Constructing the Risk-Free Yield Curve Using Symbolic Regression

Qinqin Huang

Rutgers University

Master of Quantitative Finance Program

December 7, 2025

Abstract

This paper evaluates a broad set of yield-curve construction methods, including log-linear interpolation, cubic splines, monotone–convex interpolation, Nelson–Siegel, Svensson, and a symbolic-regression–based estimator. Using U.S. Treasury and OIS data from 2020 to 2024, I compare these approaches along four dimensions: arbitrage constraints, forward-rate smoothness, calibration accuracy, and robustness to missing maturities. The empirical results show that shape-preserving and parametric models—particularly monotone–convex, Nelson–Siegel, and Svensson—deliver stable and economically consistent curves across varying market conditions. Linear interpolation is highly sensitive to the removal of individual maturities, while symbolic regression provides interpretable closed-form structures but exhibits higher pricing errors and localized irregularities without explicit constraints. Overall, symbolic regression serves as a flexible and interpretable complement to traditional methods rather than a standalone production-ready solution for yield-curve estimation.

Contents

1

Introduction

Constructing a smooth, robust, and arbitrage-free yield curve is essential for fixed-income pricing, risk management, and monetary policy analysis. Because market yields are available only at discrete maturities, interpolation or curve fitting is required, yet existing methods exhibit clear trade-offs: spline-based approaches offer high smoothness but may generate oscillatory or implausible forward rates, shape-preserving methods enforce monotonicity but limit flexibility, and parametric families such as Nelson–Siegel and Svensson remain parsimonious but may not adapt well to local curvature. Recent developments in symbolic regression provide a data-driven alternative capable of discovering interpretable functional forms, although its practical robustness in yield-curve construction is not well understood.

This study compares six methods—log-linear interpolation, cubic splines, monotone-convex interpolation, Nelson–Siegel, Svensson, and symbolic regression—using U.S. Treasury and OIS data. I evaluate their smoothness, forward-rate behavior, calibration accuracy, and robustness using a leave-one-knot-out test that assesses sensitivity to missing maturities. The results show that linear interpolation is the most sensitive to omitted data, while monotone-convex, cubic spline, and parametric models remain highly stable. Symbolic regression offers flexible functional forms but requires careful regularization to avoid overfitting. Overall, the analysis highlights the practical trade-offs across traditional, shape-preserving, parametric, and machine-learning-based yield-curve construction techniques.

2

Literature Review

Yield curve interpolation aims to construct smooth, arbitrage-free, and well-shaped term structures that reflect market data accurately while preserving theoretical consistency. A good interpolation method should avoid arbitrage, preserve monotonicity where appropriate, ensure local control, yield continuous forward curves, and provide robustness to changes in market data. Depending on the modeling objective and the pricing framework, interpolation may target different types of interest rates: spot rates, forward rates, or discount factors, each with different numerical and financial implications.

The most basic methods include linear interpolation and piecewise constant interpolation. These are easy to implement and ensure continuity in prices, but often fail to produce smooth forward rate curves. Log-linear interpolation on discount factors, as discussed in Hagan and West (2006), became widely adopted due to its simplicity and arbitrage-free guarantee. However, it may create unrealistic humps or kinks in the implied forward rate curve.

Cubic spline interpolation was introduced to improve smoothness by enforcing second derivative continuity. The natural cubic spline, proposed by McCulloch (1975), fits smooth yield curves by minimizing curvature, but it can produce spurious oscillations and even negative forward rates. To address this, the monotone convex method proposed by Hagan and West (2006) ensures shape preservation and monotonicity in discount factors, which results in a more robust forward curve, especially when market data is sparse or noisy.

Local interpolation methods such as Hermite interpolation and cubic Hermite splines offer greater control over curve shape by explicitly matching slope information at known data points. These methods allow the user to impose derivative constraints and can produce smoother forward rates without the oscillations typical of higher-degree splines.

Nelson and Siegel (1987) and Svensson (1994) proposed parametric functional forms that describe the term structure using a small number of parameters. These models are widely used in empirical finance and central bank applications due to their ability to

capture level, slope, and curvature dynamics in yields. However, they do not guarantee exact fit to market data unless overfitted and may lack flexibility under changing market regimes.

The Smith–Wilson method, introduced in Smith and Wilson (2001) and adopted under Solvency II (EIOPA, 2018), enables extrapolation of yield curves beyond observable maturities while enforcing convergence to a pre-specified ultimate forward rate. It guarantees arbitrage-free behavior but may introduce counterintuitive shapes depending on the choice of convergence speed and market anchors.

Each of these methods has trade-offs between smoothness, locality, monotonicity, and extrapolation behavior. Ultimately, the best choice depends on whether the objective is to exactly match observed quotes, preserve numerical stability during calibration, or enforce specific shape or asymptotic behavior.

Recent studies have explored machine learning (ML) as an alternative to traditional interpolation methods for constructing smooth and arbitrage-free yield curves. Traditional approaches—such as cubic splines, monotone convex interpolation, and Nelson–Siegel models—offer strong theoretical foundations and interpretability, but often struggle to capture complex, nonlinear market dynamics.

Carmona and Nadtochiy (2018) applied kernel smoothing techniques to estimate the yield curve directly from bond prices. Compared to spline methods, kernel approaches showed better local flexibility but required careful tuning of kernel bandwidths and suffered in sparse-data environments.

Richman and Scognamiglio (2022) proposed an attention-based deep learning model for multi-curve estimation. Their method outperformed classical models (e.g., VECM and Nelson–Siegel–Svensson) in out-of-sample forecasting, benefiting from its ability to capture temporal dependencies. However, the model lacked interpretability and required a large training set.

Pelger (2023) framed the problem as a conditional expectation and used neural networks to learn discount functions from cross-sectional and macroeconomic data. The ML model achieved superior accuracy and factor recovery, but did not inherently preserve financial properties like monotonicity or arbitrage-free behavior.

Although symbolic regression was not covered in the above yield-curve studies, advances in the SR literature suggest that it represents a promising compromise between traditional parametric models and black-box machine-learning approaches. Early work by Schmidt and Lipson (2009) demonstrated that genetic-programming–based SR can recover parsimonious analytical laws even from noisy data, illustrating its potential for discovering simple, interpretable structures that generalize well. More recent developments such as the AI Feynman algorithm of Udrescu and Tegmark (2020) further improved scalability and robustness by combining physics-inspired decomposition strategies with neural-network approximations and symbolic search. These advances highlight SR's abil-

ity to learn closed-form functional relationships that balance flexibility, interpretability, and extrapolation stability—properties that are particularly desirable in term-structure modeling.

However, SR remains computationally expensive, sensitive to hyperparameter choices, and difficult to constrain within strict no-arbitrage requirements. Despite these challenges, the method provides an attractive middle ground: it offers more structural flexibility than classical parametric models while retaining transparency and analytic tractability, features largely absent in deep learning or kernel-based methods.

In summary, ML-based interpolation methods provide better fit and adaptability in data-rich environments, but face persistent challenges in constraint enforcement, economic consistency, and transparency relative to traditional approaches. Symbolic regression, motivated by recent progress in interpretable machine learning, has the potential to bridge this gap by delivering flexible yet analytically coherent representations of the yield curve.

3

Methodology

3.1 Overview

This chapter presents the methodological framework for constructing the risk-free yield curve using symbolic regression (SR) and for comparing it with classical interpolation methods.

The symbols and notation used throughout the chapter are defined as follows:

Symbol	Definition
$P(t)$	Discount factor at time t
$y(t)$	Continuously compounded zero-coupon yield at time t
$f(t)$	Instantaneous forward rate at time t, defined as $f(t) = -\frac{d \ln P(t)}{dt}$
T_i	Maturities of observed market instruments
y_i	Observed yields at maturities T_i
N	Number of observed market instruments

3.2 Log-Linear Interpolation on Discount Factors

The log-linear interpolation assumes the natural logarithm of discount factors varies linearly between known maturities (Brigo & Mercurio, 2006; Hagan & West, 2006):

Let $T_i \leq T \leq T_{i+1}$ be the interval between two adjacent maturities with known discount factors $P_i = P(T_i)$ and $P_{i+1} = P(T_{i+1})$.

$$\ln P(T) = (1 - \theta) \ln P_i + \theta \ln P_{i+1}, \qquad \theta = \frac{T - T_i}{T_{i+1} - T_i}.$$

It guarantees $P(T) > 0$ and non-negative forward rates:

$$f(T) = -\frac{d \ln P(T)}{dT} = \frac{\ln P_i - \ln P_{i+1}}{T_{i+1} - T_i} \geq 0.$$

This approach is computationally efficient and arbitrage-free. However, because the forward rate is constant within each interval, the resulting forward curve lacks smoothness and may produce unrealistic kinks at knot points.

3.3 Cubic Spline Interpolation

Cubic splines approximate the term structure by fitting a piecewise cubic polynomial between adjacent maturities (McCulloch, 1975). For knots $T_1 < T_2 < \cdots < T_n$, the curve on each interval $[T_i, T_{i+1}]$ is written as

$$y(T) = a_i + b_i(T - T_i) + c_i(T - T_i)^2 + d_i(T - T_i)^3, \qquad T_i \leq T \leq T_{i+1},$$

with $h_i = T_{i+1} - T_i$. To determine the spline uniquely, the following conditions are imposed simultaneously. First, the spline interpolates the observed values,

$$y(T_i) = Y_i, \qquad i = 1, \ldots, n,$$

equivalently,

$$a_i = Y_i, \qquad a_{i+1} + b_{i+1}h_{i+1} + c_{i+1}h_{i+1}^2 + d_{i+1}h_{i+1}^3 = Y_{i+1}.$$

Second, first derivatives are continuous across all internal knots:

$$y_i'(T_{i+1}) = y_{i+1}'(T_{i+1}), \qquad i = 1, \ldots, n-1,$$

i.e.,

$$b_i + 2c_ih_i + 3d_ih_i^2 = b_{i+1}.$$

Third, second derivatives are continuous:

$$y_i''(T_{i+1}) = y_{i+1}''(T_{i+1}), \qquad i = 1, \ldots, n-1,$$

equivalently,

$$2c_i + 6d_ih_i = 2c_{i+1}.$$

Finally, two boundary conditions close the system. Common choices include the natural spline condition,

$$y''(T_1) = 0, \qquad y''(T_n) = 0,$$

the clamped spline condition with prescribed slopes,

$$y'(T_1) = m_1, \qquad y'(T_n) = m_n,$$

or the not-a-knot condition enforcing continuity of third derivatives,

$$d_1 = d_2, \qquad d_{n-2} = d_{n-1}.$$

In total, the spline has $4(n-1)$ unknown coefficients $\{a_i, b_i, c_i, d_i\}$ and is determined by n value-matching equations, $(n-2)$ first-derivative continuity equations, $(n-2)$ second-derivative continuity equations, and 2 boundary conditions, giving $4n - 4 = 4(n-1)$ equations in all. Although cubic splines generate C^2-continuous and visually smooth curves, they are non-local and may produce negative or oscillatory forward rates unless additional arbitrage constraints are imposed.

3.4   Nelson–Siegel

The Nelson–Siegel model (Nelson & Siegel, 1987) specifies the zero-coupon yield curve directly through a small number of latent factors with exponential-decay loadings. For a maturity t, the zero rate is given by

$$y(t) = \beta_0 + \beta_1 \left(\frac{1 - e^{-t/\tau}}{t/\tau} \right) + \beta_2 \left(\frac{1 - e^{-t/\tau}}{t/\tau} - e^{-t/\tau} \right),$$

where β_0 represents the long-term level of interest rates, β_1 determines the short-end slope, and β_2 captures medium-term curvature. The decay parameter τ controls the speed at which the short- and medium-term components lose influence as maturity increases.

Because the model yields closed-form expressions for zero rates and discount factors, it is computationally efficient and easy to calibrate to market instruments. Its factor loadings have clear economic interpretation, and the specification generates smooth, well-behaved term structures without artificial oscillations. These properties explain its widespread use in empirical fixed-income and macro-finance applications, including the estimation of monetary policy expectations, decomposition of yield movements, and forecasting exercises.

Nevertheless, the parsimony of the Nelson–Siegel model also limits its flexibility. With only one curvature factor, the model cannot represent yield curves with multiple humps or rapidly changing shapes. It may underfit the very short or very long ends of the curve, and during stressed market periods its exponential form can oversmooth noisy benchmark yields, resulting in poorer in-sample fit than spline-based or arbitrage-constrained techniques.

Despite these limitations, the Nelson–Siegel formula remains a leading benchmark due to its balance between smoothness, interpretability, and tractability, offering a robust low-dimensional representation of the term structure of interest rates.

The Nelson–Siegel–Svensson (NSS) model (Svensson, 1994) extends the original Nelson–

Siegel specification by adding a second curvature component, yielding a more flexible parametric representation of the zero-coupon yield curve. The zero rate for maturity t is given by

$$y(t) = \beta_0 + \beta_1 \left(\frac{1 - e^{-t/\tau_1}}{t/\tau_1} \right) + \beta_2 \left(\frac{1 - e^{-t/\tau_1}}{t/\tau_1} - e^{-t/\tau_1} \right) + \beta_3 \left(\frac{1 - e^{-t/\tau_2}}{t/\tau_2} - e^{-t/\tau_2} \right),$$

where β_0 again controls the long-term level, β_1 the slope, and β_2 and β_3 determine medium- and long-horizon curvature components. The decay parameters τ_1 and τ_2 allow the two curvature factors to peak at different maturities, enabling the model to capture richer shapes.

Like the Nelson–Siegel model, the NSS specification admits closed-form zero rates and discount factors, facilitating rapid calibration using least squares or maximum likelihood. The second curvature term substantially improves flexibility: NSS can generate yield curves featuring two humps or pronounced local deviations, a feature particularly valuable during monetary regime shifts or episodes of elevated term premia.

However, the improved flexibility introduces additional challenges. The interaction between τ_1 and τ_2 may create identification issues, especially when their values are close, and the optimization landscape becomes more nonlinear and sensitive to initialization. Moreover, like Nelson–Siegel, the Svensson model remains a purely parametric representation that does not automatically enforce absence of arbitrage, in contrast with spline-based monotone convex methods or HJM-consistent approaches.

Despite these limitations, the NSS model remains one of the most widely adopted parametric yield-curve specifications in central banks and policy institutions due to its tractability, smoothness, and superior empirical fit relative to the three-parameter Nelson–Siegel version.

3.5 Hermite Interpolation

Hermite interpolation extends standard cubic spline methods by enforcing both function values and first derivatives at the endpoints of each interval. Given discount factors P_i and local slopes f_i at adjacent maturities T_i and T_{i+1}, the interpolated discount factor on the interval is

$$P(T) = h_{00}(\theta)P_i + h_{10}(\theta)(T_{i+1} - T_i)f_i + h_{01}(\theta)P_{i+1} + h_{11}(\theta)(T_{i+1} - T_i)f_{i+1},$$

where $\theta = \frac{T - T_i}{T_{i+1} - T_i}$ and the cubic Hermite basis functions are

$$h_{00}(\theta) = 2\theta^3 - 3\theta^2 + 1, \quad h_{10}(\theta) = \theta^3 - 2\theta^2 + \theta,$$
$$h_{01}(\theta) = -2\theta^3 + 3\theta^2, \qquad h_{11}(\theta) = \theta^3 - \theta^2.$$

This construction guarantees C^1 continuity of the discount curve and provides strong local control: modifying slope or level at a knot affects only the neighboring interval. As a result, Hermite interpolants avoid the oscillatory artifacts that can arise in higher-order spline methods.

The main limitation of Hermite interpolation is its reliance on accurate slope estimates f_i. In many fixed-income markets, quotes may be sparse or subject to microstructure noise, making numerical derivatives unstable and potentially introducing artifacts into the curve. Consequently, Hermite interpolation works well when reliable forward-rate information is available, but may be less robust than monotone or convexity-preserving spline schemes in noisy market environments.

3.6   Monotone–Convex Interpolation

The monotone–convex interpolation method of Hagan and West (2006) constructs a smooth, locally controlled, and arbitrage-free discount curve by ensuring that the implied forward rates remain positive and free of spurious oscillations. On each interval $[T_i, T_{i+1}]$, the instantaneous forward rate is modeled as a quadratic function

$$f(T) = f_i + \alpha_i(T - T_i) + \beta_i(T - T_i)^2,$$

where f_i denotes the forward rate at T_i, and the coefficients α_i and β_i are chosen to satisfy C^1 continuity across knot points while enforcing monotonicity or convexity of the forward curve when warranted by the data.

The construction guarantees that the discount factors remain strictly decreasing and that forward rates do not exhibit oscillatory artifacts common in higher-order spline methods. This leads to stable PV01 and bucket sensitivities, which is critical for risk management and hedging applications. Because each interval depends only on local slope information, the method provides strong local control without sacrificing smoothness.

In practice, the monotone–convex scheme has become the standard for interpolating OIS and swap curves, where smoothness, stability, and the absence of static arbitrage are essential requirements.

3.7   Symbolic Regression Framework

3.7.1   Concept and Rationale

Symbolic regression (SR) is an interpretable machine-learning framework that searches over the space of closed-form mathematical expressions to identify functional relationships directly from data. Unlike conventional models that estimate parameters within a

predetermined architecture, SR explores compositions of elementary operators (such as $+, -, \times, \div, \exp, \log$) and functional primitives relevant for term-structure modeling (e.g., $t, e^{-t/\tau}$). This approach effectively bridges classical parametric yield-curve models (such as Nelson–Siegel or Svensson) and flexible data-driven methods by allowing the functional form itself to be discovered rather than imposed.

Early work in symbolic regression demonstrated that genetic-programming–based approaches can recover parsimonious analytical structures even from noisy data, while more recent advancements improved scalability by combining decomposition strategies, function approximation, and symbolic search. In financial applications, SR has been used to learn parsimonious and interpretable representations of term structures, offering a transparent alternative to black-box methods. Because SR yields explicit analytic formulas, the resulting curves provide economic interpretability, stable extrapolation behavior, and ease of sensitivity analysis, making it a promising tool for yield-curve estimation and interest-rate modeling.

3.7.2 Algorithmic Workflow

Algorithm 1 Symbolic Regression Workflow

Require: Dataset $\{(t_i, y_i)\}_{i=1}^{N}$, operator set $\mathcal{O}$, primitive set $\mathcal{P}$
1: Initialize population $\mathcal{F}_0$ with random expressions built from $\mathcal{O}$ and $\mathcal{P}$
2: **for** generation $g = 1, 2, \ldots, G$ **do**
3:     **for** each expression $f \in \mathcal{F}_{g-1}$ **do**
4:         Fit f to data (optimize numeric constants)
5:         Compute loss $\mathcal{L}(f)$ (e.g., RMSE)
6:     **end for**
7:     Select top-performing expressions $\mathcal{S}_g$ according to fitness
8:     Generate new candidate expressions:
9:         **Crossover:** combine subtrees from two parents
10:         **Mutation:** randomly replace operators/primitives
11:     Form new population $\mathcal{F}_g$ from $\mathcal{S}_g$ and evolved candidates
12:     Apply symbolic simplification to each $f \in \mathcal{F}_g$
13:     Apply parsimony penalty to discourage excessive complexity
14: **end for**
15: Select final model $f^\star$ based on validation loss and no-arbitrage checks
16: **return** Closed-form expression $f^\star(t)$

3.7.3   Objective Function and Constraints

To jointly control goodness of fit, smoothness, and interpretability, the SR estimator minimizes a composite objective,

$$\mathcal{J}(f) = \mathrm{RMSE}(f) + \lambda_1 \int_0^{T_{\max}} \big(f''(t)\big)^2 \, dt + \lambda_2 \, \mathrm{Complexity}(f),$$

where the second term penalizes excessive curvature in forward rates and the third enforces parsimony by discouraging unnecessarily large expressions.

3.7.4   Model Validation

To ensure that the estimated discount curve is economically meaningful, numerically stable, and free of static arbitrage, I evaluate the fitted curve using several complementary validation checks. These diagnostics assess not only the quality of fit to market instruments, but also the structural properties of the term structure.

1.   Positivity of discount factors.   Discount factors represent present values and must remain strictly positive for all maturities:

$$P(t) > 0 \quad \forall \, t \geq 0.$$

I evaluate $P(t)$ on a dense maturity grid to ensure that no negative values arise, even under interpolation or extrapolation.

2. Monotonicity and no-arbitrage.   An arbitrage-free term structure must be non-increasing:

$$P'(t) \leq 0.$$

I compute discrete slopes $P(t_{k+1}) - P(t_k)$ across the maturity grid to detect upward movements, which would imply static arbitrage opportunities.

3. Forward-rate regularity.   The instantaneous forward rate,

$$f(t) = -\frac{d \ln P(t)}{dt},$$

should be well-defined, continuous, and free of spikes. I inspect forward-rate curves for discontinuities, extreme local variation, or implausible short-term behavior, all of which may indicate local overfitting.

4. Smoothness and shape stability. A realistic discount curve should avoid unnecessary oscillations. I assess smoothness through the continuity of $P(t)$ and $f(t)$, and check curvature patterns using second-difference diagnostics. Excessive curvature or oscillatory patterns indicate numerical instability or insufficient regularization.

5. Calibration accuracy. The fitted curve should reproduce market instruments with small pricing errors:

$$\text{Error}(T_i) = |P_{\text{model}}(T_i) - P_{\text{mkt}}(T_i)|.$$

Exact matching is not required—especially under noisy market quotes—but errors should be uniformly small and stable across maturities.

6. Robustness and sensitivity analysis. A reliable model should respond smoothly to small perturbations of input data. I conduct:

- *Leave-one-out tests*: removing one maturity and re-estimating the curve;

- *Shock tests*: shifting selected yields by small amounts (e.g., ± 1 bp);

- *Comparative analysis*: examining resulting changes in $P(t)$ and $f(t)$.

A stable model should exhibit only minor and smooth adjustments, without structural breaks or instability.

Overall, these validation criteria ensure that the estimated discount curve is monotone, smooth, robust, and aligned with market instruments, making it suitable for valuation, hedging, and risk-management applications.

4

Results

4.1 Data Description

The data used in this study consists of daily observations of U.S. Treasury yields and Overnight Indexed Swap (OIS) rates as well as treasuries trading data. The data source is WRDS (Wharton Research Data Services). The sample period spans from January 1, 2020 to December 31, 2024. The frequency of interpolation is quarterly as the computation cost is high.

4.2 Positivity and Monotonicity Checks

A valid discount curve must remain strictly positive and monotonically non-increasing to prevent static arbitrage. I evaluate these conditions point-wise along each fitted curve. Table 4.1 summarizes the results.

All methods satisfy the positivity constraint, as discount factors remain above zero for all maturities. However, only linear interpolation, monotone–convex interpolation, and the symbolic-regression curve maintain global monotonicity. In contrast, cubic spline, Nelson–Siegel, and Svensson methods generate local violations of monotonicity, particularly around data gaps or curvature changes. These upward "bumps" imply temporarily increasing discount factors, which are theoretically inconsistent and reflect inherent flexibility in these smooth parametric families.

The symbolic-regression curve is monotone in this sample, but this occurs without an explicit constraint. The learned closed-form expression happened to produce a globally decreasing discount function; monotonicity is therefore not guaranteed and may fail in other datasets unless explicit penalties or constraints are imposed.

Table 4.1: Positivity and Monotonicity Checks

Method	Positive	Non-Increasing
Symbolic regression (yield)	True	True
Linear interpolation	True	True
Cubic spline	True	False
Nelson–Siegel	True	False
Svensson	True	False
Monotone convex	True	True

4.3   Forward-Rate Regularity

To assess whether each interpolation method produces smooth and economically plausible forward curves, I compute two diagnostics: (i) the volatility of forward-rate increments (the average absolute change between adjacent maturities), and (ii) the maximum single-step change. Smaller values indicate smoother and more stable forward-rate dynamics.

Table 4.2 reports results for a representative market date (2020-03-31), during which the yield curve exhibited strong stress and irregularities. Despite the volatile environment, all methods achieve low incremental variation, with forward-rate volatility centered around 0.005.

Nelson–Siegel and Svensson exhibit the smallest maximum jumps, reflecting the stabilizing effect of their parsimonious parametric structure. Linear interpolation displays the largest maximal change due to its piecewise-linear construction. Symbolic regression produces forward curves that are similarly smooth, although minor fluctuations may arise near densely clustered maturities, where the learned analytic expression becomes more sensitive.

Table 4.2: Forward-rate smoothness metrics on 2020-03-31

Method	Volatility	Max Change
Symbolic regression (yield)	0.005151	0.000372
Linear interpolation	0.005043	0.005198
Cubic spline	0.005665	0.000294
Nelson–Siegel	0.005243	0.000043
Svensson	0.004972	0.000146

Forward-rate plots across several dates further confirm that all methods produce continuous and well-behaved forward-rate curves. Parametric models remain especially stable, whereas linear interpolation introduces visible kinks, and symbolic regression—though

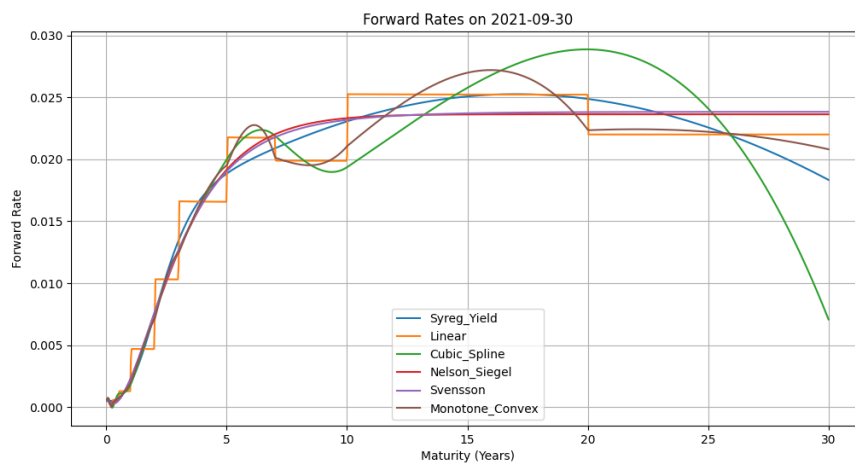generally smooth—may exhibit small local irregularities near key maturities.

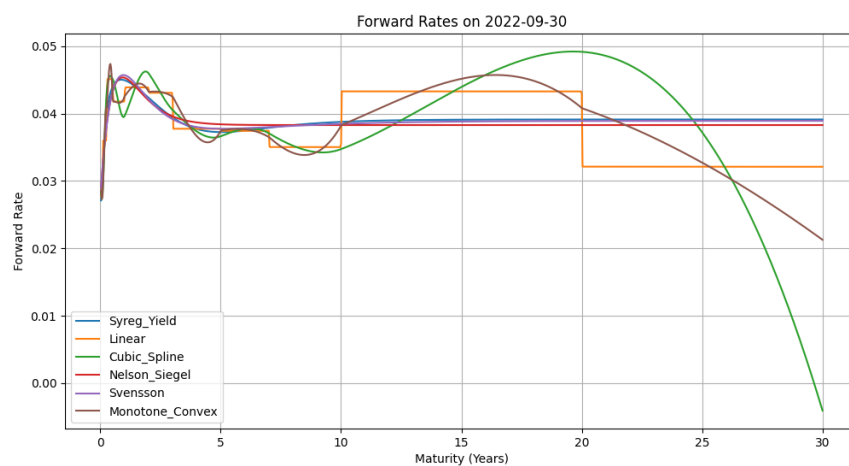Figure 4.1: Forward rates on 2021-09-30.
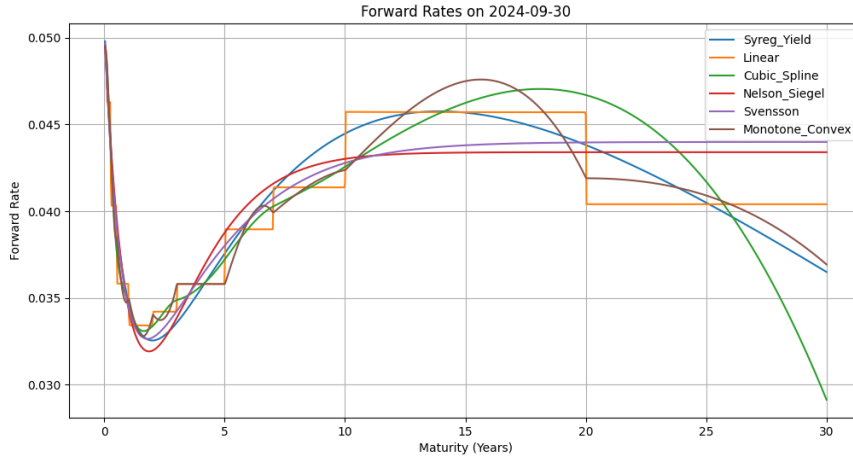

Figure 4.2: Forward rates on 2022-09-30.

Figure 4.3: Forward rates on 2024-09-30.

4.4   Smoothness and Shape Stability

To examine the behavior of different interpolation methods under varying market conditions, I selected two representative dates from the recent U.S. Treasury term structure. The first, 2024–09–18, exhibits a pronounced post-pandemic yield inversion with elevated short rates and depressed long-term yields. The second, 2025–02–11, reflects a more typical upward-sloping curve in which yields rise monotonically with maturity. These two contrasting shapes provide a useful stress test for assessing smoothness and shape stability across interpolation approaches.

Across both dates, linear interpolation produces piecewise-linear segments with visible kinks at key maturities, indicating limited smoothness and potential instability in forward rates. Cubic splines achieve higher smoothness overall but tend to overshoot between sparse long-maturity points, particularly in the 2024–09–18 case where the long end is nearly flat; forcing the curve to pass through every key point amplifies local oscillations. Monotone convex interpolation yields stable and economically plausible shapes in both environments, with strong local control and no artificial wiggles, though it may appear slightly rigid when the market curve contains sharp curvature.

Parametric models behave differently. Nelson–Siegel retains global smoothness but, due to its restrictive functional form, may fail to reproduce the true long-end level when observations are sparse—as seen on 2025–02–11. The Svensson extension allows additional curvature and fits better in both regimes, though the extra flexibility can introduce mild shape instability if the two decay parameters interact unfavorably.

Symbolic regression, while offering the advantage of fully data-driven functional discovery, shows mixed behavior. In regions with well-separated key maturities, it produces smooth and interpretable curves comparable to parametric models. However,

when adjacent points lie close together—especially around short maturities or during yield inversion—the learned expression may introduce localized irregularities or mild non-smoothness due to the model's effort to fit sharp transitions. This highlights a practical trade-off: symbolic regression provides structural flexibility and interpretability, but may require additional regularization or constraints to ensure stable behavior around dense clusters of input maturities.

Overall, the comparison across these two dates shows that interpolation methods differ materially in their ability to maintain smoothness, avoid spurious curvature, and produce robust shapes under changing market conditions. Yield inversion, in particular, stresses which models can accommodate sharp slopes without generating oscillations, while the upward-sloping curve tests long-end extrapolation and stability in sparsely observed regions.
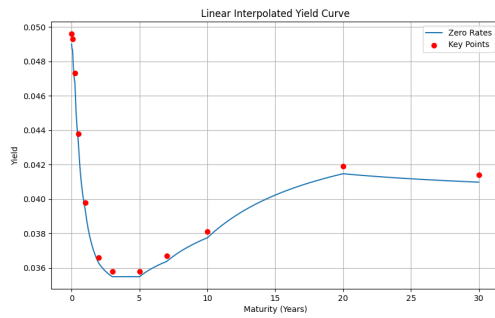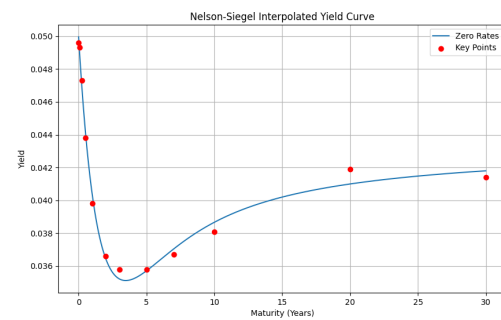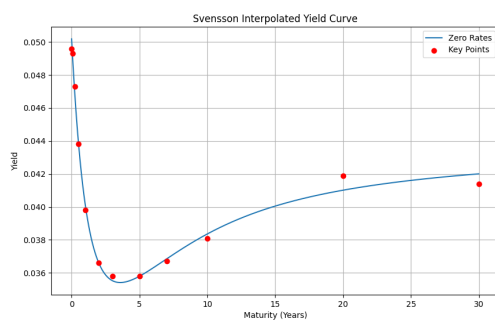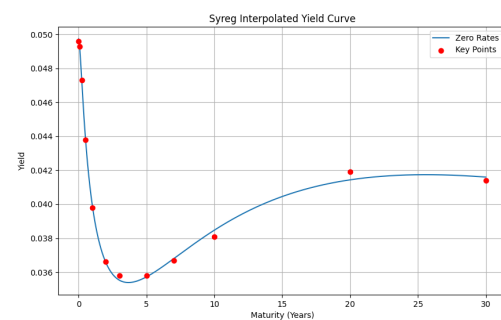
(a) Linear interpolation
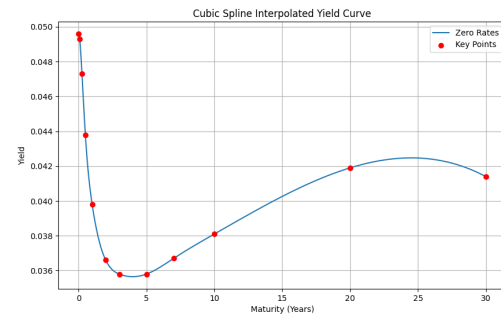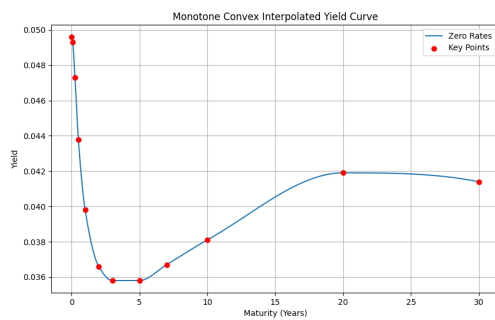
(b) Cubic spline interpolation

(c) Monotone convex interpolation
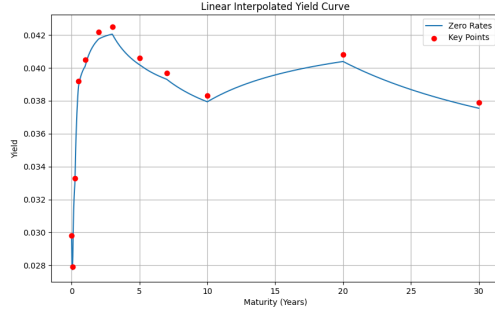
(d) Nelson–Siegel interpolation
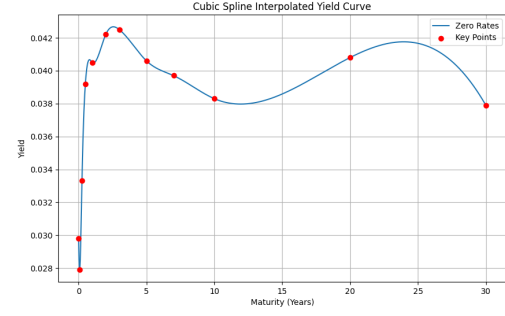
(e) Svensson interpolation
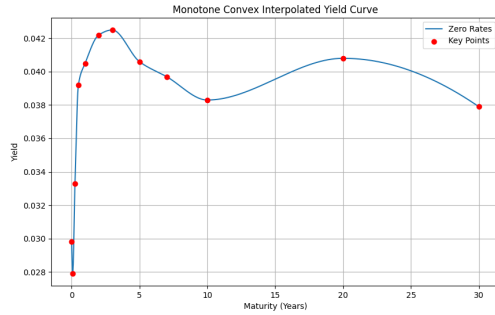
(f) Symbolic regression interpolation

Figure 4.4: Yield curve estimates on 2024-09-30 using six interpolation methods: linear, cubic spline, monotone convex, Nelson–Siegel, Svensson, and symbolic regression.
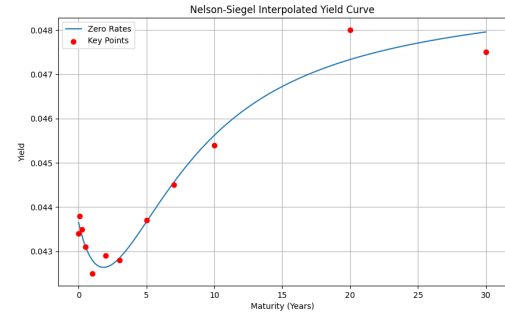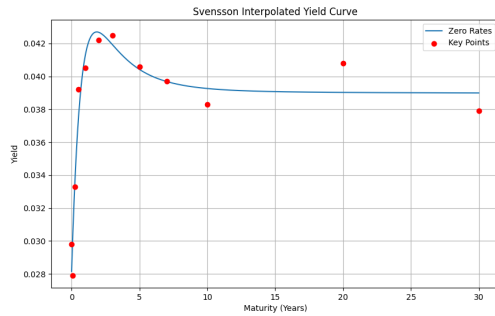
(a) Linear interpolation
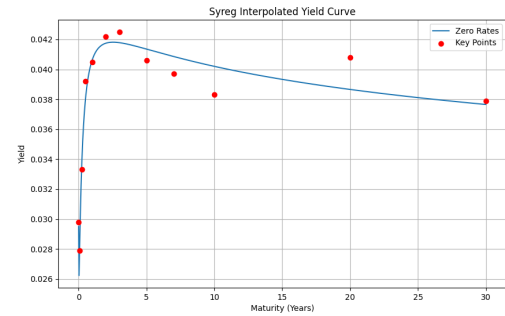

(b) Cubic spline interpolation

(c) Monotone convex interpolation

(d) Nelson–Siegel interpolation

(e) Svensson interpolation

(f) Symbolic regression interpolation

Figure 4.5: Yield curve estimates on 2022-09-30 using six interpolation methods: linear, cubic spline, monotone convex, Nelson–Siegel, Svensson, and symbolic regression.

4.5 Calibration Accuracy

Table 4.3 reports the average pricing errors across all evaluation dates. Overall, the differences across methods are small: MAE values fall within a narrow range of roughly 2.63–2.75 dollars and RMSE between 7.55–7.70 dollars. Linear and monotone–convex interpolation achieve the lowest mean absolute errors, while Nelson–Siegel, Svensson, and cubic spline perform comparably.

Symbolic regression fitted on yields exhibits slightly higher errors than the other methods, which reflects its emphasis on closed-form simplicity rather than exact replication of

quotes. Nevertheless, its MAE and RMSE remain close to the benchmarks, demonstrating that the method is still able to reproduce market prices reasonably well despite its structural constraints.

Table 4.3: Average Pricing Errors Across Sample Dates

Method	MAE	RMSE
Nelson–Siegel	2.7224	7.5510
Svensson	2.7049	7.5637
Cubic spline	2.6416	7.6768
Linear	2.6347	7.5773
Monotone convex	2.6354	7.5989
Symbolic regression (yield)	2.7508	7.6975

4.6   Robustness and Sensitivity Analysis

The leave-one-knot-out robustness test reveals meaningful differences across interpolation methods. Linear interpolation exhibits the largest sensitivity: removing a single key maturity produces an average MSE fluctuation of 1.051×10^{-3} and an MAE fluctuation of 7.87×10^{-4}, indicating that the method relies heavily on the presence of individual maturities.

Syreg_Yield performs moderately better, with average fluctuations of 6.40×10^{-5} (MSE) and 6.40×10^{-4} (MAE). However, its dependence on specific knot points is still notably stronger than that of the smoother or shape-preserving methods.

Cubic Spline, Nelson–Siegel, Svensson, and Monotone–Convex all show essentially zero MSE fluctuations when individual maturities are removed. Among these, Monotone–Convex achieves the smallest MAE fluctuation (3.0×10^{-6}), followed by Cubic Spline (8.0×10^{-6}). Svensson and Nelson–Siegel produce slightly larger MAE fluctuations (6.1×10^{-5} and 1.04×10^{-4}, respectively), but still exhibit strong robustness overall. These results indicate that shape-preserving or parametric curve families are substantially more stable under local data omissions than non-smooth alternatives such as the Linear or Syreg_Yield methods.

Table 4.4: Leave-One-Knot-Out Robustness Results

Method	Average MSE Fluctuation	Average MAE Fluctuation
Syreg_Yield	6.40×10^{-5}	6.40×10^{-4}
Linear	1.051×10^{-3}	7.87×10^{-4}
Cubic_Spline	0.0	8.0×10^{-6}
Nelson_Siegel	0.0	1.04×10^{-4}
Svensson	0.0	6.1×10^{-5}
Monotone_Convex	0.0	3.0×10^{-6}

5

Conclusion

This study evaluates six yield-curve construction methods using U.S. Treasury and OIS data from 2020–2024. While all approaches produce positive discount factors and broadly smooth forward rates, their performance differs in monotonicity, robustness, and calibration accuracy. Traditional methods such as monotone–convex interpolation and the Nelson–Siegel–Svensson family deliver stable, smooth, and economically consistent curves across market conditions. Linear interpolation is simple but exhibits the highest sensitivity to omitted maturities.

Symbolic regression (SR) offers clear advantages in interpretability, producing closed-form expressions that resemble parametric term-structure models. In several cases, SR generates smooth discount and forward curves without requiring an imposed functional form. However, the empirical results also show important weaknesses: SR displays slightly higher pricing errors than benchmark methods, exhibits sensitivity to closely spaced maturities, and can introduce local irregularities around the short end or inverted segments. Without explicit constraints, monotonicity and no-arbitrage properties are not guaranteed, making SR less reliable across datasets. In addition, computational cost and hyperparameter sensitivity limit its practical robustness.

Overall, SR functions better as a complementary modeling tool—useful for discovering interpretable functional structures—rather than as a stand-alone production curve. Future improvements should incorporate explicit shape constraints, stronger regularization, systematic hyperparameter search, and hybrid approaches that combine SR with established monotone or spline-based frameworks.

Bibliography

Brigo, D., & Mercurio, F. (2006). *Interest rate models: Theory and practice* (2nd). Springer.

EIOPA. (2018). *Technical documentation of the methodology to derive the ufr* (tech. rep.). European Insurance and Occupational Pensions Authority.

Hagan, P. S., & West, G. (2006). Interpolation methods for curve construction. *Applied Mathematical Finance*, *13*(2), 89–129.

McCulloch, J. H. (1975). The tax-adjusted yield curve. *Journal of Finance*, *30*(3), 811–830.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of Business*, *60*(4), 473–489.

Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, *324*(5923), 8185.

Smith, A., & Wilson, T. (2001). *Fitting yield curves with long term constraints* (tech. rep.). BACOB Research Paper.

Svensson, L. E. O. (1994). Estimating and interpreting forward interest rates: Sweden 1992–1994. *IMF Working Paper*, *94*(114).

Udrescu, S.-M., & Tegmark, M. (2020). AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, *6*(16), eaay2631.