Fixed-Income Relative Value from a Quantitative Perspective

Silverio J. Vasquez

WorldQuant University

Abstract

Fixed-income relative value strategies are used by hedge funds to manage a significant amount of wealth. There are various implementation strategies, one of which led to the infamous downfall of Long-Term Capital Management. This study aims to identify alpha for yield spreads between two fixed income securities by leveraging various supervised learning quantitative models. Each strategy is judged on its own merit before the top performers are incorporated into two different portfolio construction techniques: the naive equal-weighted portfolio and another employing hierarchical risk parity. The naive portfolio slightly outperforms, contrasting findings in other research studies.

Keywords: *fixed income, relative value, macro, quantitative analysis, random forest, logit, support vector machine, principal component analysis, hierarchical risk parity, dendrogram*

1. Introduction

Assets under management in the hedge fund industry increased to over $5 trillion in 2023[1], according to BarclayHedge. Hedge funds employ numerous strategies to generate competitive risk-adjusted investment returns. Fixed-income relative value is among these strategies with about $315 billion in assets under management as of 2020Q1[2] and is the strategy explored in this project. The goal of this project is to combine a fixed-income relative value framework with quantitative analysis to maximize risk-adjusted returns.

Some hedge fund strategies, such as global macro and managed futures, have experienced a loss in alpha since the Global Financial Crisis due to a larger portion of returns getting explained by traditional market factors such as momentum (Sullivan, 2023). Hedge funds aim to provide diversified returns, but if traditional market factors are prevalent in their returns, then this reduces the diversification benefit for a large investor invested in multiple hedge funds. Fixed-income relative value, by design, pairs one security versus another. This helps increase its diversification benefit by reducing correlations to traditional market factors as well as macroeconomic factors (Andrea, 2021).

In pairs trading, the investor is generally long (the investment increases as the value of the security increases) one security and short (the investment increases as the value of the security declines) another. The strategy was created in the mid-1980s at Morgan Stanley where researchers identified pairs of securities via statistical models (Gatev, 2006). The investment return is the net gain or loss between the two securities and is what helps minimize correlations to traditional market factors.

Part of this study will focus on how to construct a portfolio of different relative value trades by comparing the naive portfolio (equal weights of all trades) to hierarchical risk parity (López de Prado, 2016).

[1] https://www.barclayhedge.com/solutions/assets-under-management/hedge-fund-assets-under-management/hedge-fund-industry
[2] https://www.preqin.com/insights/research/factsheets/q1-2020-hedge-fund-asset-flows

2. Literature Review

2.1 Fixed-Income Relative Value

There are numerous strategies that are categorized as fixed-income relative value. The infamous Long Term Capital Management hedge fund employed a fixed-income relative value strategy with long positions in undervalued bonds and short positions in overvalued bonds (Edwards, 1999). Their downfall was the amount of leverage and, surprisingly, poor risk management given the number of renowned academics working at the hedge fund. Another form of fixed-income relative value involves basis risk, which is the difference between the price of a Treasury futures contract and the price of the cheapest-to-deliver underlying bond (Di Maggio, 2020). These trades, however, require the use of cash and borrowing via repurchase agreements (repo)[3]. This introduces complexity, however, in calculating the actual return, which is why this project will keep it simple by focusing purely on the difference in yields across fixed income securities.

Other attempts at fixed-income relative value include butterfly trades, which is when one goes long a bond of a specific maturity and shorts two other bonds of different maturities (all within the same yield curve, e.g., US Treasuries). The duration of the long position is matched to the net duration of the short positions to avoid unnecessary interest rate risk. Fontaine and Nolin (2017) explored this via a model-free approach to avoid sampling estimation errors and model misspecification. This study avoids butterfly trades to avoid the complexity of more than two positions per trade. Fontaine and Nolin (2017) focused on relative value along the yield curves of various countries, but the authors did not explore how these different trades could have been combined in a way to produce the best risk-adjusted return possible.

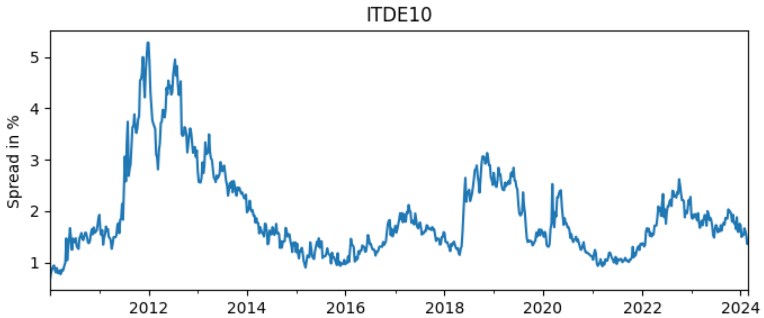
The relative value strategy employed in this study seeks to capitalize on the difference in yields between two securities. If the securities belong to the same yield curve, then the difference in yield between a longer maturity security and a shorter maturity security, e.g., the yield difference between the 10-year and 2-year US Treasury securities, is called a yield curve. The changes in a yield curve can be driven by growth and/or inflation concerns. The classic example is an inverted yield curve, where the difference (spread) is negative, which has been a good predictor of recessions in the US (Estrella, 1996). Constant maturity yields are used in this study, which are the same as Treasury yields but created in a way to avoid the change in yield as a bond approaches its maturity date. Diagram 1 illustrates the US 30Yr-2Yr curve.

[3] Investors can use the repo market for short-term collateralized loans allowing them to increase their leverage. LTCM utilized the repo market to increase their leverage ratio to very high levels surpassing 20-to-1 (Edwards, 1999).

Diagram 1: US 30-Year Constant Maturity Yield less the US 2-year Constant Maturity Yield

Another example of fixed-income relative value is the spread between a security with minimal credit risk and another with more credit risk. Since these securities are of different credit risk, they do not belong to the same yield curve. The spread between the Italian 10-year sovereign yield and the German 10-year sovereign yield is a popular example where a widening of this spread indicates financial concerns by investors (ECB, 2008). Diagram 2 depicts how this spread increases during periods of high macro risk like the Eurozone debt crisis in 2011.

Diagram 2: Italian 10-year yield less German 10-year yield



This study will leverage three different supervised learning statistical models and one unsupervised learning statistical model to create trading strategies. Further details will be covered in the methodology section, but the statistical models are introduced below.

2.2 Principal Component Analysis (unsupervised learning)

Principal component analysis (PCA) is an unsupervised learning technique, i.e., all features and no target variable, used to reduce a high dimensional dataset to a lower dimension. The feature set is standardized (removing the mean and scaling by standard deviation) via the StandardScaler Python package from the sklearn library[4] before applying the Python package PCA from the sklearn library[5]. Eigenvectors, which are sensitive to scaling, are calculated and used to calculate principal components, which are orthogonal to each other, i.e., principal component 1 is orthogonal to principal component 2. The actual computation is done via Singular Value Decomposition (SVD), which is a more robust algorithm that yields the same result as when PCA is applied using a covariance matrix after centering along the columns (Wall, 2003). The principal components are in descending order of importance in capturing the variation across all

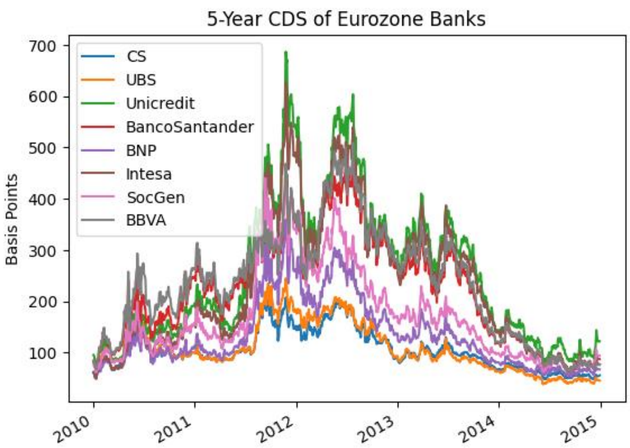
[4] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
[5] https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

features, i.e., principal component 1 captures more of the variation across all features than principal component 2 (Gareth, 2023).

PCA has been used to decompose the main components of a yield curve (Litterman, 1991), but it can also help identify if a specific bond or swap tenor is either trading rich or cheap relative to the same securities across maturities. If a PCA analysis identifies that the yield of a 5-year US Treasury note is trading rich (low yield, high price) compared to the US Treasury yield curve across maturities, then an investor can short the 5-year US Treasury note. This approach, however, only has one leg and exposes the investor to interest rate risk. The investor can go long the 2-year and 10-year US Treasuries to neutralize the interest rate risk and ensure that return and risk only come from the relative value identified. This is similar to the approach used by Fontaine and Nolin (2017), but it applies a purely statistical approach and only uses two positions (not three which are required in butterfly trades).

This study leverages PCA to identify the main components explaining the variation across the 5-year Credit Default Swaps of eight European banks: Credit Suisse, UBS, Unicredit, Banco Santander, BNP, Intesa, Société Générale, and BBVA. A Credit Default Swap (CDS) represents an insurance contract against the event of a default of an underlying asset. They are traded over the counter across various maturities, but the 5-year maturity is one of the most liquid (Giglio, 2014). The CDS spread, the price of the insurance protection, is used as a risk gauge with a higher spread suggesting higher default probability. (Giglio, 2014). During the Eurozone debt crisis of 2011, these CDS spreads increased markedly indicating higher risk of default (Diagram 3).

Diagram 3: 5-year CDS Spreads



The top three principal components (PC) identified explain 91% of the variation for the eight CDS during the initial training period (Diagram 4). One can use the output of PCA to create linear combinations. The steps below are done for each row (time observation) in the dataset to calculate the linear combination of all the principal components and the scaled feature set.

Step 1: define a principal component as a linear combination (dot product) of features {x1, …., xp} from the scaled feature set, **X**, and the loadings (coefficients), $\boldsymbol{\phi}$, from the PCA decomposition (Gareth, 2023).

$$PC_1 = \phi_1 x_1 + \phi_2 x_2 + \ldots + \phi_p x_p \qquad \textit{Equation 1}$$

where p = number of columns in **X.**

Step 2: calculate the dot product of the PCs and the proportion of the variance explained by each PC (W_i), where i = number of principal components

$$Index\_d = PC_1W_1 + PC_2W_2 + PC_3W_3 \qquad \textit{Equation 2}$$

Index_d is a time series of deltas, because **X** is a series of differences in basis point (1/100 of 1%) from one observation to another. The final index, the Eurozone Risk Index, is the cumulative sum of Index_d. Diagram 5 visualizes how well the Eurozone Risk Index (black line) captures the aggregate variation across all the CDS over time.

Diagram 4: Principal Components

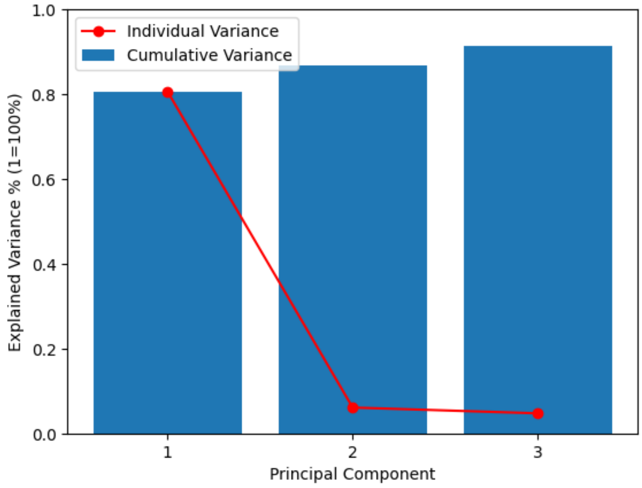
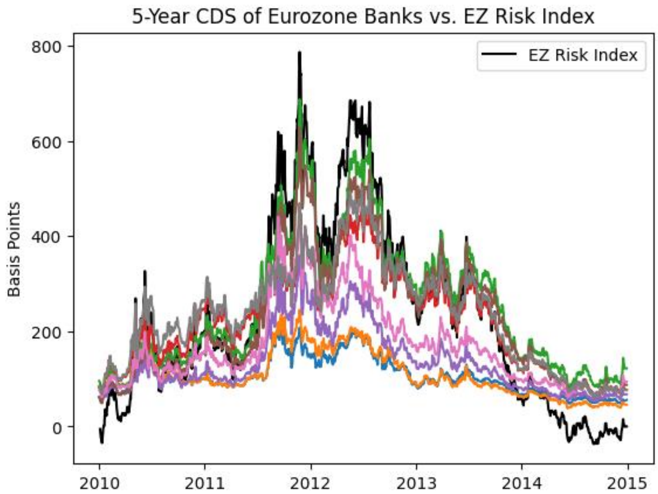


Diagram 5: Eurozone Risk Index



2.3 Statistical Models

Three supervised learning statistical models will be utilized – Logit Model, Random Forest Classifier (RFC), and Support Vector Classifier (SVC) for classification. The target variable is categorical and explained further in the methodology section. The Logit model was chosen to capture the linear relationship between the feature set and the target variable. Sometimes nonlinear models introduce complexity without added benefit. Financial data, however, do not always follow linear relationships, which is why two nonlinear models will also be used. Logit and RFC also have the benefit of identifying variable importance, which is helpful when researching the historical behavior of a particular model. SVC can also provide variable importance when using a linear kernel, but the radial ("RBF") kernel is employed in this study. At least two out of the three models can be audited for variable importance. Each model will provide a long, short, hold signal which will be aggregated via majority vote outlined in the methodology section.

2.3.1 Logistic Regression

The Logistic (Logit) model identifies a linear relationship between the target variable and the features dataset similar to Ordinary Least Squares (OLS). The nature of the target variable is the main difference between Logit and OLS models. The target variable is categorical for the former and continuously distributed for the latter.

The Logistic model applies the logarithm function to both sides of the linear equation, as defined below, limiting the target into either a 0 or 1 (Gareth, 2023). Therefore, a unit rise in X would result in a β_1 change to the log odds of the target variable, ceteris paribus.

$$log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X \quad \textit{Equation 3}$$

There are other supervised learning models that handle a categorical target variable, such as Naive Bayes, but the Logit model was chosen due to ease of use and familiarity.

2.3.2 Support Vector Classifier

The Support Vector Classifier (SVC) model is a nonlinear supervised model that attempts to classify the target variable into categories by splitting the data in various hyperplanes. This entails maximizing the margin surrounding the hyperplane that correctly classifies the highest number of observations (Gareth, 2023). This approach is imperfect and not all observations will be classified correctly. One can determine how much slack is allowed via the "C" non-negative hyperparameter, i.e., the L2 (squared) tuning parameter. This parameter helps control the bias-variance tradeoff, whereby a high value for C means more errors (violations) are allowed and decreases the variance at the expense of a higher bias, or underfitting (Gareth, 2023). The converse is true for lower values of C, i.e., higher variance and lower bias (overfitting). The mathematical intricacies underlying SVC are out of scope for this project.

2.3.3 Random Forest Classifier

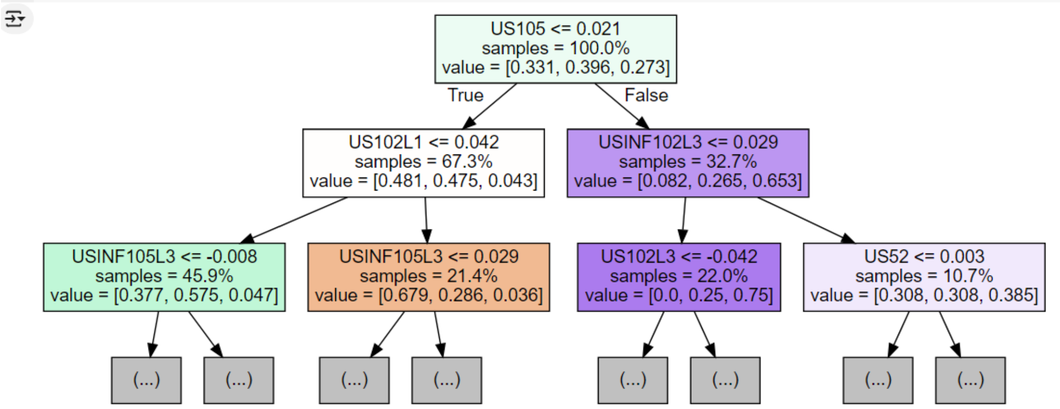
The Random Forest Classifier (RFC) model is a nonlinear supervised model that leverages the bagging of random trees to identify the target variable.

Diagram 6, below, visualizes how the RFC model split the trees in trying to fit a predictive model for the target variable (the US 10yr - 2yr yield curve). The first row is the root node which uses 100% of the samples. The first split is based on when the first predictor (the US 10yr - 5yr yield curve) is less than or equal to 0.021% (or 2.1 basis points). The value list comprises percentages for the three categories the model is aiming to predict: long, hold, and short. This first split states that there is a 39.6% probability of the target variable being a hold from simply using this one feature.

The second row has two branches based on where the target variable yield is after splitting the data with respect to the value of the first feature. The left-hand branch states that 67.3% of the samples fall into this branch based on the US 10yr - 5yr yield curve being less than or equal to 0.021%. The rest goes to the right-hand branch. One can confirm all the samples are used across the two branches by summing the respective samples (67.3% and 32.7%, respectively) which total to 100%.

The algorithm continues to generate splits based on a stopping criterion (Gini impurity is the default and what was used). Another stopping criterion is the number of trees, estimators, in a model. This threshold is a hyperparameter that one can optimize (further detailed in the methodology section).

Diagram 6: Decision Tree



2.4 Portfolio construction

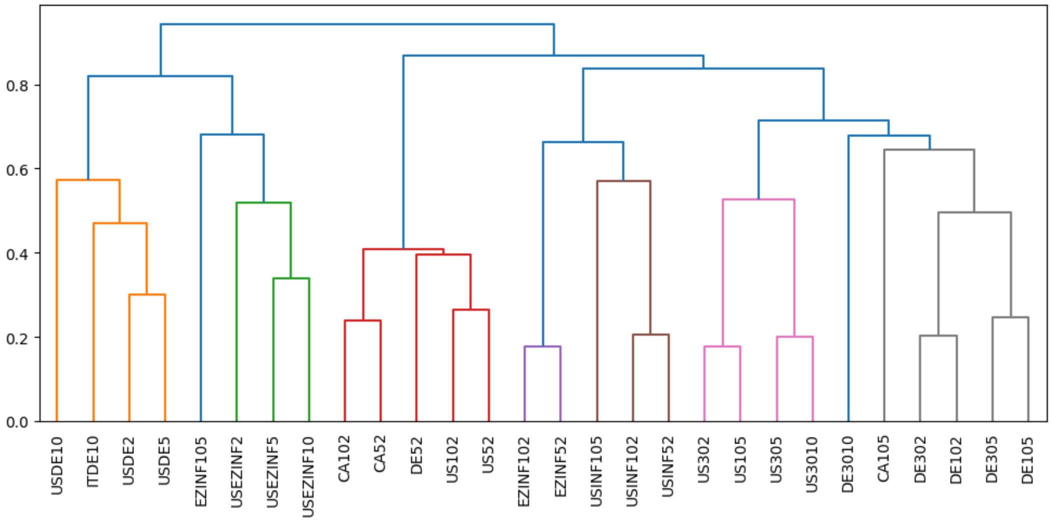
Markowitz's mean-variance optimization remains a popular approach to construct portfolios. This approach, however, suffers from sensitivity to the input data (López de Prado, 2016). This study seeks to further the practical application of fixed-income relative value by combining it with hierarchical risk parity (HRP), which applies risk parity at a cluster level to incorporate correlations between positions (López de Prado, 2016). The correlation matrix (Figure 1 in Appendix) shows there are meaningful relationships between a myriad of spreads, which introduces the possibility of optimizing the risk-adjusted return of a portfolio of spreads.

Hierarchical clustering is used to group stocks that have higher correlations. This approach measures the distance between values, which is defined as the distance between correlation coefficients. A large (farther) distance implies more differences between two securities. Given a matrix of correlation coefficients, the distance will be calculated as follows (López de Prado, 2016):

$$distance = \sqrt{\frac{1}{2} * (1 - \rho_{i,j})} \quad \textit{Equation 4}$$

where $\rho_{i,j}$ is the correlation coefficient between security i and security j. The dendrogram in diagram 7 below identifies various clusters (color coded) based on the distance measure defined above. These clusters help inform the "HRPOpt" algorithm in the PyPortfolioOpt Python package (Martin, 2021).

Diagram 7: Dendrogram for a selected number of fixed-income spreads



3. Methodology

3.1 Data

This study will leverage financial market data such as short-term interest rates (STIR) futures for the US and Eurozone, inflation swaps across various tenors for the US and Eurozone, credit default swaps (CDS) for European banks, and constant maturity yields across various tenors for the US, Canada, UK, Germany, and Italy. The dataset is weekly, follows the US holiday calendar, and covers the period between January 2010 and December 2022. All data are sourced from Bloomberg. Figure 1 in the appendix illustrates the high correlations among numerous variables.

3.2 Target and Feature Sets

The target variables will be the weekly change in percentage point of numerous fixed-income spreads one period ahead. Table 1 in the appendix lists all the top eight target variables after backtesting each individual strategy.

$$\gamma_{t+1} = F(\chi_t \beta_t) \quad \textit{Equation 5}$$

where γ_{t+1} is the weekly change one period ahead and $F(\chi_t \beta_t)$ represents the model with the inputs as of the current period.

The feature set will consist of the top securities whose absolute Pearson correlation coefficients are above 0.7 over the training period. A maximum of 10 features per target variable is implemented for parsimony. Each model for each target variable will be trained numerous times during a walk-forward backtest. This allows for dynamic updates to the feature set helping avoid stale parameter selection during different market regimes, e.g., a parameter initially trained between January 2010 and December 2012 may fall out of favor during 2020-2022.

3.3 Models

Three statistical models will be utilized – Logit Model, Random Forest Classifier (RF), and Support Vector Classifier (SVC) for classification. The feature dataset will hold continuous variables, but the target variable will be categorical:

- +1: if the weekly increase is greater than 3 basis points (1/100 of 1%)
- -1: if the weekly decline more than negative 3 basis points
- 0: otherwise

Each model will predict the position to implement to capture the change in spread over the subsequent week. The ultimate position will be decided by an equal vote across all three models, i.e., the position will be the mode of all three predictions. In the case that each model produces a different position, the default will be a 0.

3.3.1 Workflow

Each model will be trained initially using weekly data between January 2010 and December 2014. The subsequent 52 weeks of data points will be used to validate the models. The RFC and SVC models will undergo some hyperparameter tuning during the validation phase. The penalty term “C” will be tuned for the SVC model, and the number of trees will be tuned for the RFC model.

Each target variable will have its respective feature set. This feature set will be lagged up to four weeks to capture any lagged effects in the data. Lags of the dependent (target) variable are also included to capture any serial correlation, i.e., momentum, given some of the first differenced data showed some questionably significant lags in the autocorrelation plots (see Figure 2 in the appendix for some examples). Albeit the Augmented Dickey Fuller test (Dickey, 1979) confirmed all target variables are stationary because the null hypothesis of a unit root was rejected at the 5% significance level.

Every trade seeks to make a profit from a relative value perspective. Each trade is structured as two legs – a leg is one side of a trade. One leg represents a long position, while the other leg represents a short position. Since this analysis uses yields and not prices, it assumes that both legs have the same DV01(dollar value of 1 basis point move). If the DV01 of each leg is not the same, then one of the legs will be more sensitive to the move in interest rates.

Assume a trade is to capture a widening of the US 2s10s swap curve (the term spread – difference in yields -- between US 10-year swap and US 2-year swap) from the 2-year yield declining by 5 basis points (bps), say, from 3.95% to 3.90%, and the 10-year yield rising by 5 bps, say, from 4.20% to 4.25%. Note that in this simple example, both legs are moving by the same number of bps. In this trade, one makes money when the 2-year swap rate declines and when the 10-year swap rate rises.

There is a risk that the whole curve has a parallel move either up or down, while the trade is betting on a widening of the curve (an increase in term spread). An investor is exposed to this risk if the DV01s (duration) are not equal, i.e., assume the 2-year swap leg has a $5K DV01, the 10-year swap leg has a $10K DV01, and the whole curve moves down by 5 bps. This means that then the position will generate a $25K gain on the 2-year leg, but a $50K loss on the 10-year leg. If both legs had the same DV01, this parallel move (an unwanted risk) would be neutralized with no impact on returns.

3.4 Backtest

A walk-forward analysis is used to backtest each trade. Diagram 8 below illustrates a sample walk-forward analysis. The analysis displayed below uses a rolling 3-year window of observations to train a model. It then uses the subsequent year’s worth of data to validate the model, and then the year after that to test it. This process is similar to the one employed in this study except for one exception: the test period is only one time period ahead instead of a whole year. The models in the study will predict a position of either long/hold/sell {1, 0, -1} for the duration of one week (the frequency of the data). The string of predicted positions will be used to calculate individual trading returns per strategy.

Diagram 8: A walk-forward analysis

Train			Validate	Test		
Year 1 = 2010	Year 2 = 2011	Year 3 = 2012	Year 4 = 2013	Year 5 = 2014		
Next iteration	2011	2012	2013	2014	2015	
	Next iteration	Train			Validate	Test

A position established in the prior week will persist if the forecasted next position is 0 (hold). That is, if the previous week's position was a +1 and the signal is a 0, then the position will remain +1.

Yields are more important in fixed income, unlike with equities where returns are more important. All the signals are based on yield spreads in percentage point terms. This prevents the use of percent change as a form of calculating returns on a position. The following assumptions are made for each position for the walk-forward analysis:

- There is a portfolio with assets under management (AUM) of US$ 1 million in cash
- Each trade is made with a size of $10,000 DV01
    - Note: a losing position that moves 100 bps will wipe out the portfolio.
- The weekly $ profit/loss is calculated as the $10k DV01 multiplied by the number of bps the position moves over the week.
- The weekly profit/loss is added to the cash, and the weekly return is the percent change in the cumulative AUM balance.

4. Results

4.1 Individual Strategy Performance

The top eight strategies are presented in table 2 below. These strategies achieved the highest Sharpe Ratio (assuming a zero risk-free rate) defined as annualized mean of return per unit of annualized standard deviation. A ratio of 1.0x is deemed acceptable with anything greater deemed very good. While other performance statistics are presented in the table, the selection of the top eight strategies were solely chosen based on their respective Sharpe Ratio. Plots of cumulative returns are presented in Figure 3 in the Appendix. Table 2 includes the following performance statistics:

- **Mean** is the annualized average rate of return

$$\frac{\sum_{i=1}^{n} x_i}{n} * 52 \qquad \textit{Equation 6}$$

    - where x_i is the weekly return

- **Vol** is the annualized standard deviation

$$\sqrt{\frac{\sum (x_i - \mu)^2}{n}} * \sqrt{52} \qquad \textit{Equation 7}$$

    - where x_i is the weekly return, μ is the mean return over the entire rolling 1-year period, and *n* is the total number of returns per period, i.e., 52

- **MDD %** is the maximum drawdown as a percent of peak holding (investment) value.

$$\frac{Trough\ Equity\ Value}{Peak\ Equity\ Value} - 1 \qquad \textit{Equation 8}$$

- **Sortino Ratio** is the annualized mean return per unit of downside annualized standard deviation, i.e., the standard deviation calculation performed on only negative returns (Rollinger, 2013). Like the Sharpe, a value of 1.0 or higher is preferred.

$$\frac{r_p - r_f}{\sigma_d} \qquad \textit{Equation 9}$$

    - where r_p is the return of the strategy, r_f is the risk-free assumed to be zero, and σ_d is the standard deviation of negative returns

- **Calmar Ratio** is the annualized mean return divided by the MDD % (Young, 1991). Like the others, a value of 1.0 or higher is preferred.

$$\frac{r_p - r_f}{MDD\%} \qquad \textit{Equation 10}$$

Table 2: Performance statistics of individual strategies

Targets	Spreads	Mean %	Vol %	Sharpe	MDD %	Sortino	Calmar
US102	US 10Y-2Y	18.83	16.33	1.15	-24.88	2.09	0.76
US52	US 5Y-2Y	11.97	20.92	0.57	-37.48	0.86	0.32
US302	US 30Y-2Y	20.07	24.2	0.83	-38.5	1.43	0.52
DE102	German 10Y-2Y	19.16	16.98	1.13	-24.96	1.76	0.77
DE52	German 5Y-2Y	15.16	30.91	0.49	-45.62	0.72	0.33
CA102	Canada 10Y-2Y	16.37	35.79	0.46	-72.37	0.61	0.23
ER6-12	Euribor 6th contract less 8th contract (effectively a 3Y - 1.5Y yield curve)	15.85	20.64	0.77	-44.59	1.09	0.36
SFR4-8	US SOFR 4th contract less 8th contract (effectively a 2Y - 1Y yield curve)	19.98	38.94	0.51	-57.86	0.67	0.35
ITDE10	Italian 10Y - German 10Y	0	100	nm	-100	nm	nm

nm = not meaningful

The Italian 10Y spread over the German 10Y produced poor results in the backtest. This strategy was dropped from the portfolio construction section.

4.2 Portfolio Construction Results

Combining strategies into a portfolio can outperform an individual strategy due to diversification benefits. All the individual trading signals are aggregated into two portfolios to try to enhance overall performance. One portfolio construction approach is equal weighted, and the other is constructed using hierarchical risk parity.

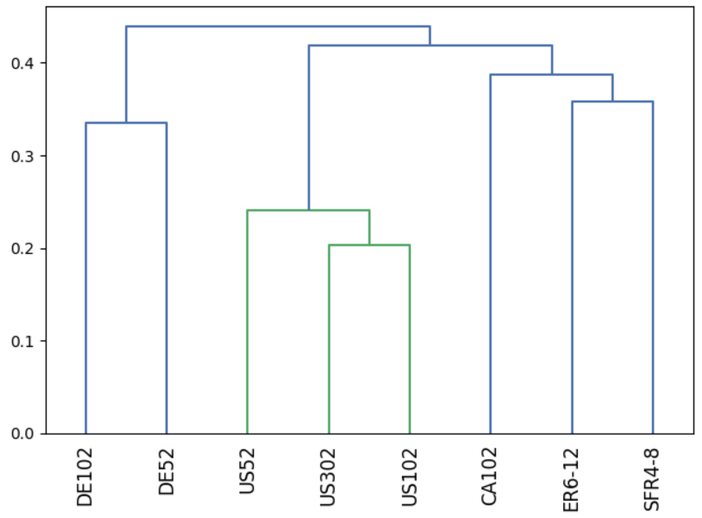
The HRP portfolio uses weights that were calculated annually and held constant for the duration of that specific year. The prior year's 52 weeks of returns were used in the HRP algorithm to calculate the weights. This caused the weights to shift throughout time. Table 3 below outlines how the weights shifted between 2014 and 2022, and diagram 9 displays their corresponding dendrograms. Plots of cumulative returns per strategy are in figure 4 in the Appendix.

Table 3: Weights Changed Over Time

Strategy	Weights - 2014	Weights - 2022
US102	8.43%	2.89%
US52	14.81%	7.67%
US302	8.61%	1.46%
DE102	12.65%	11.27%
DE52	16.11%	31.96%
CA102	19.59%	3.59%
ER6-12	10.82%	33.10%
SFR4-8	8.97%	8.06%

Diagram 9: Dendrograms

2014



2022

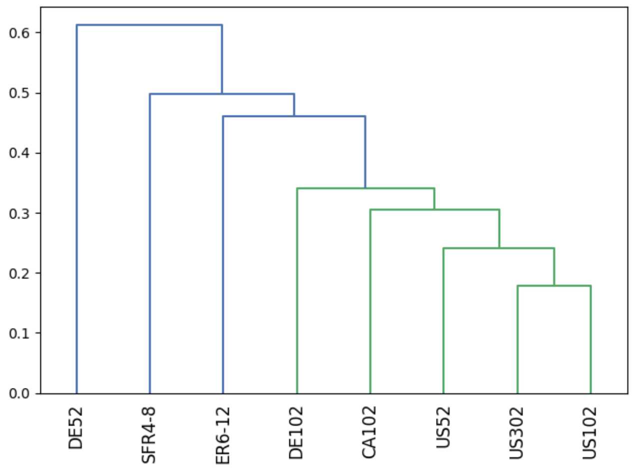


Table 4 below outlines the results. The equal-weighted portfolio outperforms the HRP portfolio across all performance statistics. This can be partly explained by the large weights the HRP portfolio places on a handful of securities in the later years.

Table 4: Equal-Weighted Portfolio Outperforms on a Risk-Adjusted Basis

Portfolio	Mean %	Vol %	Sharpe	MDD %	Sortino	Calmar
Equal Weighted	17.14	14.59	1.17	-23.6	1.72	0.73
HRP	18.67	17.24	1.08	-33.9	1.58	0.55

The better distributed equal-weighted portfolio likely benefitted from diversification resulting in lower volatility and drawdown. The HRP portfolio assigned a higher weight to some holdings at the expense of others. This helped the HRP portfolio achieve a slightly higher mean return but also led to higher volatility.

5. Conclusion

This study found that there is merit to combining fixed-income relative strategies with quantitative models for both alpha research and portfolio construction. While only two individual strategies produced Sharpe Ratios of over 1.0 in the backtest, both portfolio construction approaches produced performance statistics that were largely better than the individual strategies.

The backtest results of the individual strategies would benefit from further research due to a myriad of factors that are worth revisiting. The model selection, hyperparameter optimization, and feature selection all incorporate assumptions and perhaps some bias. The results of the portfolio construction analysis are interesting in that it counters other studies where HRP portfolios tend to outperform (López de Prado, 2016). This section would also benefit from further investigation. All three supervised learning models used in this project require parameter estimation, which introduces complexity, sampling uncertainty, and possible model misspecification. This study could be enhanced by exploring other models (both parametric and nonparametric) to predict long/short positions, as well as other approaches to portfolio construction not considered in this project.

Appendix

Figure 1: Correlation heatmap across various term spreads

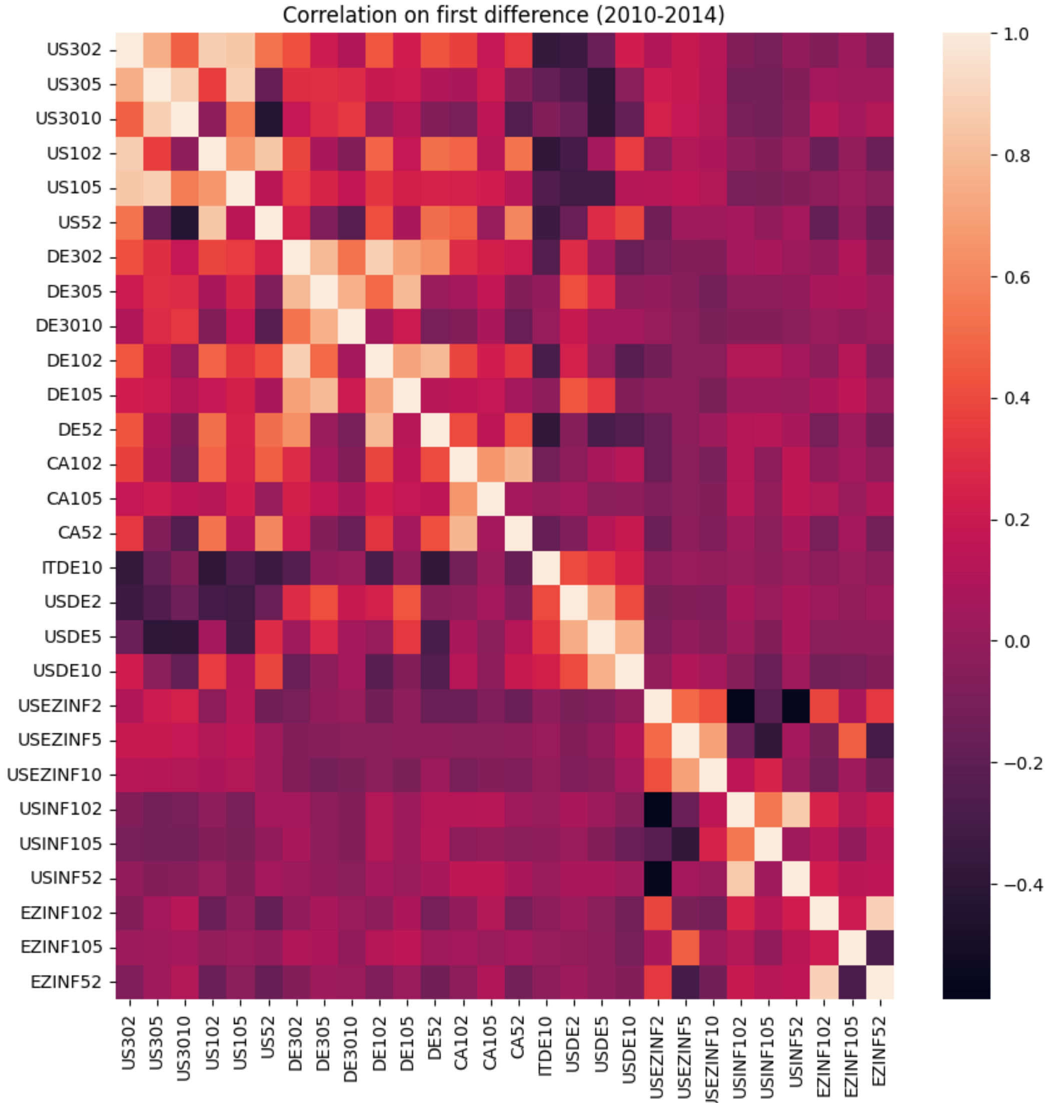
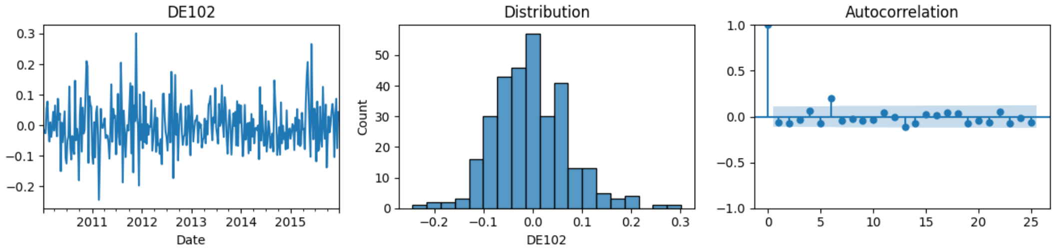


Table 1: Strategy variables (fixed-income spreads)

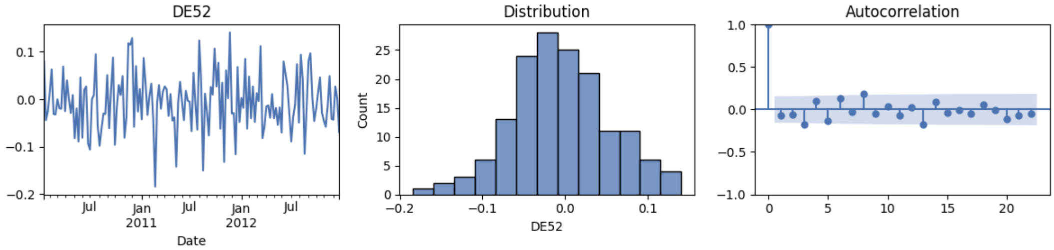
Strategy	Positions
US102	US 10yr less 2yr spread
US52	US 5yr less 2yr spread
US302	US 30yr less 2yr spread
CA102	Canada 10yr less 2yr spread
DE102	German 10yr less 2yr spread
DE52	German 5yr less 2yr spread
ITDE10	Italy 10yr less Germany’s 10yr
ER6-12	Euribor 6th contract less 8th contract (effectively a 3Y - 1.5Y yield curve)
SFR4-8	US SOFR 4th contract less 8th contract (effectively a 2Y - 1Y yield curve)

Figure 2: Distribution and autocorrelation plots for selected spreads

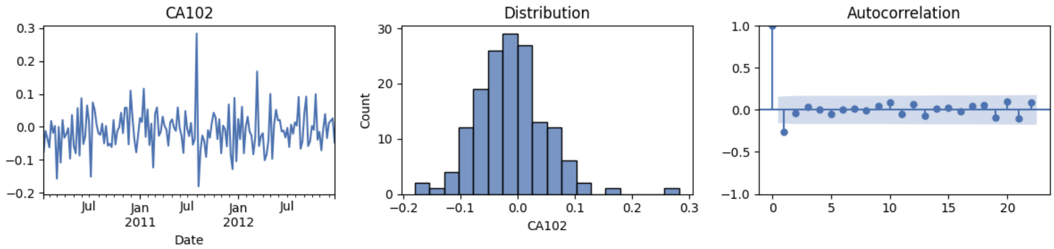
Spread between German 10-Yr and 2-Yr yields



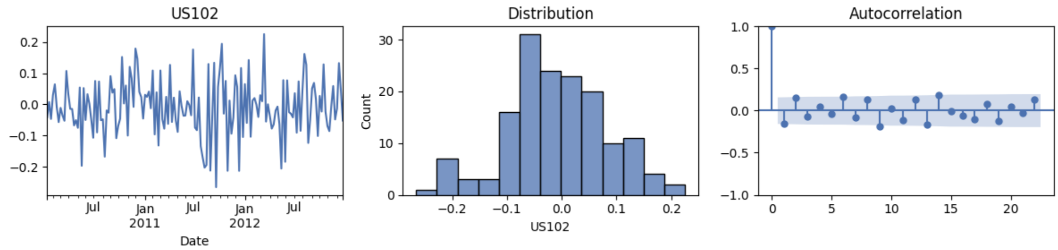
Spread between German 5-Yr and 2-Yr yields



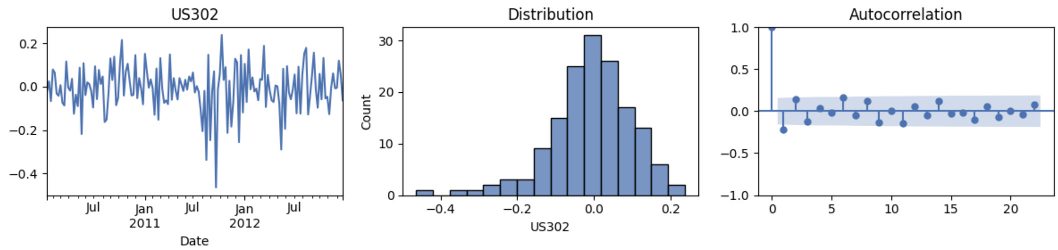
Spread between Canada 10-Yr and 2-Yr yields



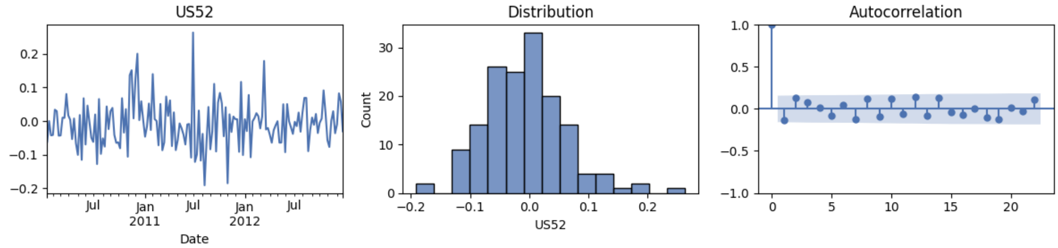
Spread between US 10-Yr and 2-Yr yields



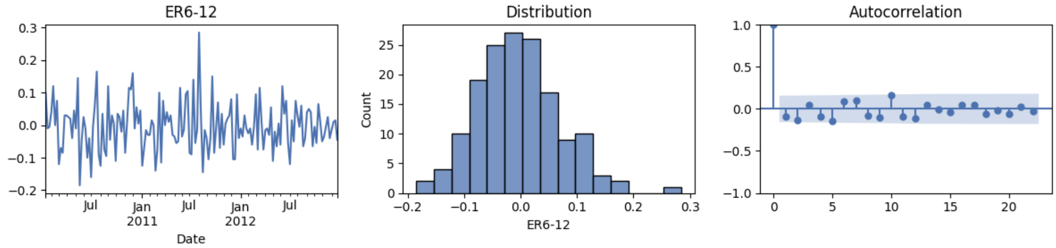
Spread between US 30-Yr and 2-Yr yields



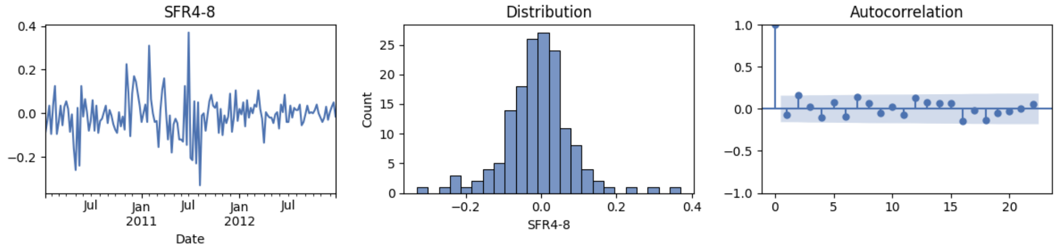
Spread between US 30-Yr and 2-Yr yields



Spread between Euribor 6th contract and 12th contract (3Yr - 1.5Yr spread)



Spread between SOFR 4th contract and 8th contract (2Yr - 1Yr spread)



Spread between Italian 10-Yr and German 10-Yr

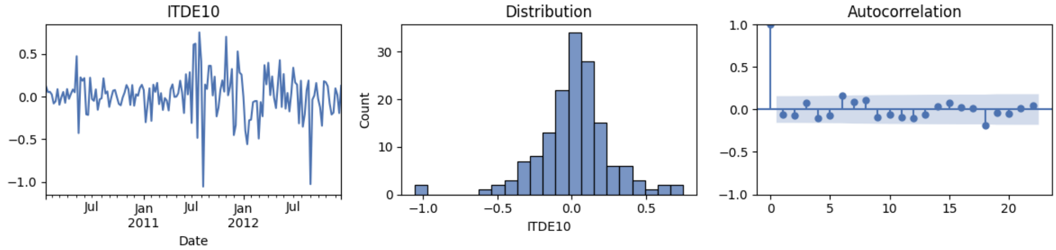
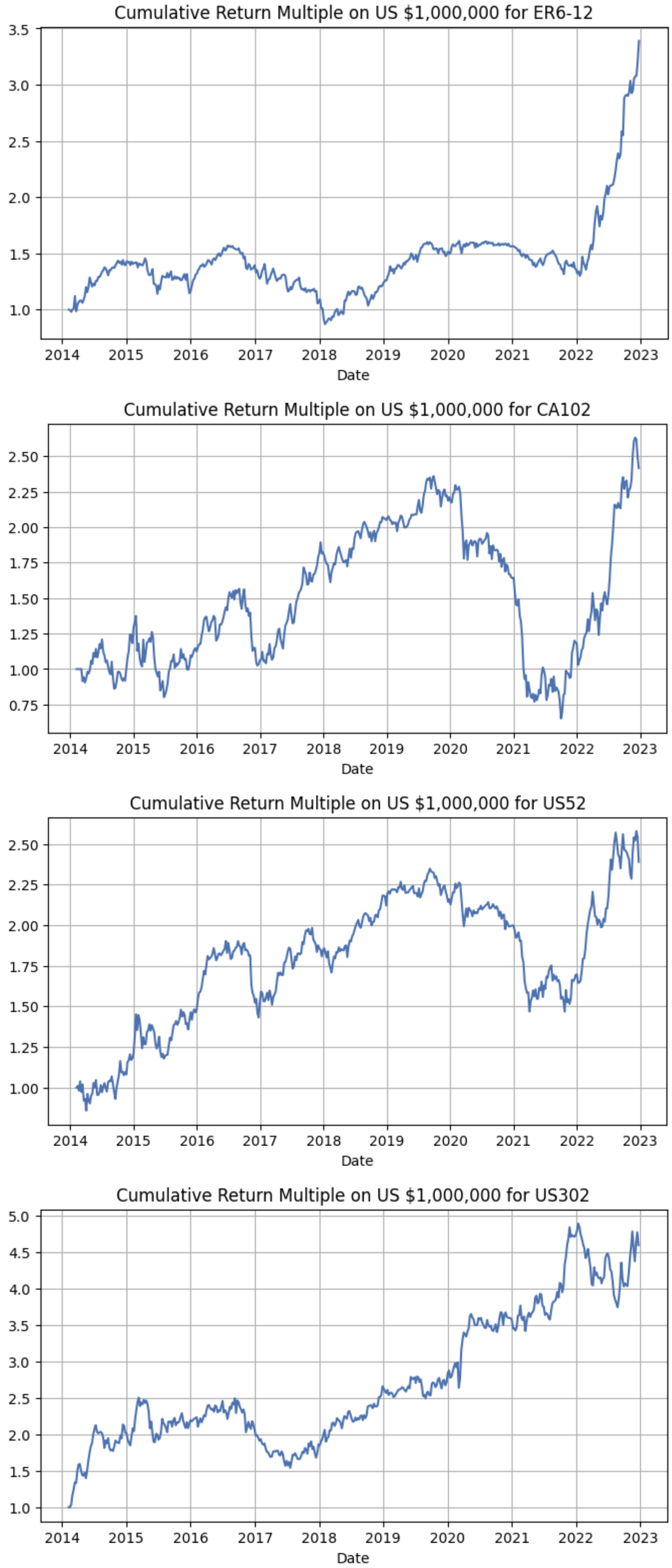
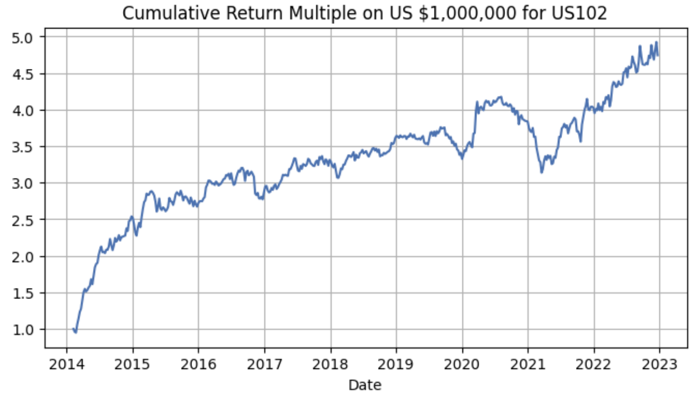
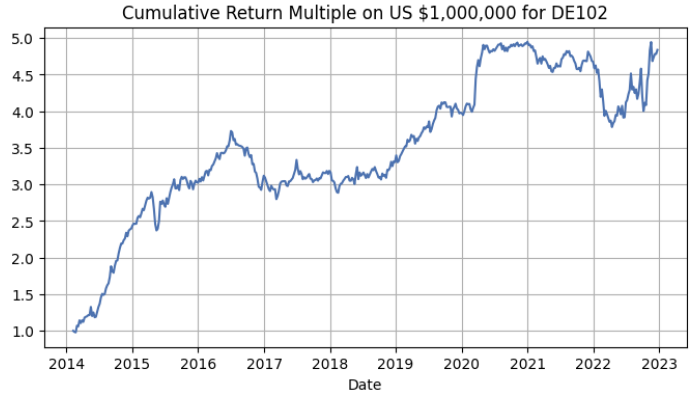


Figure 3: Cumulative returns for the top 8 performing spreads



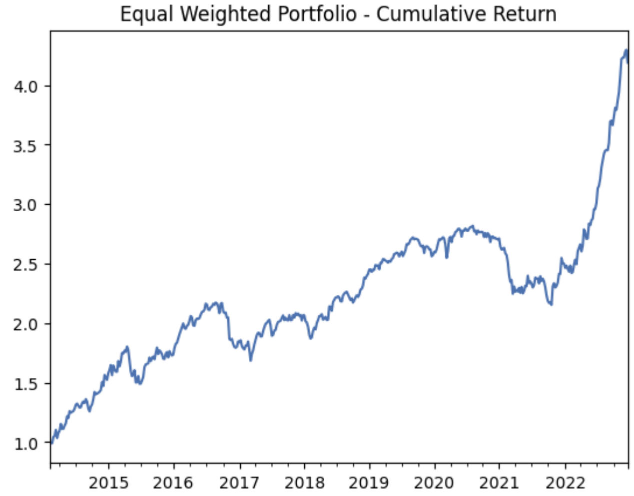

Cumulative Return Multiple on US $1,000,000 for SFR4-8
3.5
3.0
2.5
2.0
1.5
1.0
0.5
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Date

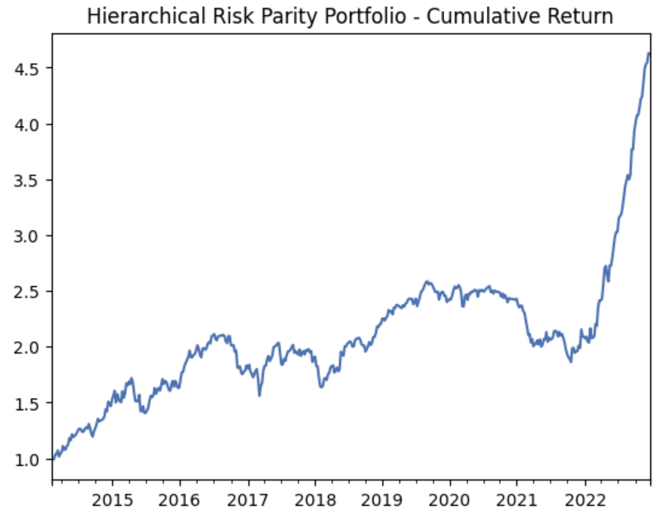

Cumulative Return Multiple on US $1,000,000 for US102
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Date


Cumulative Return Multiple on US $1,000,000 for DE102
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Date


Cumulative Return Multiple on US $1,000,000 for DE52
2.50
2.25
2.00
1.75
1.50
1.25
1.00
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Date

Figure 4: Cumulative return plots for each portfolio construction method





Acknowledgments

This paper was created as part of a WorldQuant University degree program towards an MSc in Financial Engineering. This paper is reproduced with the consent and permission of WorldQuant University.

References

[1] Ardea Investment Management. “A Primer on Interest Rate Markets and Relative Value – Part 1: Yield Curve Opportunities and Risks.” 2021 https://www.ardea.com.au/wp-content/uploads/ARDE-202101_Rates-Market-Primer-_fin.pdf

[2] Correlation Heatmaps with Hierarchical Clustering. https://www.kaggle.com/code/sgalella/correlation-heatmaps-with-hierarchical-clustering

[3] Di Maggio, Marco. “The Role of Hedge Funds in the 2020 Treasury Market Turmoil.” September, 2020. http://dx.doi.org/10.2139/ssrn.3698415

[4] Dickey, D. & Wayne Fuller. “Distribution of the Estimators for Autoregressive Time Series With a Unit Root.” *Journal of the American Statistical Association*, vol. 74, issue 366, 1979, pp.427-431.

[5] Eurozone Central Bank (ECB). “Recent Widening in Euro Area Sovereign Bond Yield Spreads.” *Eurozone Central Bank Monthly Bulletin*, November 2008. https://www.ecb.europa.eu/pub/pdf/other/mb200811_focus02.en.pdf

[6] Edwards, Franklin, R. "Hedge Funds and the Collapse of Long-Term Capital Management." *Journal of Economic Perspectives*, vol. 13, no. 2, Spring 1999, pages 189–210.

[7] Elliott, Robert J., John Van Der Hoek, & William P. Malcolm. “Pairs trading.” *Quantitative Finance*, vol 5(3), 2005, pp. 271–276. https://doi.org/10.1080/14697680500149370

[8] Estrella, Arturo and Frederic S. Mishkin. “The Yield Curve as a Predictor of U.S. Recessions.” *Current Issues in Economics and Finance*, vol. 2, no. 7, June 1996. https://www.newyorkfed.org/research/current_issues/ci2-7.html

[9] Fontaine, Jean-Sebastien, Guillaume Nolin. “Measuring Limits of Arbitrage in Fixed-Income Markets.” *Bank of Canada Staff Working Paper*, vol. 44, 2017. https://www.bankofcanada.ca/2017/10/staff-working-paper-2017-44/

[10] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor. “An Introduction to Statistical Learning with Applications in Python.” New York. Springer, 2023.

[11] Gatev, Evan, William N. Goetzmann, K. Geert Rouwenhorst. "Pairs Trading: Performance of a Relative-Value Arbitrage Rule," *Review of Financial Studies*, Oxford University Press for Society for Financial Studies, vol. 19, no. 3, 2006, pp. 797-827. https://www.nber.org/papers/w7032

[12] Giglio, Stefano. “Credit Default Swap Spreads and Systemic Financial Risk”. *European Systemic Risk Board*. 2014. https://www.esrb.europa.eu/pub/pdf/other/cds_spreads_2014.pdf

[13] Hess, Kurt. “Bond Relative Value Models and Term Structure of Credit Spreads: A Simplified Approach.” June 2004. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7a95a5b2435a5b40d507b0ad3ddbeb75d45cd95f

[14] Litterman, Robert, José Scheinkman. “Common Factors Affecting Bond Returns.” *Journal of Fixed Income,* June 1991.

[15] López de Prado, Marcos. “Building Diversified Portfolios that Outperform Out-of-Sample.” *Journal of Portfolio Management*, May 2016. https://doi.org/10.3905/jpm.2016.42.4.059.

[16] Martin, R. A. “PyPortfolioOpt: portfolio optimization in Python.” *Journal of Open Source Software*, vol. 6, no. 61, 2021, pages pp. 3066. https://doi.org/10.21105/joss.03066

[17] Parker, Dean and Moritz Schularick. “The Term Spread as a Predictor of Financial Instability.” *New York Federal Reserve, Liberty Street Economics*, November 2021. https://libertystreeteconomics.newyorkfed.org/2021/11/the-term-spread-as-a-predictor-of-financial-instability/

[18] Rollinger, Thomas and Scott T. Hoffman. “Sortino Ratio: A Better Measure of Risk.” *Futures Magazine*, February 2013, pp. 40-42.

[19] Sullivan, Rodney N, and Matthew Wey. “Global Macro and Managed Futures Hedge Fund Strategies: Portfolio Differentiators?.” *Darden Business School Working Paper*, November 2023. https://ssrn.com/abstract=4625978

[20] Wall, Michael E., Andreas Rechtsteiner, Luis M. Rocha. “Singular Value Decomposition and Principal Component Analysis.” *A Practical Approach to Microarray Data Analysis*, edited by D.P. Berrar, W. Dubitzky, M. Granzow, Kluwer: Norwell, MA, 2003, pp. 91-109.

[21] Young, Terry W. “Calmar Ratio: A Smoother Tool.” *Futures (Cedar Falls, Iowa)*, vol. 20, no. 11, 1991.