Yield Curve Modelling with Nelson-Siegel-Svensson

US Treasury Yields

A yield curve is a line that plots yields or interest rate bonds that have equal credit quality, but different maturity dates (eg. 1 month, 2 month… 30 year)

The US Treasury Yield curve is special because it is regarded as the benchmark curve. It is used often because it has no credit risk, meaning other bonds, such as corporate bonds can be benchmarked against it.

Distinct Behavior

Short term volatility: The short run is extremely volatile as it is heavily influenced by the Fed policy changes.

Midterm dip: Investors expectation of the Fed to cut rates, causes this dip.

20-year hump: The 20-year yield was reintroduced in 2020; however, the lack of demand for the yield causes it to require higher yields to attract buyers because of the lack of liquidity.

LR > SR: Normal yield curve indicating a healthy market.

Steep curve: When the spread between short and long yields widens, it means there is increased expectation in market growth and inflation.

Flat curve: When the spread between short and long yields tightens, it means there is economic uncertainty and potential slowdown.

SR > LR: Investors see short term bonds are riskier compared to long term bonds, meaning there are recessionary fears.

The Model

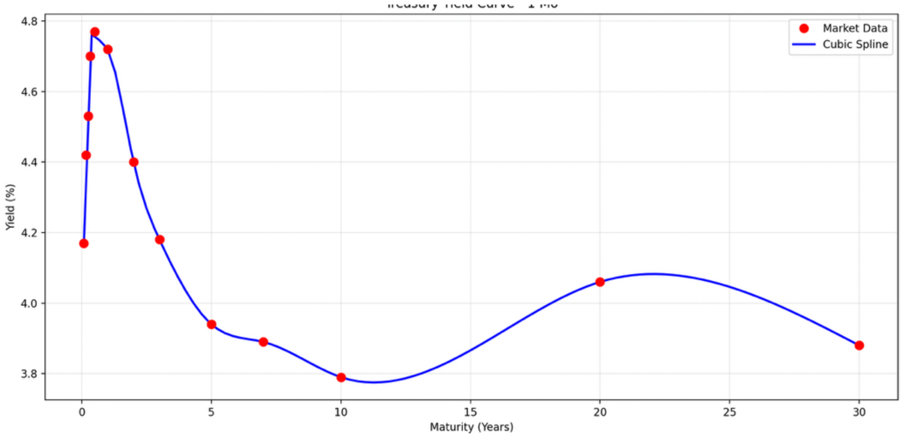
Using the 13 US Treasury yield data points since 2023 to create a curve using interpolation:

Nelson Siegel

Formula based, economically accurate parameters to build a curve with parameters β_0, β_1, β_2, λ_0.

Cubic Spline Interpolation

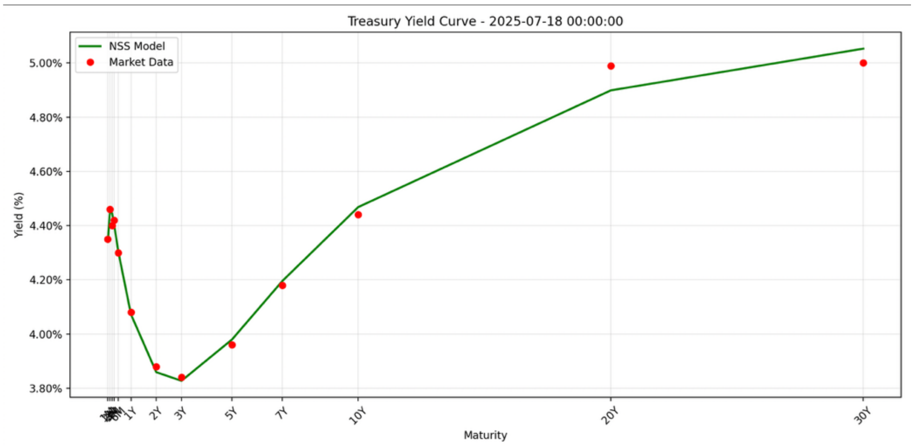
Maintains datapoints as first and second derivatives and creates a smooth and flexible curve by fitting cubic polynomials between each pair of points.



The issue with cubic spline: Over optimizes the yields, because there are no economic based parameters, meaning the curve built is arbitrary. Furthermore, if there are outliers in the yield data the spline will pass through the outliers creating oscillations.

The issue with Nelson Siegel: Having 4 parameters is hard to accurately build a curve that has 13 maturity points. Creating a curve that is both accurate for short term and long-term behaviors with limited parameters is challenging. If I try get the model to optimize well in the short run, the curve cannot follow the 10-year dip and 20-year hump. Vice versa if I try to focus on the long term yields the curve cannot follow the short-term volatility.

Nelson Siegel Svensson



Maturity	Market	NSS Model	Error	R^2
1M	4.35%	4.34%	0.01%	0.99
2M	4.46%	4.45%	0.01%	0.99
3M	4.40%	4.45%	0.05%	0.99
4M	4.42%	4.40%	0.02%	0.99
6M	4.30%	4.30%	0.00%	0.99
1Y	4.08%	4.07%	0.01%	0.99
2Y	3.88%	3.86%	0.02%	0.99
3Y	3.84%	3.83%	0.01%	0.99
5Y	3.96%	3.98%	0.02%	0.99
7Y	4.18%	4.20%	0.02%	0.99
10Y	4.44%	4.47%	0.03%	0.99
20Y	4.99%	4.90%	0.09%	0.99
30Y	5.00%	5.05%	0.05%	0.99

Thus, I resulted in using the Nelson Siegel Svensson model. It is an extended version of the original Nelson Siegel model; however, it has 2 additional parameters (β_3, λ_1) to handle 13 tenures. The extra parameters allow for more flexibility to follow the long-term yield data points while I focus on the short-term volatility.

Optimization Process

Nelder Mead Algorithm

Pattern search optimization algorithm with the following characteristics:

Derivative-Free Optimization (gradient free): Gradient based calculations can get stuck in local minima. This is not good because it means it is less flexible and cannot follow the volatile US treasury yields.

Handles Non-Linear Parameters well: The parameters used in NSS have complex interactions.

The Nelder Mead optimization uses a shape structure called the simplex (n+1) to find the lowest point. In this case, it uses a 7-vertex model as the Nelson Siegel Svensson has 6 parameters. Each vertex is evaluated using the error function. The vertex with the worst function value is removed and replaced with another point with a better value obtained by the error function. The algorithm performs transformations (reflection, expansion, contraction) to generate new candidate vertices. The error function once again evaluates the new vertices. The process continues until there is a minimal error.

Error Function

It uses a residual difference between an observed value and the value predicted by the model. In this case, it takes the market value vs the predicted value from the Nelson Siegel Svensson model. This happens for all 13 maturities, which are summed up to create a single number representing the error.

Running Issues with the Nelder Mead

While using the Nelder Mead function, I ran into a few issues with dealing with the parameters. Manually creating bounds were critical as the parameters without bounds would go "crazy". For example, Day 1: Parameters might be [2.5, -1.2, 0.8, 1.5, 0.5, 2.1], then on Day 2: Parameters explode to [45000, -0.0001, 890, 0.00003, 67000, -23000]. With bounds parameters would stay within economically reasonable and meaningful ranges.

Another issue I ran into was the optimization being stuck at a local minima. To handle this, I manually adjusted the lambda for the model to move around multiple local minima and look at overall global minima.

Fine Tuning the NSS

Working with the Nelson Siegel Svensson was like tuning a guitar for the first time. Every time a parameter was changed, another parameter had to be changed, followed by the curve changing behavior. Using the knowledge of each parameter and the output from the plots, I was able to fine tune the curve.

Efficiency

Vectorized Data Preparation

I converted 2 years' worth of data from a CSV file into a pandas data frame. The data frame allows for sorting and converting to a NumPy array. The NumPy array allows for less memory usage and vectorized operations. As I had to create a curve for each day, the NumPy array vectorization allowed for all thirteen tenures to be optimized with the Nelson Siegel Svensson at one go.

Warm Start

Each day requires different parameters because the yields change daily, solved by iterating each day. However, to keep the optimal parameters I decided to reuse the previous day's parameters on the current day for a "warm start". As yields do not change drastically daily, the process allowed the optimization to be more efficient.

Applications

Relative Value Trading

The curve illustrates relative value opportunities. On a given day, if yields are above the curve, the bonds are undervalued meaning there is an opportunity to buy (higher yields → lower price), and vice versa.

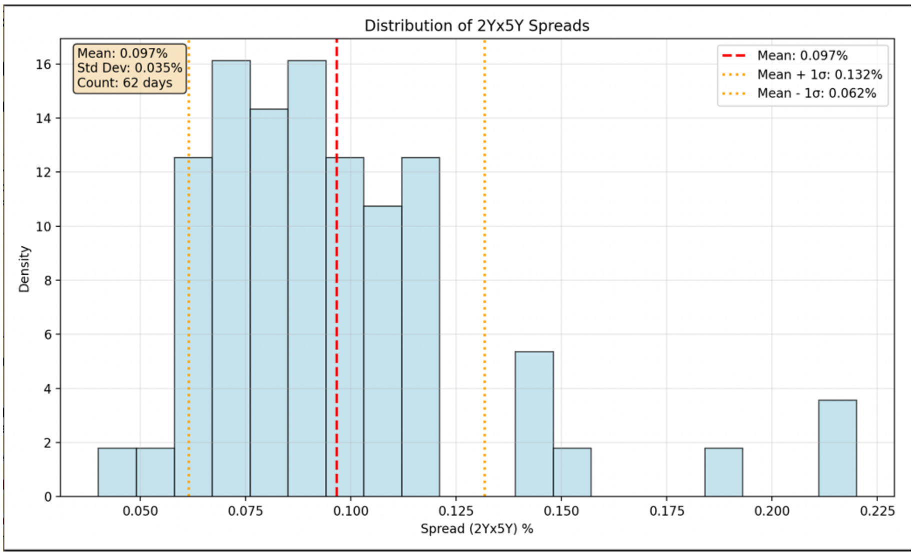
Spread Analysis

Mean Reversion

Using the past 3 months of data of 2y and 5y yields, I was able to get the spread (5y-2y) and find the average spreads.

The average resulted to be 0.097% (9.7 bps) and using the data I was able to calculate the standard deviation ±0.035% (±3.5 bps). The average can be used as the profit target and the surrounding 0.90 – 1 is partial profit taking.

This implies that mean reversion is using historical data, using the calculated mean and standard deviation to find trends and opportunities in the market. For example if the current price is above the average it is expected to fall.



Z Score Analysis

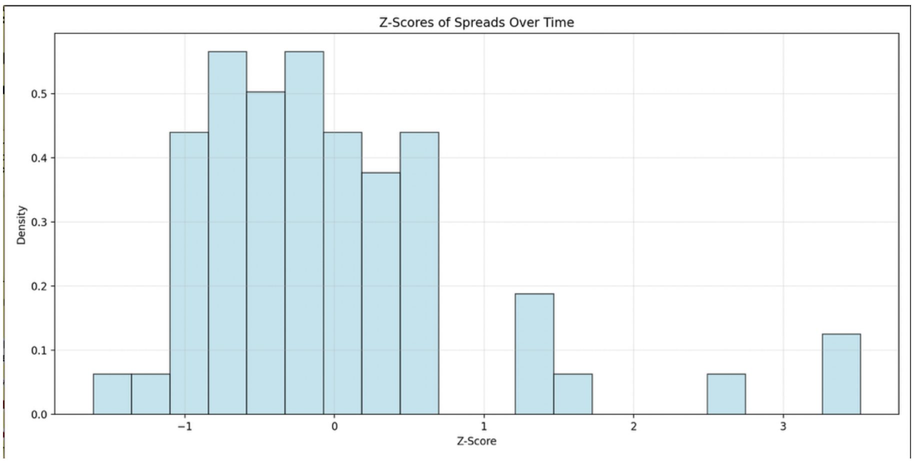
Converting the standard deviation to a z score (the distance between data point and the mean of a dataset), and creating a histogram, I can see that the normal range is between a score of -1 and 1.

The outside scores (Z > 2.5 or Z < -2.5) are trade opportunities and warnings.

If Z > 2.5:

Sell 2y and buy 5y and go long because the spread is wide. The expectation is the 5 year to decrease faster relative to the 2-year meaning there is trade opportunity to buy the 5 year and sell the 2 year.

If Z > 2.5: Buy 2y and sell 5y and go short because the spread is narrow. The expectation is the 5 year to increase faster relative to the 2-year meaning there is trade opportunity to sell the 5 year and buy the 2 year.



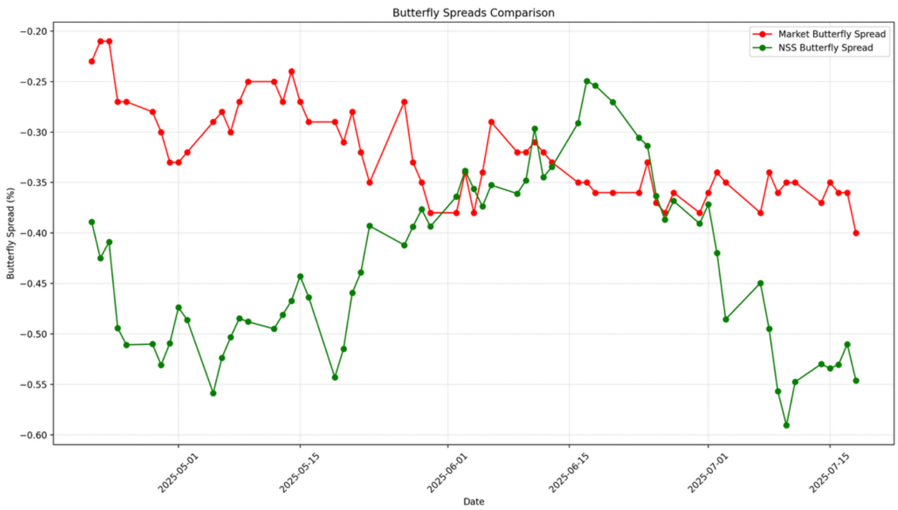
Butterfly Spread

A 2s5s10s butterfly trade is a fixed income trade that is designed to benefit from the changes in the yield curve. The curvature trade compares the 5Y against the movement of the 2Y and 10Y.

Using the NSS curve and the market, I can calculate 2 separate butterflies for 2s5s10s, which I can find the difference for trading opportunities.

If **market butterfly > NSS butterfly**: buy the body, sell the wings as market curve shows more curvature

If **NSS butterfly > market butterfly**: sell the body, buy the wings as the market curve is flat



Hedging

The wings of the butterfly need to be weighted to hedge such risks:

Unit Butterfly: The most common where both wings are equal and the body offsets the wings. For example, -1,2,-1.

Durational Risk: The goal is to make the butterfly duration neutral by weighting legs by DV01. The wings are weighted to hedge against parallel curve shifts.

Directional Risk: The goal is to limit exposure to steepening or flattening.

Regression Optimized: Using historical data to teach the model market behavior and minimize P&L variance.

There is a trade-off, it is not possible to hedge for both durational and directional so traders must choose their priority.

Finding Accuracy (R^2)

R^2 in NSS

Uses the error function output and the total sum of squares to calculate the accuracy of the optimized curve. The R^2 value in this case measures the quality of the distribution, thus the closer the value of R^2 is the better quality.

R^2 in Mean Reversion

Using linear regression, I was able to calculate when I can use mean reversion. The metric R^2 measures how much spread variation is explained by the linear time trend.

R^2 Close to 1: Spreads are trending, predictable, and have a stronger linear relationship.

R^2 Close to 0: There is oscillation, the relationship is non-linear, meaning there is an opportunity for mean reversion.

The slope obtained through line created by linear regression can be interpreted too.

Positive slope: widening

Negative slope: tightening

Minimal slope and minimal R^2 would be the optimal combination to execute mean reversion.

The spread analysis conducted resulted in a R^2 of 0.8889 and slope of 0.8878. The R^2 value shows 88.9% of spread changes follow a straight line over time and mean reversion is not recommended.

Appendix: Technical Terms

Models and Techniques

Nelson Siegel Model: A parametric yield curve model using 4 parameters (β_0, β_1, β_2, λ_0) to fit the term structure of interest rates with economic interpretation.

$$y(\tau) = ß_0 + ß_1\left(1 - e^{-\frac{\tau}{\lambda_0}}\right) + ß_2\left(1 - e^{-\frac{\tau}{\lambda_0}}\right)\left(-\frac{\tau}{\lambda_0}\right) - e^{-\frac{\tau}{\lambda_0}}$$

- τ: maturity
- β_0: long term yield level (10y – 30y)
- β_1: short vs long term (spread)
- β_2: curvature
- λ_0: decay, how effective the maturity is

Nelson Siegel Svensson (NSS) Model: Extended version of Nelson Siegel with 6 parameters (β_0, β_1, β_2, β_3, λ_0, λ_1) providing additional flexibility for fitting yield curves.

$$y(\tau) = ß_0 + ß_1\left(1 - e^{-\frac{\tau}{\lambda_0}}\right) + ß_2\left(1 - e^{-\frac{\tau}{\lambda_0}}\right)\left(-\frac{\tau}{\lambda_0}\right) - e^{-\frac{\tau}{\lambda_0}} + ß_3\left(1 - e^{-\frac{\tau}{\lambda_1}}\right)\left(-\frac{\tau}{\lambda_1}\right) - e^{-\frac{\tau}{\lambda_1}}$$

- β_3: secondary curvature
- λ_1: decay at secondary curvature

Cubic Spline Interpolation: Mathematical technique that fits piecewise cubic polynomials between data points while maintaining smooth first and second derivatives.

Mean Reversion: Statistical property where values tend to return to their long-term average over time.

Optimization and Algorithms

Nelder Mead Algorithm: Derivative-free optimization method that uses a simplex (geometric shape) to find minimum values of a function.

Residual: The difference between an observed value and the value predicted by the model.

$$y_i - \hat{y}_i$$

Error Function: The error function gives the optimizer a single value to minimize. The smaller the error function means a better fit.

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Simplex: In optimization, a geometric figure with n+1 vertices used to search for optimal solutions in n-dimensional space.

Local Minima: Points where a function reaches its lowest value within a small neighborhood, but not necessarily the global minimum.

Global Minima: The absolute lowest point of a function across its entire domain.

Warm Start: Optimization technique where previous solutions are used as starting points for new iterations to improve efficiency.

Bounds: Constraints placed on parameters to keep them within economically reasonable ranges.

Data Processing

Vectorized Operations: Computational technique that applies operations to entire arrays simultaneously, improving efficiency.

NumPy Array: Python data structure optimized for numerical computations with lower memory usage than lists.

Pandas DataFrame: Python data structure for handling structured data with rows and columns, similar to a spreadsheet.

Financial Terms

Yield Curve: Graph showing the relationship between bond yields and their time to maturity.

Basis Points (bps): Unit of measurement for interest rates, where 1 basis point = 0.01%.

Spread: The difference between yields of two different bonds or maturities (e.g., 5y-2y spread).

Relative Value: Investment strategy that identifies mispriced securities by comparing them to similar assets.

Tenure: The time remaining until a bond's maturity.

2s 5s 10s Butterfly: A 2s5s10s butterfly trade can be structured as a duration-neutral fixed income strategy that simultaneously buys and sells 2-year, 5-year, and 10-year bonds in specific proportions to profit from changes in yield curve curvature, while being hedged against parallel interest rate movements.

Duration: Duration measures the percentage change in a bond’s price for a 1% change in yield.

DV01 (Dollar Value per 01bp): The dollar P&L change for a 1 basis point yield move, per $1MM notional.

- 2Y: ~$200 per $1MM
- 5Y: ~$450 per $1MM
- 10Y: ~$850 per $1MM (approximate values)

Statistical Measures

Linear Regression: Statistical method for modelling the relationship between variables using a straight line.

Slope: In regression analysis, indicates the rate of change in the dependent variable per unit change in the independent variable.

Standard Deviation: Statistical measure of the amount of variation in a dataset.

$$\sigma = \sqrt{\frac{\sum(x_i - \mu^2)}{n}}$$

Z Score: Standardized score indicating how many standard deviations a data point is from the mean.

$$Z = \frac{x - \mu}{\sigma}$$

Market Terms

2y, 5y, 10y, 20y: Common abbreviations for 2-year, 5-year, 10-year, and 20-year Treasury bonds.

Undervalued/Overvalued: Securities trading below or above their fair value respectively.

Trade Opportunities: Market conditions that present potential for profitable trading.

R^2 (R-squared): Coefficient of determination that measures the proportion of variance in the dependent variable explained by the model, ranging from 0 to 1.

$$R^2 = 1 - SSR/TSS$$

Sum of Squared Residuals (SSR): Total squared differences between observed and predicted values; measures model error.

Total Sum of Squares (TSS): Total squared differences between observed values and their mean; measures total variation in data.

Trending vs. Oscillating Markets: Trending markets show persistent directional movement (high R^2), while oscillating markets show mean-reverting behavior around an average (low R^2)