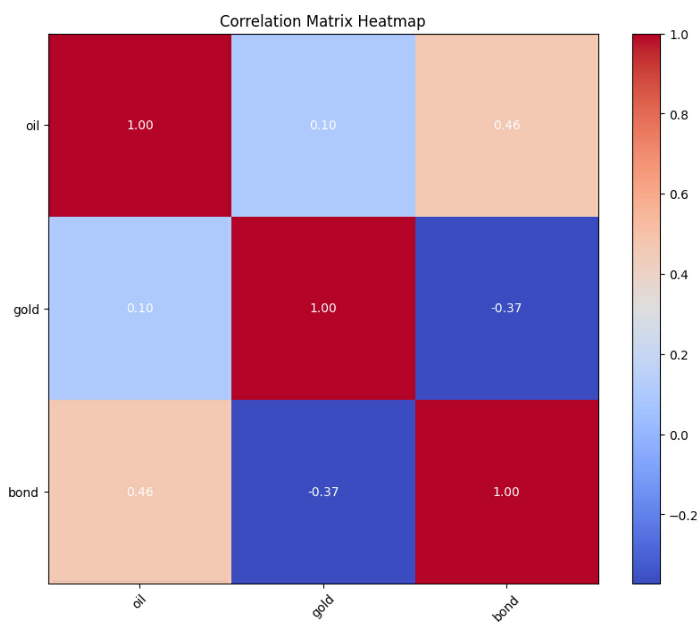
Gold	1409.230619	309.66498	-0.90	0.13

The dataset comprises 4,089 observations. The predictive variables are the bond, gold, and oil price lags. Depending on the model employed, these were tested in both raw levels and returns. The respondent variable is the bond yield, either in its level or return form, depending on the specific model used.

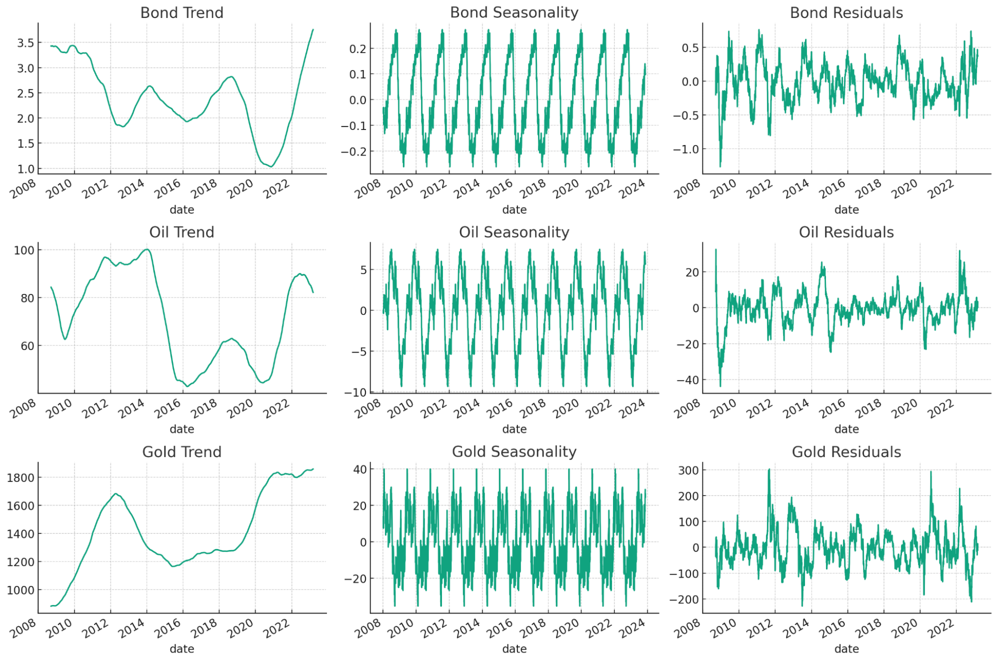
Below are the scatterplots of predictors to the next period's value of bond price.



Based on these plots, we see that bonds and oil have a positive correlation, bonds and gold seem to have a negative correlation that is more noisy, and bonds have a positive autocorrelation. And below is the contemporaneous price heatmap, which matches these findings:



ii). Trend, Seasonality, and Residuals



The trend shows some fluctuation for bond, oil, and gold prices, with a notable dip around 2016 before rising again. Seasonality represents the seasonal effect or the recurring fluctuations at regular intervals. A consistent pattern repeats yearly for bond, oil, and gold prices, suggesting a vital seasonal component. Residuals are the irregularities or the 'noise' left after the trend and seasonality have been removed from the data, representing the short-term unexplained variance. Due to market shocks or other unpredictable factors, volatility exists in bond, oil, and gold prices.

iii). Stationary

Original Data	Test Statistic	P-value
Bond	-1.8895	0.3370
Oil Price	-2.5969	0.0936
Gold Price	-1.2142	0.6675
Oil_log	-2.93	0.042
gold_log	-1.469	0.549

All original data p-value, excluding oil_log (the natural logarithm of the oil prices), is more significant than 0.05, which is non-stationary. Oil_log is stationary, but it is very close to the threshold. The original data need to be differenced to process the model, a data transformation technique used in time series analysis, primarily to remove a time series' trends and seasonality, making it stationary.

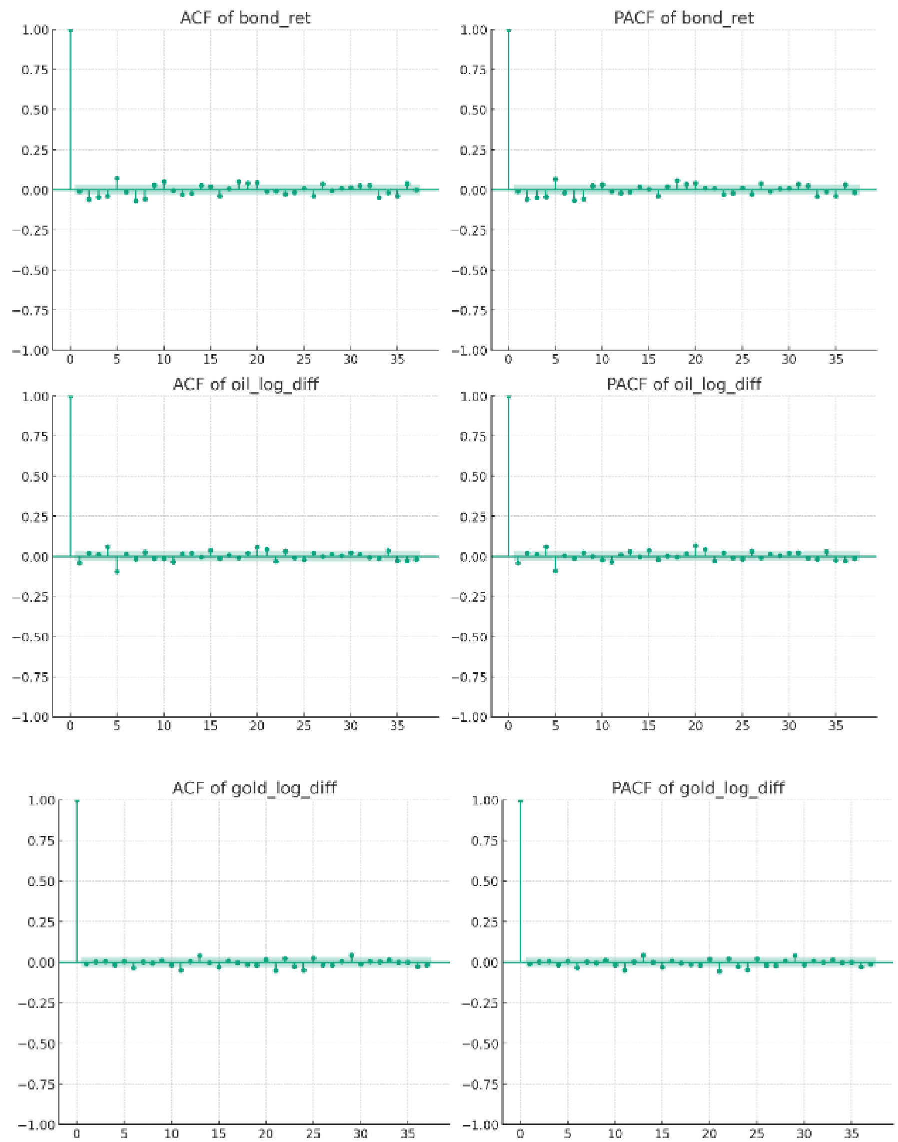
Differenced Data	Test Statistic	P-value
Bond_diff	-12.6067	1.69e-23
Oil_diff	-17.2637	5.92e-30
Gold_diff	-14.5240	5.47e-27
Oil_log_diff	-11.992	3.50e-22
gold_log_diff	-12.8445	5.52e-24
Bond_ret	-11.6967	1.60e-21

All differenced data and bond return p-values are smaller than 0.05, which is stationary.

B: Presentation and Interpretation of Results

i). GARCH Model

a). ACF &PACF



Since the autocorrelation function (ACF) or partial autocorrelation function (PACF) spikes for individual lags lie within confidence bounds(the green shaded area around zero), the autocorrelation at those lags is not statistically significant from zero at the 5% significance level. These plots suggest that the bond returns and differenced logarithm of the gold price and oil price do not have significant autocorrelation, satisfying one of the requirements for using a GARCH model. This implies that any observed volatility clustering is not due to autocorrelation in the mean but potentially due to autocorrelation in the variance, which the GARCH model aims to capture.

The fitted model is a standard GARCH(1,1) with no ARIMA components in the mean equation (ARFIMA(0,0,0)). The distribution used is the standard t-distribution (std). The standard GARCH(1,1) model is fundamental for capturing volatility clustering in financial time series. It's often the first model used in empirical studies due to its simplicity and effectiveness in many scenarios. Starting with a simpler model makes interpreting the results and understanding the data dynamics easier. The decision not to include ARIMA components suggests that the time series mean (bond returns) does not show significant auto-regressive (AR) or moving average (MA) behaviors, which is reflected by ACF and PACF plots.

b). Model Results

LogLikelihood	10017.45			
Parameter	Estimate	Std. Error	t value	Pr(>|t|)
mxreg1	0.213351	0.014774	14.44056	<0.00001
mxreg2	-0.368541	0.033701	-10.9355	<0.00001
alpha1	0.056761	0.014450	3.92801	0.000086
beta1	0.936643	0.016646	56.26841	<0.00001

The coefficient for the differenced logarithm of oil price (mxreg1) is positive and statistically significant, indicating that increases in oil price volatility are associated with increased bond price volatility. The coefficient for the differenced logarithm of the gold price (mxreg2) is negative and statistically significant, suggesting that increases in gold price volatility are associated with decreased bond price volatility. Parameters of the GARCH model, alpha1 and beta1, are significant, indicating the presence of volatility clustering in bond yields. The log-likelihood value is high, suggesting a good model fit. These coefficients are highly significant, with p-values effectively at zero. This means a strong relationship between oil and gold market volatility and treasury bond yield volatility.

Weighted Ljung-Box Test on Standardized Residuals:

Lag	Statistic	p-value
Lag[1]	9.188	0.002436
Lag[2*(p+q)+(p+q)-1][2]	9.610	0.002424
Lag[4*(p+q)+(p+q)-1][5]	10.136	0.008494

Weighted Ljung-Box Test on Standardized Squared Residuals:

Lag	Statistic	p-value
Lag[1]	8.864	0.002908
Lag[2*(p+q)+(p+q)-1][5]	12.227	0.002428
Lag[4*(p+q)+(p+q)-1][9]	21.799	0.00006862

Weighted ARCH LM Tests:

ARCH Lag	Statistic	Shape	Scale	P-Value
Lag[3]	0.004619	0.500	2.000	0.945816
Lag[5]	5.754454	1.440	1.667	0.068590
Lag[7]	13.704906	2.315	1.543	0.002401

Sign Bias Test:

Test	t-value	p-value	Significance
Sign Bias	3.925	0.00008832	***
Negative Sign Bias	6.397	0.0000001756	***
Positive Sign Bias	1.210	0.2263	
Joint Effect	42.496	0.00000003149	***

Adjusted Pearson Goodness-of-Fit Test:

Group	Statistic	p-value
20	45.99	0.0004976
30	60.69	0.0005070
40	95.91	0.00001069
50	112.59	0.0000006506

The Ljung-Box tests on standardized and squared residuals show significant p-values, indicating that there might still be some autocorrelation in the residuals. The ARCH LM test indicates potential remaining ARCH effects, especially at lag 7. The Sign Bias Test results indicate the presence of a sign bias in the model, particularly a negative sign bias. The Pearson Goodness-of-Fit test shows significant p-values, suggesting that the model may not perfectly capture the returns distribution. The presence of autocorrelation and ARCH effects in the residuals and the negative sign bias indicate that the model could be improved by considering different lag structures or using a more complex GARCH variant.

ii). ARIMA Model Performance and Selection

a). ARIMA(1,1,0):

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                  DSG10   No. Observations:                  127
Model:                 ARIMA(1, 1, 0)   Log Likelihood                  21.828
Date:                Mon, 27 Nov 2023   AIC                            -39.657
Time:                        23:37:06   BIC                            -33.984
Sample:                    01-01-2008   HQIC                           -37.352
                         - 07-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.2265      0.074      3.061      0.002       0.081       0.372
sigma2         0.0414      0.003     13.458      0.000       0.035       0.047
===================================================================================
Ljung-Box (L1) (Q):                   0.10   Jarque-Bera (JB):               152.53
Prob(Q):                              0.75   Prob(JB):                         0.00
Heteroskedasticity (H):               0.31   Skew:                            -1.15
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.87
===================================================================================
```

This model showed a log likelihood of 21.828 with AIC and BIC values of -39.657 and -33.984, respectively. The absence of MA (Moving Average) and additional AR (AutoRegressive) terms indicates a simple model focusing on the differencing aspect to render the series stationary.

ARIMA(1,1,1): Exhibiting a higher log likelihood of 24.200, this model combines both AR and MA components. The improvement in AIC (-42.401) and BIC (-33.892) over the ARIMA(1,1,0) model suggests a better fit. The AR coefficient at -0.3944 (p=0.046) indicates a slight but significant negative correlation with the previous value, while the MA coefficient at

0.6784 (p < 0.001) suggests a stronger and highly significant influence of the moving average component.

ARIMA(0,1,1) (listed twice with identical values): This model, with a log likelihood of 22.894 and AIC and BIC values of -41.787, focuses solely on the MA component (coefficient of 0.3076, p < 0.001). Its performance metrics are relatively close to those of ARIMA(1,1,0), but with a significantly better fit indicated by the lower AIC and BIC values.

b). Error Metrics

The Mean Absolute Error (MAE) of 1.087, Mean Squared Error (MSE) of 1.636, and Root Mean Squared Error (RMSE) of 1.279, along with the high Mean Absolute Percentage Error (MAPE) of 84.21%, suggest that the model, while statistically significant, may have limitations in accurately forecasting the 10-year Treasury bond yields. The high MAPE, in particular, indicates a considerable average deviation between the predicted and actual values.

Interpretation and Implications

c). The ARIMA(1,1,1)

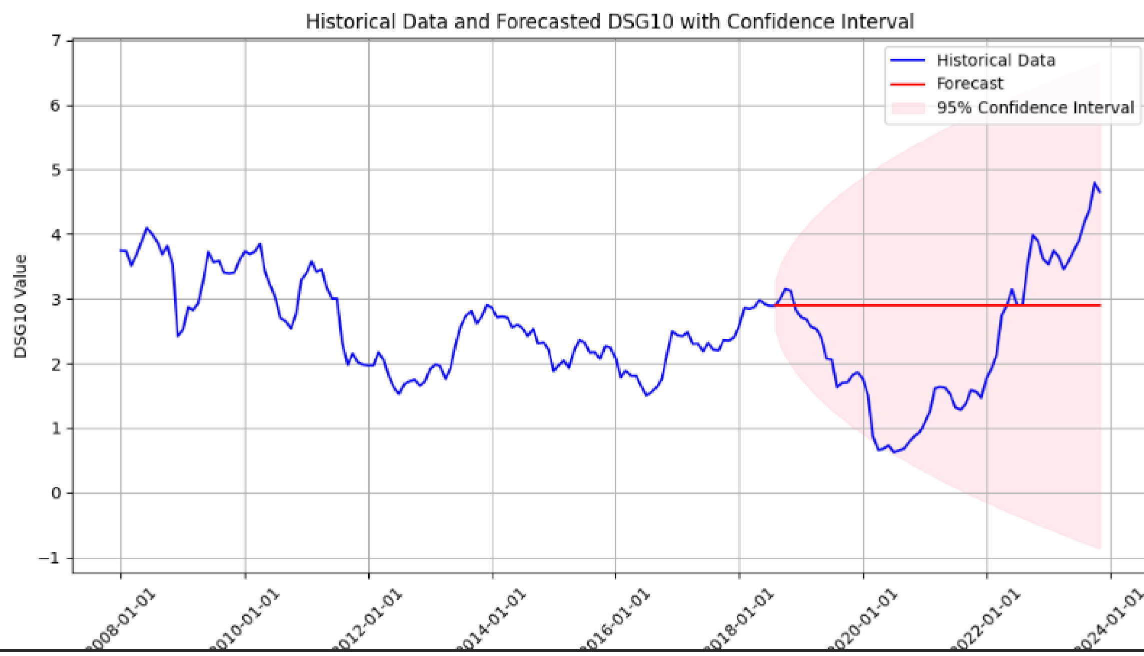
```
                               SARIMAX Results
==============================================================================
Dep. Variable:                  DSG10   No. Observations:                  127
Model:                 ARIMA(1, 1, 1)   Log Likelihood                  24.200
Date:                Mon, 27 Nov 2023   AIC                            -42.401
Time:                        23:37:07   BIC                            -33.892
Sample:                    01-01-2008   HQIC                           -38.944
                         - 07-01-2018
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.3944      0.198     -1.994      0.046      -0.782      -0.007
ma.L1          0.6784      0.168      4.031      0.000       0.349       1.008
sigma2         0.0398      0.004      9.730      0.000       0.032       0.048
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):               100.88
Prob(Q):                              0.98   Prob(JB):                         0.00
Heteroskedasticity (H):               0.34   Skew:                            -1.09
Prob(H) (two-sided):                  0.00   Kurtosis:                         6.81
===================================================================================
```

model's combination of AR and MA components seems to provide a more nuanced understanding of the time series data, capturing both the immediate past value's influence and the moving average trend.

The statistical significance of the AR and MA coefficients in the ARIMA(1,1,1) model suggests that both past values and moving average trends are important in understanding the fluctuations in the 10-year Treasury bond yields.

The presence of a negative AR coefficient and a positive MA coefficient could indicate a tendency for the bond yields to adjust in the opposite direction of their previous value while maintaining a trend influenced by recent averages.

d). Recommendations for Future Research



Given the limitations of the high MAPE, future research might explore more complex ARIMA models or integrate additional variables that could affect bond yields, such as macroeconomic indicators or external market factors.

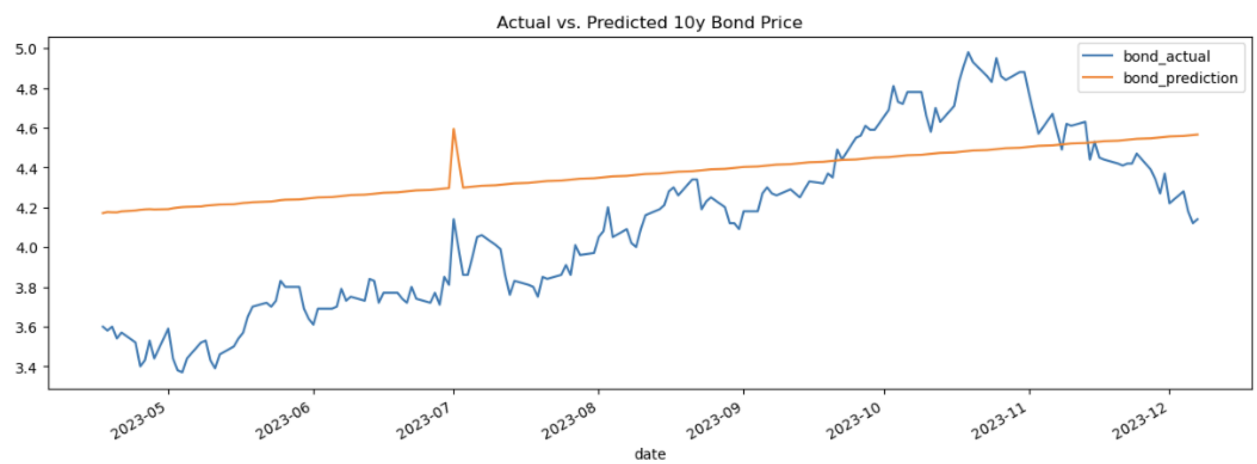
The analysis could be extended to include a broader range of historical data or to consider the impact of specific economic events on model performance.

iii) Vector AutoRegressive (VAR) model

Model	MSE	MAPE	MAE
VAR(1)	0.31	13.00	0.49
VAR(7)	0.32	13.38	0.51
VAR(14)	0.34	10.67	0.46

The multivariate VAR model performed slightly better than the univariate models. We tested for different lags (1, 2, 14) on train test splits to evaluate their out-of-sample performance. The best-performing VAR model had 14 lags and a mean absolute percentage error of 10.67%. However, there may have been model misspecifications as the determinant of the maximum likelihood estimate of the error covariance matrix is 0.00017, potentially suggesting issues related to multicollinearity or model specification. Most of the model's parameters were largely insignificant; however, to capture longer-term dynamics in the model that could have helped estimate 10-year bond yields, we included larger lags.

To decide the model lags, common lags that may have yielded accurate short-term predictions were chosen: lag 1 and lag 7. Lag 14 was chosen because it minimized AIC, which offers a trade-off between goodness of fit of the model and model complexity and outperformed the other models.



For future research, one should explore a more comprehensive VAR model that includes inflation, which is more closely correlated with 10y treasury bond yields, consumer sentiment, which can help explain the noise in the relationship between gold and bond yields, and last market volatility as it can significantly impact bond yields. This expanded analysis could offer deeper insights into the intricate interplay among economic indicators and market variables, enriching our understanding of bond yield determinants and improving forecast accuracy for our VAR models.

iv) Machine Learning models

For machine learning models, the features tested in all models were bond, oil, and gold price data lagged at 1, with the responder again being bond price data. The models were run with both returns, calculated as $\frac{price_{t-1}}{price_t} - 1$, and raw price levels. For all the machine learning models, the first chronological 80% of the data was used in training, and the last 20% used for testing. In model fitting, time-series cross validation was used so as to avoid overfitting the models. For evaluating the results, we check Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percent Error (MAPE). All of these metrics are calculated using the test set only. For returns, MAPE is not applicable since many times we would be dividing by 0 to calculate that.

a) XGBoost

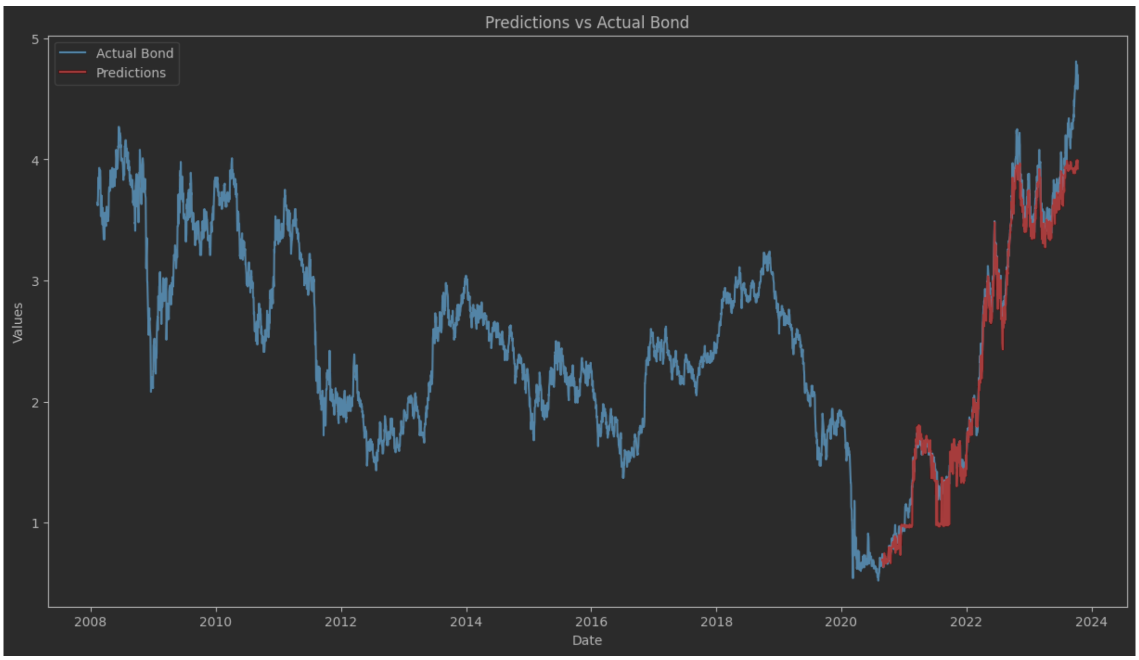
The first machine learning model tested was XGBoost. A hyperparameter sweep was performed using GridSearch, with the following parameters tested:

num_estimators	max_depth	learning_rate	subsample
50	3	0.01	0.8
100	5	0.1	0.9
200	7	0.2	1

The best combination found is the one highlighted in green. Using this model, the results were as follows:

Responder	RMSE	MAE	MAPE
1-Day Returns	0.0421	0.0275	N/A
Levels	0.1727	0.1172	5.2864

The prediction graph for this model is visualized:



b) Decision Tree

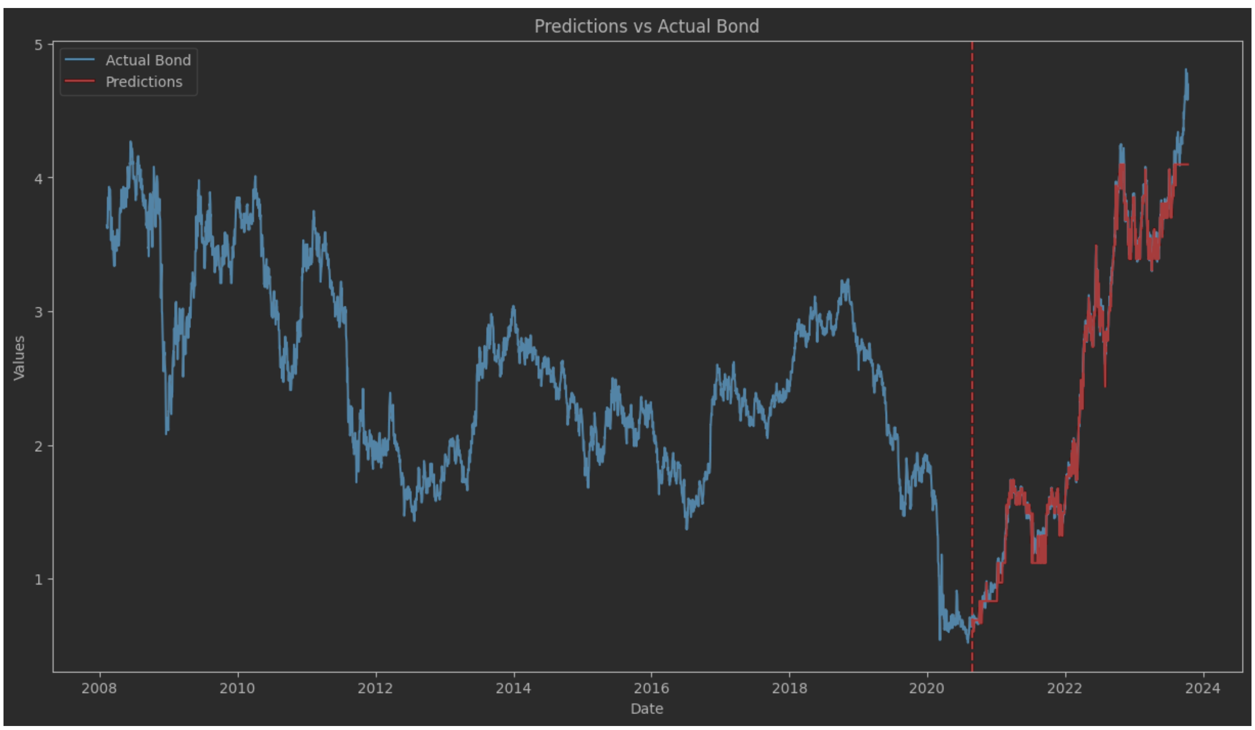
Decision tree was tested similarly to XGBoost, with the following values tested for hyperparameter tuning:

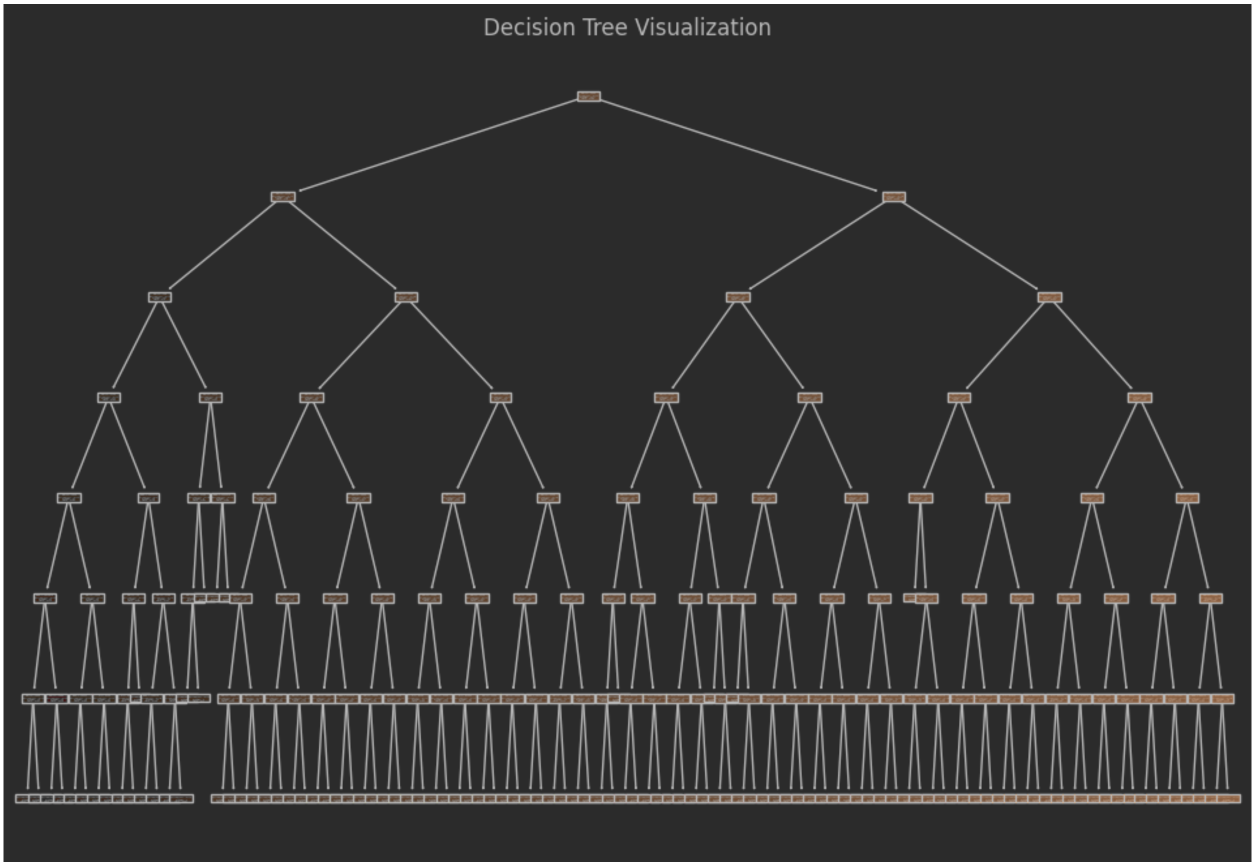
max_depth	min_samples_split	min_samples_leaf
3	2	1
5	5	2

7	10	4
10		
20		

Again the best combination is highlighted. The results summary for this model is below.

Responder	RMSE	MAE	MAPE
1-Day Returns	0.0405	.0255	N/A
Levels	0.1208	.07891	3.79





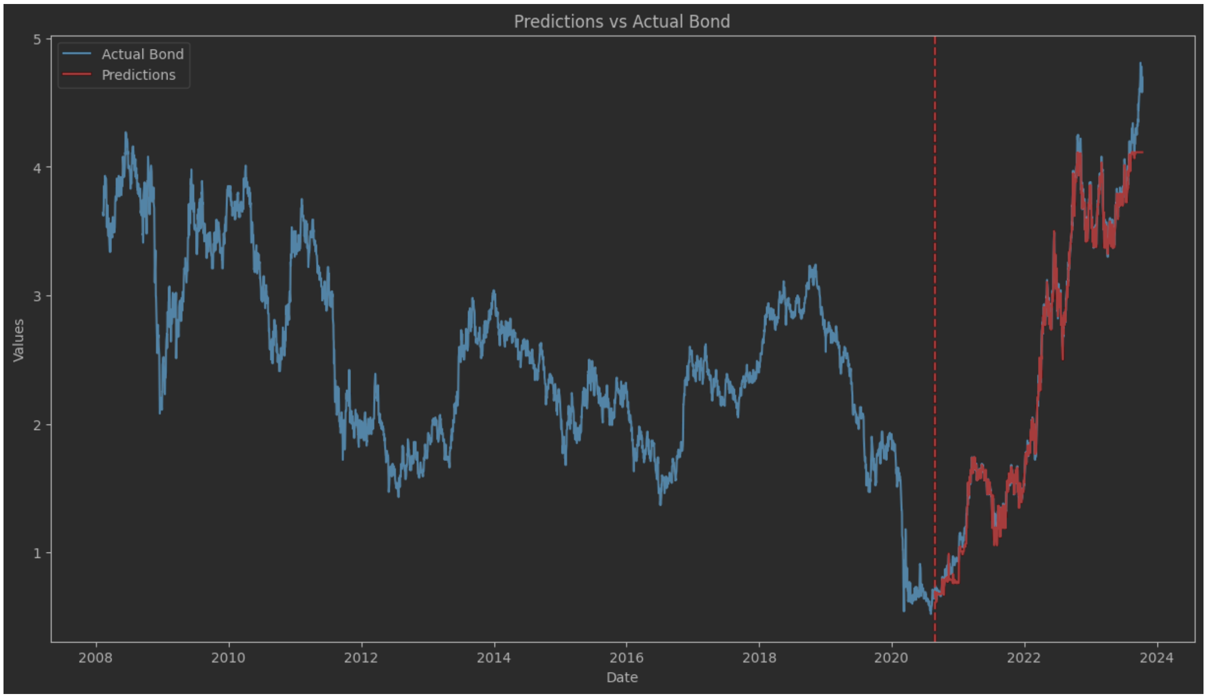
Above is a visualization of this decision tree. The labels are too small to read, however, it gives a good idea of the overall structure used.

c) Random Forest

The same process was done with the random forest, with the following hyperparameters used:

num_estimators	max_depth
50	3
100	5
200	7

And the projection is as follows:



Interestingly, for all of the model predictions thus far, from the graph we see that the model does a pretty good job at predicting, but then towards the end seems to flatline instead of increasing as the actual prices do.

d) Lasso Regression

For Lasso Regression, we did not perform any hyperparameter sweeps. Instead, we used the default model, which was already better than the other models in terms of the results, summarized below:

Responder	RMSE	MAE	MAPE
1-Day Returns	0.02853	0.02109	N/A
Levels	0.06277	0.04748	2.15

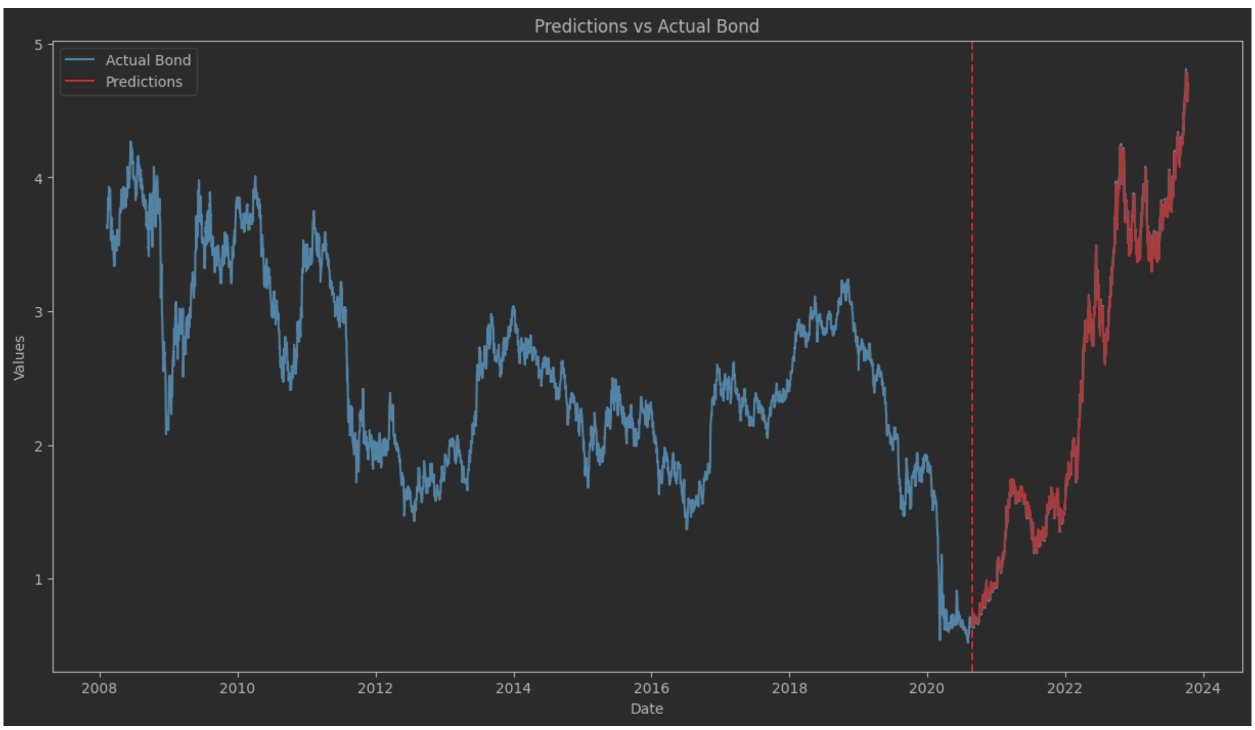
Interestingly, the Lasso Regression performed relatively the same for predicting returns but was a lot better at predicting raw price levels, which is the main focus of this project.

Following is the table of coefficients for each variable:

Variable	Coefficient
oil_lag_1	.0126
gold_lag_1	-.0017
bond_lag_1	.2040

We see that the coefficient of lagged bonds is the most important, which is as expected. We also see that the coefficients match up with the expectations from the scatterplots we presented earlier, with oil and bonds being positively correlated but gold and bonds being negatively correlated.

Therefore the predictions visualization is as follows:



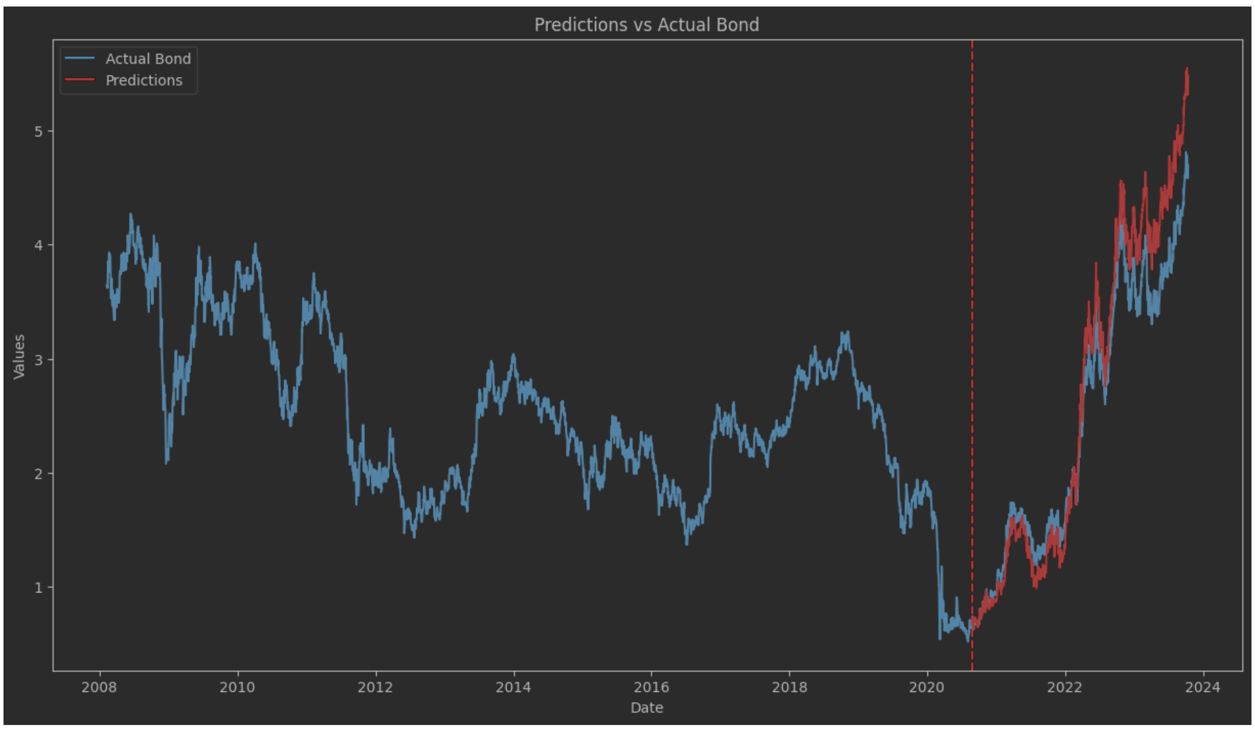
Something interesting to note is that the predictions for Lasso Regression, unlike the other models thus far, correctly predicts the uptick at the end as opposed to flatlining.

e) Neural Network

The Neural Network was implemented using a Multilayer Perceptron (MLP). Below is the table of the results.

Responder	RMSE	MAE	MAPE
1-Day Returns	0.02854	0.02113	N/A
Levels	0.3559	0.2813	10.48

It was about the same at predicting for returns, but for levels, it was much worse than the other models.



We see that at the end it overpredicts a lot, although it gets the general direction correct (i.e. if we were to shift down the predictions at the end, it seems like they would be pretty close).

V. Conclusion.

In investigating the use of gold and oil to predict 10-year treasury prices, multiple models from different categories were tested and compared. In terms of general model performance, it was observed that traditional models such as ARIMA, GARCH, and VAR fell short in predicting bond yields, and were outdone by machine learning models. Overall, for the machine learning models, we saw that the Neural Network was the worst, and had multiple times the error values of most of the other models. The reason for this is probably because the data was quite noisy. For instance, the relationship between gold and bonds can be explained by a confounding variable, inflation, which is highly correlated with both of them. Random Forest, Decision Tree, and XGBoost were all relatively similar, with XGBoost being the worst of the three. This is expected since they are all related in terms of implementation and all use decision trees as part of the model. The best model was Lasso Regression, which is the simplest machine learning model we used.

To enhance the models, other factors could be included, such as inflation rates, market volatility indices, the S&P index, consumer confidence index, yield curve dynamics, and monetary policy influence. Additionally, rolling windows could be used for model fitting, for instance, a rolling regression as opposed to just fitting the coefficients to the training dataset and using it on the testing dataset. Another possible way to improve the models would be to select the training and testing sets more carefully, as right now the split somewhat coincided with the major economic changes spurred by COVID in 2020.

References

Altavilla, C., Giacomini, R., & Ragusa, G. (2017). Anchoring the yield curve using survey expectations. Journal of Applied Econometrics., 32(6), 1055–1068. https://doi.org/10.1002/jae.2588

Baghestani, H. (2010), Forecasting the 10-year US Treasury rate. J. Forecast., 29: 673-688. https://doi-org.libproxy2.usc.edu/10.1002/for.1165

Baur, D. G., & McDermott, T. K. J. (2016). Why is gold a safe haven? Journal of Behavioral and Experimental Finance, 10, 63–71. https://doi.org/10.1016/j.jbef.2016.03.002

Brooks, R., & Gray, J. B. (2004). History of the forecasters. Journal of Portfolio Management, 31(1), 113-117. Retrieved from http://libproxy.usc.edu/login?url=https://www.proquest.com/scholarly-journals/history-forecasters/docview/195596612/se-2

Chen. (2009). The prediction of Taiwan 10-year government bond yield. WSEAS Transactions on Systems, 8(9).

Dai, Z., & Kang, J. (2021). Bond yield and crude oil prices predictability. Energy Economics, 97, 105205-. https://doi.org/10.1016/j.eneco.2021.105205

Gupta, R., Shahzad, S. J. H., Sheng, X., & Subramaniam, S. (2021). The role of oil and risk shocks in the high‐frequency movements of the term structure of interest rates: Evidence from the U.S. Treasury market. International Journal of Finance & Economics, 28(2), 1845–1857. https://doi.org/10.1002/ijfe.2511

Sumner, Steven and Johnson, Robert and Soenen, Luc, Spillover Effects Among Gold, Stocks, and Bonds (April 8, 2011). Journal of CENTRUM Cathedra, Vol. 3, Issue 2, pp. 106-120, 2010, Available at SSRN: https://ssrn.com/abstract=1806036