Predicting the Net Asset Value of Distressed Bonds Using Machine Learning Techniques

Camila Asprilla

Writing Sample

May 21, 2024

Abstract:

This paper analyzes the price movements of the iShares Convertible Bond ETF by BlackRock (ticker: ICVT) and proposes a machine learning-based approach to predict its Net Asset Value (NAV), which represents the accumulated value of each holding in the security. The ICVT ETF is distinct due to its inclusion of debt securities issued by companies in financial distress, often referred to as "junk bonds," rated by Moody’s from C to Ba1. These bonds carry a higher risk of default but compensate investors with yields above current market returns. Convertible bonds combine features of both debt and equity components, presenting complex and challenging price dynamics.

By analyzing historical price data and using multiple features, this study applies multiple linear regression to develop predictive models for the NAV of ICVT. Regression models such as Random Forest, Simple Linear Regression, and ensemble methods are employed to capture the relationships between input variables and the ETF's price movements. Machine learning-based predictive models provide valuable insights into the factors driving performance, offering investors a tool for making informed decisions.

1. Introduction

Background:

Distressed debt arbitrage involves purchasing the debt of financially distressed companies that are unable to meet their debt obligations in the short term. Investors in this strategy buy debt securities at a discount and profit from the potential recovery or restructuring of these companies. This requires extensive analysis of various financial and operational metrics to identify lucrative opportunities. This paper focuses on predicting the price of a security comprising 334 holdings within the iShares Convertible Bond ETF (ICVT).

Most fixed-income securities in the ICVT ETF are issued by technology services companies, with Palo Alto Networks Inc., a leading cybersecurity firm, being the largest holding. Despite the growth potential, many of these companies are at a maturity stage where their financial stability is uncertain. To finance ongoing growth, these companies issue debt securities, often including a conversion feature allowing the debt to be converted into equity at a predetermined rate.

Institutional investors like BlackRock manage these convertible fixed-income securities within a diversified investment pool using strategies such as reinvesting the principal repayment into new fixed-income securities upon maturity. This approach helps maintain portfolio balance and ensure a continuous stream of income with consistent returns. Convertible bonds provide regular interest payments or "coupon payments," offering bondholders a steady income stream. Unlike regular corporate bonds, convertible bonds provide investors with the flexibility to convert their securities into shares of the issuing company. This provides investors with a choice between receiving a steady interest income with the option of principal repayment upon maturity or holding shares of stock, which offer the potential for capital appreciation and dividend payments.

Objectives:

The primary objective of this paper is to analyze and predict the Net Asset Value (NAV) of the iShares Convertible Bond ETF (ICVT) applying a supervised machine learning algorithm strategy.

Thesis Statement:

This paper aims to develop a robust machine learning model utilizing multiple linear regression to predict the Net Asset Value (NAV) of the iShares Convertible Bond ETF (ICVT) using nine years of historical data up to May 17, 2024.

2. Methodology

Research Design:

The research design incorporates statistical analysis, alongside financial metrics, to construct multiple features essential for the model.

Variables:

1. **Dependent Variable:**
    - Net Asset Value (NAV): Comprises the 334 holdings in the ETF.
2. **Independent Variables:**
    - **Dates:** Fixed values.
    - **Lags:** Additional features calculated using iterations of the current data, estimates.
    - **Rolling Average:** Calculates NAV over specified periods (5 and 10 days) to smooth price fluctuations and identify long-term trends.

3. Data Collection

Sources:

The dataset utilized for this analysis is sourced directly from the BlackRock webpage dedicated to the ETF. BlackRock provides comprehensive metrics, insights, and numerical data to ensure transparency with their investors.

4. Data analysis with R:

1. Load and clean the data:

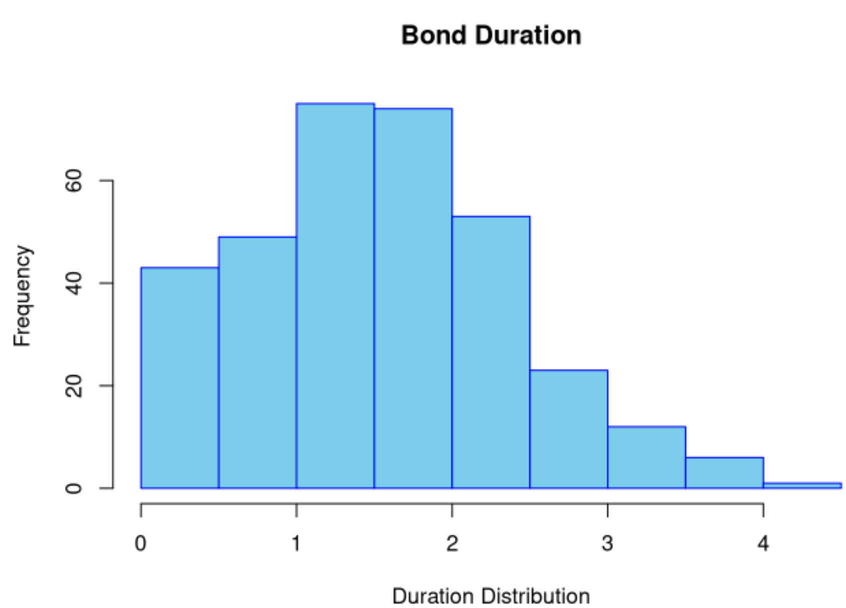
```
#Load the data
data <- read.csv("iShares_ICVT_Holdings.csv")

#Display the data
head(data)
```

	Name <chr>	Sector <chr>	Asset.Class <chr>	Market.Value <chr>	Weight.... <dbl>
1	PALO ALTO NETWORKS INC	Technology	Fixed Income	32,309,594.18	1.63
2	ROYAL CARIBBEAN GR	Consumer Cyclical	Fixed Income	28,946,002.16	1.46
3	WESTERN DIGITAL CORPORATION 144A	Technology	Fixed Income	21,679,413.30	1.09
4	MICROSTRATEGY	Technology	Fixed Income	21,247,616.98	1.07
5	FORD MOTOR COMPANY	Consumer Cyclical	Fixed Income	20,068,070.99	1.01
6	PG&E CORPORATION 144A	Electric	Fixed Income	19,549,003.54	0.99

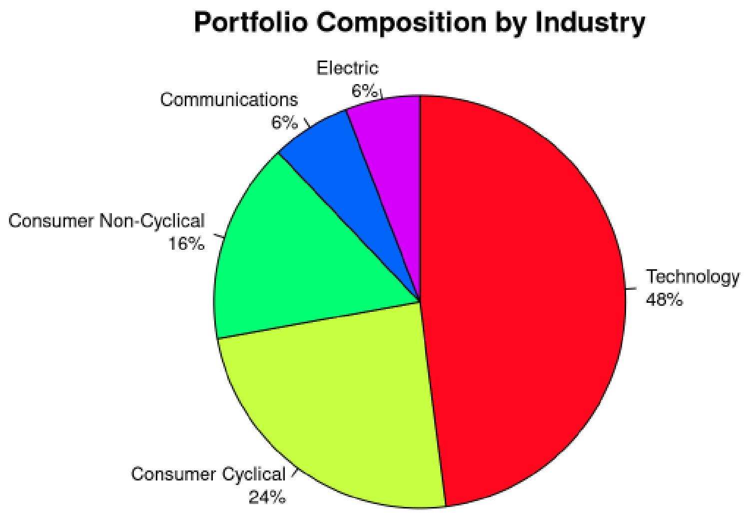
6 rows | 1-6 of 24 columns

2. Histogram: Data distribution of the bond duration in the ETF

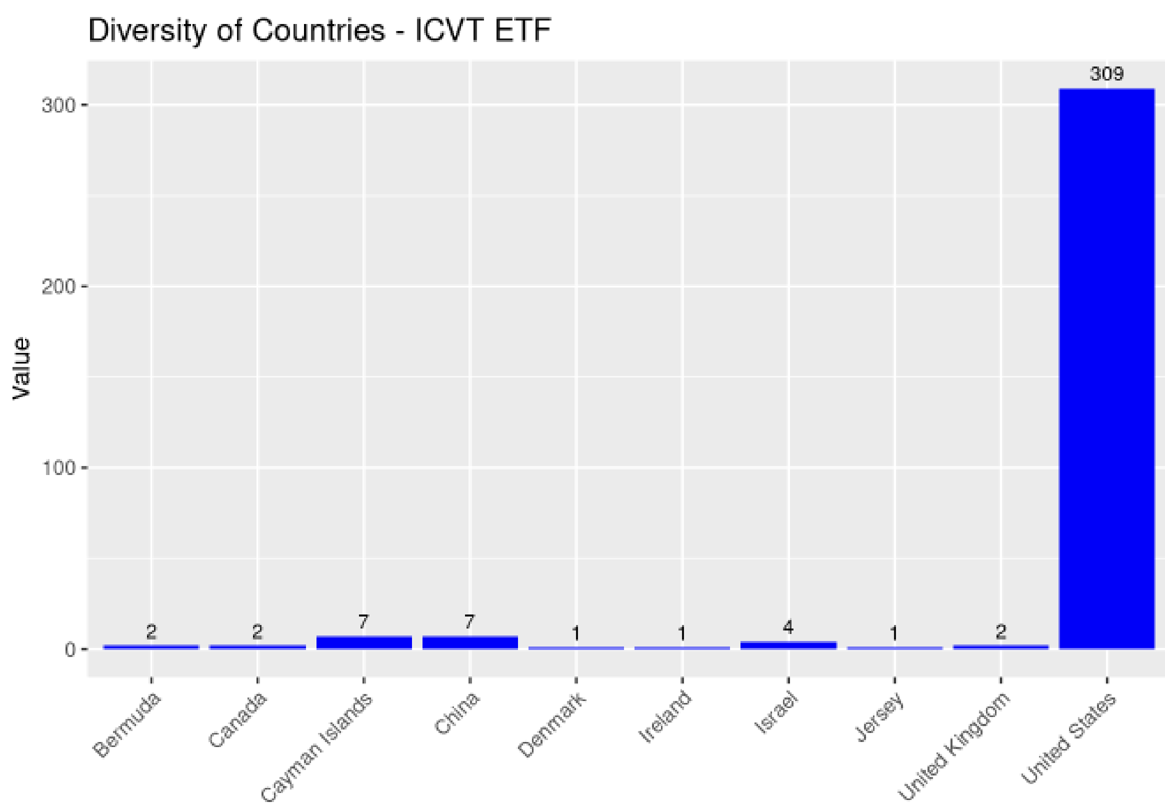


The histogram showcases how most of the bonds in the ETF will mature within 1 and 2.5 years.

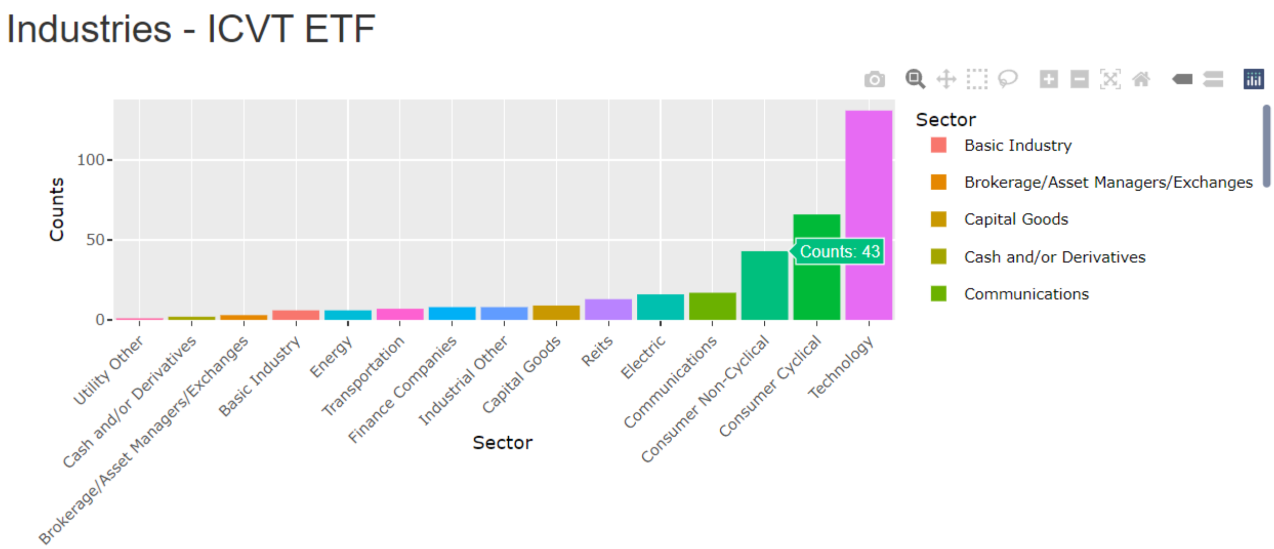
3. Pie chart: Top 5 sectors in the ETF



4. Bar graph: Geographic diversification of the ETF

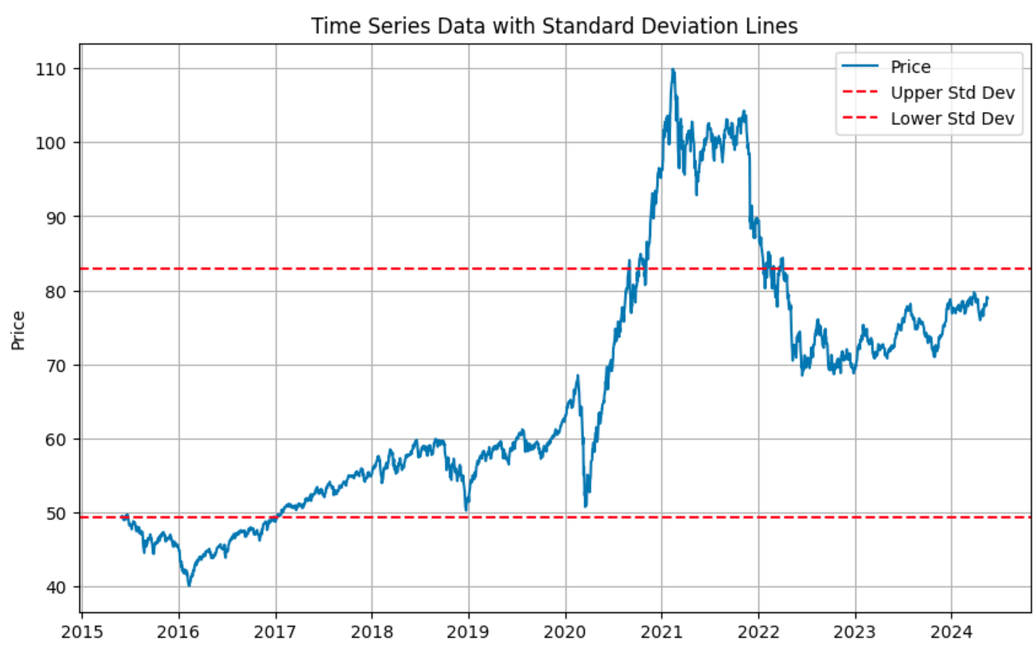


5. Interactive R-shiny app to showcase all the industries in the ETF:



6. **Machine Learning Modeling with Python:**

1. Line chart: NAV price behavior from 2015-2024 (as of May 17)



The standard deviation provides insight into the variability of the data. In this context, the ETF NAV has fluctuated between $50.00 and approximately $80.00 since 2017.

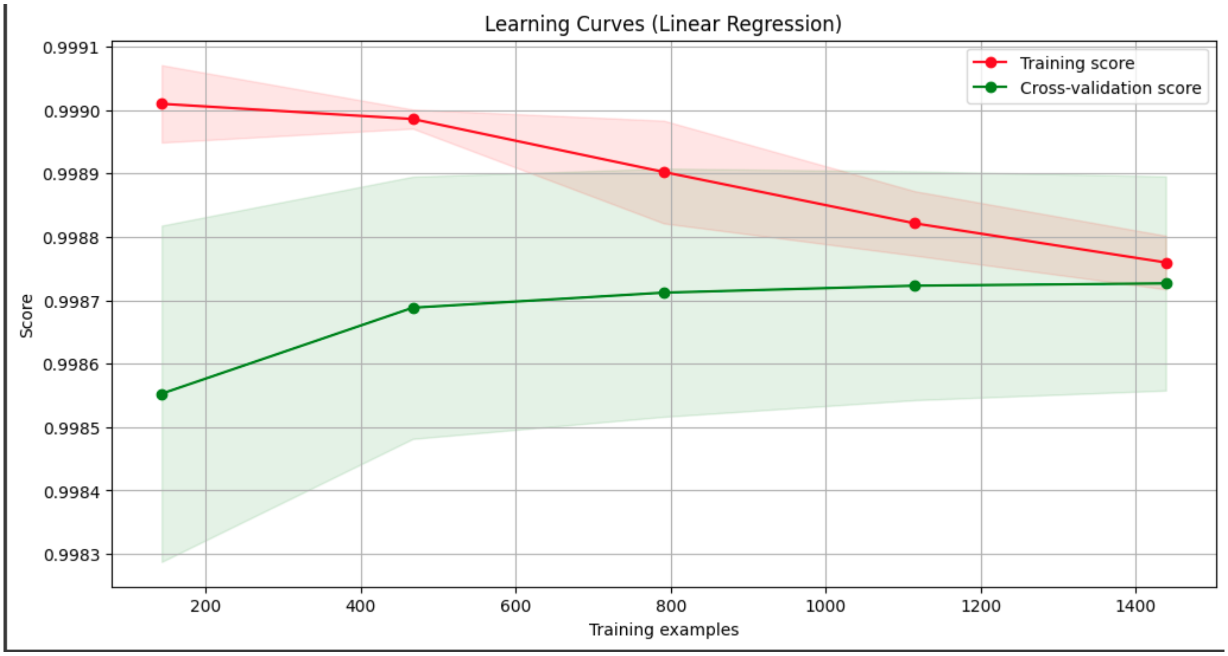
7. Results:

The table below presents the actual prices from the dataset alongside the predictions generated using the multiple linear regression algorithm. Additionally, it includes the difference between each predicted price and the corresponding actual price. Please note that the dates are not sequential; this is intentional, as the model was trained on randomly ordered data to reduce bias and enhance the robustness of the predictions.

	Actual Price	Predicted Price	Difference
Date			
2015-07-27	47.43	47.83	0.40
2022-06-06	73.89	74.43	0.54
2017-03-07	50.74	50.66	-0.08
2023-05-12	71.83	72.12	0.29
2017-10-05	54.95	55.03	0.08
...	...	...	...
2023-11-20	73.92	73.62	-0.30
2024-01-18	77.03	77.24	0.21
2022-09-09	74.30	74.80	0.50
2021-12-30	89.52	89.35	-0.17
2019-12-13	61.21	61.68	0.47

Model Summary:

The learning curve illustrates the relationship between the training scores and the cross-validated scores. When these curves begin to converge, it indicates that adding more data will no longer enhance the performance of the model. This suggests that the algorithm is optimized in its current state. The merging of these 2 curves also indicates a good generalization of the model and low bias, meaning that the model can be used on different datasets and should still perform at the same level.

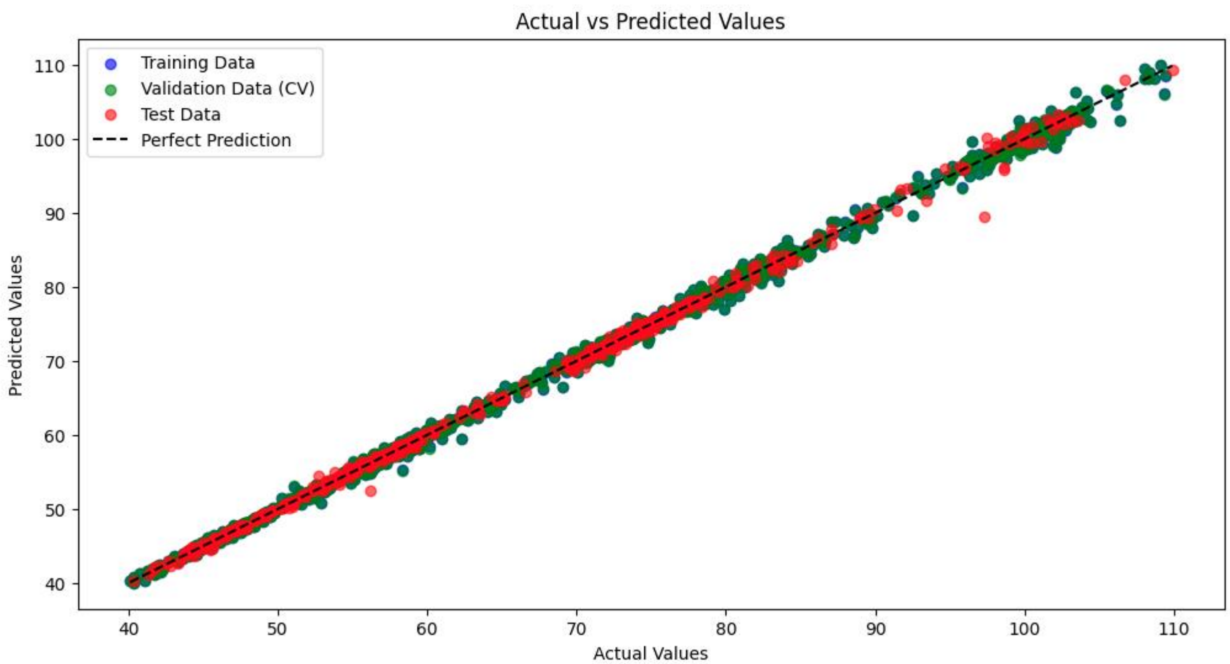


Statistics Interpretation:

```
Cross-Validation R^2 Scores: [0.99886018 0.9989765  0.99861235 0.99867265 0.99851139]
Mean Cross-Validation R^2 Score: 0.9987266156272241
Standard Deviation of Cross-Validation R^2 Scores: 0.00016886351845565409
Mean Squared Error: 0.4665100283343398
R^2 Score: 0.9981396319379283
```

1. **A cross-validation R^2 score, its average and the standard deviation of the same metric** indicate the performance of a model trained and tested on different data subsets. In this case, all scores were above 0.99, demonstrating high consistency and strong predictive accuracy across the entire dataset. This implies that the model should accurately predict the ETF's Net Asset Value (NAV) with minimal error across all samples in our dataset.

2. **Mean squared error (MSE)** measures average squared differences between predicted and actual values (NAV), reflecting regression model accuracy, a number closer to 0 is ideal.
3. R^2 is a statistical measure that indicates the variability in the dependent variable (NAV) that is explained by the independent variables, with values ranging from 0 (no explanatory power) to 1 (perfect explanatory power). After being optimized and validated, the model used to predict the NAV had a score of 0.998, which is reflected on the graph below:



Implications:

This analysis lays the foundation for employing convertible arbitrage with junk bonds, focusing on modeling the price movements of the Net Asset Value (NAV) in a distressed debt security. Convertible arbitrage is a sophisticated investment strategy utilized by investors to exploit perceived mispricing between a convertible bond and its underlying equity. This approach requires its investors to purchase the convertible bonds while short selling the underlying stock to mitigate directional market risk, aiming to capitalize on pricing disparities between the two securities.

Limitations:

1. **Data availability:** The dataset on ICVT consisted of only 2,258 data points, representing all available information presented by BlackRock. Generally, larger datasets enhance model robustness and applicability to similar datasets.
2. **Survivorship Bias:** The model may be affected by the exclusion of matured securities in convertible bond ETFs, as they require rebalancing.
3. **Macro-Economic Bias:** Recent interest rate hikes impose a substantial impact on debt securities, potentially distorting the perceived risk of junk bonds, a phenomenon known as Interest Rate Sensitivity. As rates rise, bond values typically decrease while conversion premiums tend to rise.

8. Conclusion:

While convertible junk bonds are associated with higher risk due to the lower credit quality of the issuers, they offer several benefits that make them attractive investments. These benefits include higher yields compared to traditional bonds, the potential for equity upside through conversion features, downside protection provided by the bond's fixed-income component, portfolio diversification benefits, the potential for credit improvement as issuers strengthen their financial position, and strategic flexibility in adapting to changing market conditions. These advantages make convertible junk bonds appealing to investors willing to accept higher risk in exchange for potentially higher returns and other strategic benefits.

When pooled into an ETF, convertible bonds may appear less risky than individual securities suggest. Despite price fluctuations over time, investing in such ETFs often demonstrates appreciation, thanks to a diversified portfolio of convertible bonds. This diversification helps mitigate the impact of individual bond defaults or price volatility, enhancing the overall risk-return profile of the investment.

Summary:

	Actual Price	Predicted Price	Difference
count	450.000000	450.000000	450.000000
mean	65.670756	65.678778	0.008022
std	15.853098	15.847645	0.683711
min	40.300000	40.330000	-7.830000
25%	53.687500	53.742500	-0.210000
50%	60.395000	60.490000	0.050000
75%	75.307500	75.457500	0.300000
max	109.890000	109.390000	2.650000

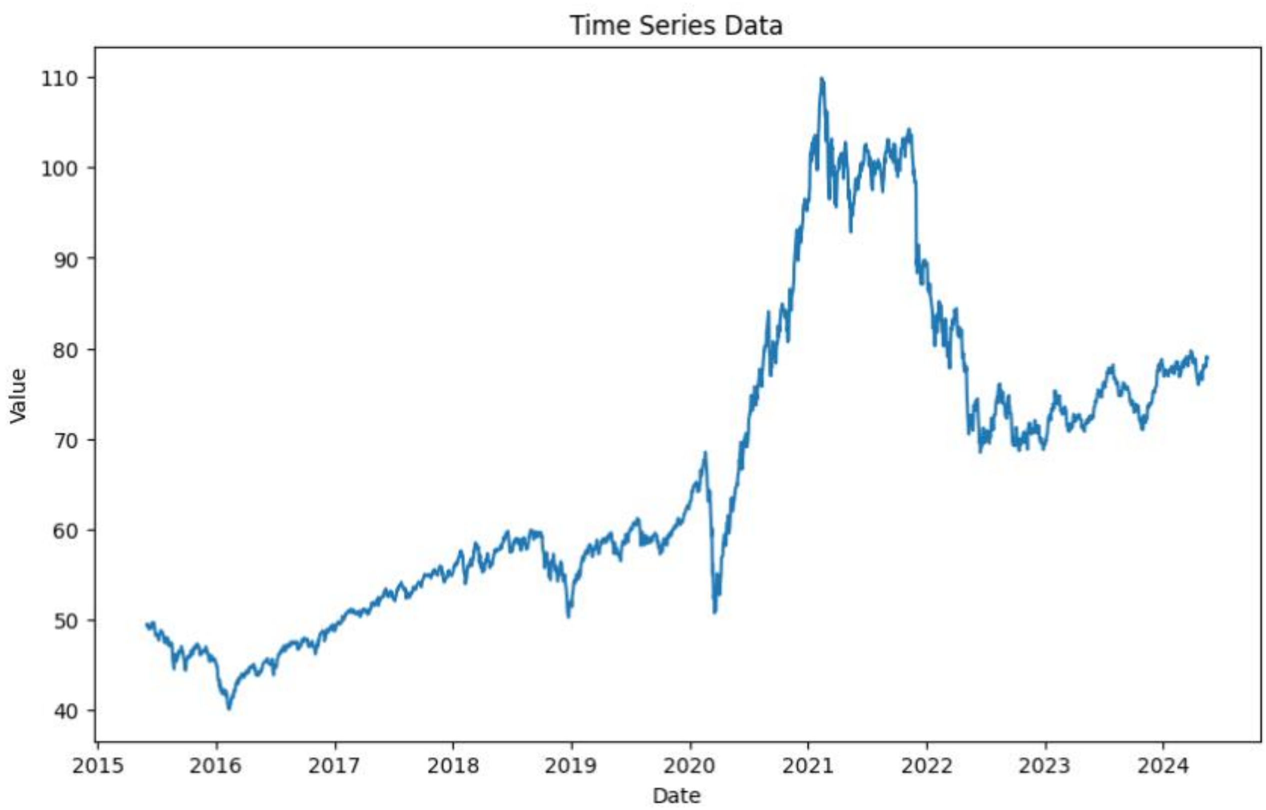
```
# Calculate the difference between the min and max values for the actual prices
min_value = results["Actual Price"].min()
max_value = results["Actual Price"].max()
difference = max_value - min_value

print(f"The difference between the max and min values in for the actual price is: ${difference}")
```

```
The difference between the max and min values in for the actual price is: $69.59
```

The chart above illustrates the NAV price predictions compared to the actual data points in the dataset. Utilizing 450 of these prices, I developed a machine learning model capable of forecasting across 9 years of data with only minimal discrepancies. An investor who purchased the ICVT “shares” in 2015 and held them until May 17, 2024, could have seen a NAV appreciation of over $30.00, validating the premise of my study. Investing in distressed debt securities offers a high reward, even when pooled into investment vehicles like ETFs.

Recent interest rate hikes have facilitated higher returns for investors in the portfolio. While the stock price significantly influences overall price movements due to the convertible bond nature, portfolio management rebalancing strategies ensure that this pooled investment vehicle mirrors the price changes of a standard ETF.



Recommendations:

Assessing the credit risk of each holding is a crucial step in an institutional investors' due diligence process for creating ETF securities. Continuous analysis of issuer creditworthiness and default likelihood is vital, particularly for ETFs with numerous holdings like the one in this research. Leveraging data provided by BlackRock directly proves more efficient than analyzing individual debt securities. Ideally, this analysis should include a summary of each holding in the investment pool. Remarkably, utilizing available information proved sufficient for the machine learning model to predict NAV price movements with only a few cents of discrepancy.

9. References:

For space saving purposes, the following references will include the links to their source only.

https://investor.vcm.com/insights/investor-learning/the-difference-between-etf-price-and-value

https://www.BlackRock.com/us/individual/products/272819/ishares-convertible-bond-etf

https://www.BlackRock.com/us/individual/literature/fact-sheet/icvt-ishares-convertible-bond-etf-fund-fact-sheet-en-us.pdf

https://www.moodys.com/sites/products/productattachments/ap075378_1_1408_ki.pdf

https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=dimensionally-rolling-moving-averages

https://www.investopedia.com/ask/answers/071414/whats-difference-between-moving-average-and-weighted-moving-average.asp

https://www.investopedia.com/ask/answers/071414/whats-difference-between-moving-average-and-weighted-moving-average.asp

https://www.investopedia.com/terms/l/learning-curve.asp#:~:text=A%20learning%20curve%20is%20measured,more%20proficient%20at%20the%20task.

https://www.investopedia.com/terms/j/junkbond.asp

https://www.morganstanley.com/im/publication/insights/investment-insights/ii_highyieldbondsinrisingrates_en.pdf

https://www.scikit-yb.org/en/latest/api/model_selection/learning_curve.html#:~:text=A%20learning%20curve%20shows%20the,varying%20number%20of%20training%20samples.

https://scikit-learn.org/stable/modules/cross_validation.html

https://www.sciencedirect.com/science/article/abs/pii/S0169207007000052

10. Appendices:

You can find the code used for this writing sample in the following links:

- Python:https://colab.research.google.com/drive/1HR7yxYmHUI1LxxJfs_oQtRZ-z51ZniVQ?usp=sharing
- R: https://posit.cloud/content/8237838

If you need access to the data sets, please contact me via e-mail at: mariiacamila31@gmail.com