Factor-Based Neural Network Forecasting of the German Yield Curve

rijksuniversiteit groningen

EUROPEAN CENTRAL BANK

Menno Westenbrink

S4847261

Supervisors:

Alexander Düring
European Central Bank

Artem Tsvetkov
Rijksuniversiteit Groningen

January 30, 2026

Contents

1 Introduction

The yield curve, which describes the relationship between interest rates and time to maturity for government bonds, plays a central role in modern finance. It serves as a benchmark for pricing fixed-income securities, reflects market expectations about future economic conditions, and responds dynamically to monetary policy decisions. For central banks such as the European Central Bank, with whom this research is conducted, accurate yield curve forecasts are essential for forward-looking policy decisions. Improved forecasting models enable policymakers to better anticipate future market developments and adjust monetary policy instruments proactively when necessary. For bond portfolio managers, these forecasts are equally important for making informed investment decisions, managing interest rate risk, and identifying relative value opportunities across different maturities. However, forecasting the yield curve presents significant challenges due to its high dimensionality and complex, time-varying dynamics.

Modeling and forecasting yields across numerous maturities simultaneously is computationally demanding and statistically inefficient. Moreover, yields are highly correlated as they respond to common macroeconomic factors. Factor models such as Nelson and Siegel (1987) and Litterman and Scheinkman (1991) have become the dominant framework by exploiting the empirical finding that a small number of factors can explain the vast majority of yield variation across maturities, effectively reducing a high-dimensional problem to a low-dimensional one.

Established approaches combine dimensionality reduction with time-series models to forecast these factors Diebold and Li (2006). While this framework has proven effective, it has two important limitations that motivate this research. First, traditional models assume linear relationships between past and future factor values, yet there is growing evidence that yield curve dynamics exhibit nonlinear patterns Kondratyev (2018). Second, these models rely exclusively on historical yield data, ignoring forward-looking market information that may contain valuable signals about future rate movements.

This thesis extends the established factor-based forecasting framework by addressing both limitations. This approach retains dimensionality reduction through principal component analysis presented in Litterman and Scheinkman (1991) but replaces linear forecasting models with artificial neural networks (ANNs), which can capture complex nonlinear relationships. Furthermore, we augment the standard set of predictors with forward-looking market variables: implied volatility from interest rate derivatives and market positioning data. Implied volatility from options on medium-term rates reflects market participants' expectations about future rate uncertainty and potential regime changes, while positioning data provides insights into informed trading flows and potential supply-demand imbalances that may drive future price movements.

The contribution is threefold. First, we investigate whether ANNs can improve out-of-sample forecasts of yield curve factors relative to the linear VAR benchmark, with particular attention to periods of elevated volatility or structural change when nonlinear dynamics may be most relevant. Second, we examine whether incorporating forward-looking market variables, specifically butterfly implied volatility and futures positioning data, adds predictive power beyond what can be extracted from historical yields alone. Third, we provide a comprehensive empirical evaluation of this augmented forecasting framework using out-of-sample tests, analyzing forecast accuracy across different horizons, market regimes, and specific yield curve movements. While prior research suggests that both neural networks and market-based variables hold promise for yield curve forecasting, the extent to which they improve upon established benchmarks remains an open empirical question that this thesis aims to address.

The remainder of this thesis is organized as follows. Section 2 reviews the relevant literature, tracing the evolution from parametric yield curve models to factor-based forecasting frameworks, and examining the application of machine learning methods to fixed-income markets. Section 3 describes the data sources and construction of the German zero-coupon yield curve. Section 4 compares two dimensionality reduction approaches—PCA and Nelson-Siegel—evaluating their reconstruction accuracy and testing the stability assumptions critical for forecasting applications. Section 5 investigates the forecastability of factors and betas within both frameworks using artificial neural networks, comparing performance against linear benchmarks. Section 6 concludes with a discussion of findings and directions for future research.

2 Literature Review

The literature on yield curve modeling and forecasting has evolved considerably over the past decades, progressing from parametric curve-fitting methods to factor-based forecasting frameworks, and more recently incorporating machine learning techniques. This review traces this evolution, beginning with foundational work on dimensionality reduction, proceeding through the development of dynamic factor forecasting models, and concluding with recent applications of neural networks to yield curve analysis. This review is structured to highlight how each contribution addresses specific aspects of the yield curve forecasting problem.

2.1 Parametric Dimensionality Reduction: The Nelson-Siegel Model

The challenge of modeling yield curves across numerous maturities was first addressed systematically by Nelson and Siegel (1987), who proposed a parsimonious functional form capable of generating the diverse shapes observed in practice. The Nelson-Siegel model represents the yield curve at time t for maturity τ as:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t}\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3,t}\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \tag{1}$$

where $\beta_{1,t}$, $\beta_{2,t}$, and $\beta_{3,t}$ are time-varying parameters, and λ is a fixed decay parameter that determines the maturity at which the loading on the curvature factor reaches its maximum. The elegance of this specification lies in its factor interpretation: $\beta_{1,t}$ represents the level (long-term yield), $\beta_{2,t}$ represents the slope (short-term vs. long-term spread), and $\beta_{3,t}$ represents the curvature (medium-term hump). By reducing an N-dimensional yield curve to just three parameters, Nelson and Siegel (1987) demonstrated that complex yield curve shapes could be parsimoniously represented while maintaining economic interpretability.

The key insight of the Nelson-Siegel model is that it achieves dimensionality reduction through an economically motivated functional form rather than through statistical methods. The three factors correspond to parallel shifts (level), tilting (slope), and bowing (curvature) of the yield curve, movements that practitioners had long recognized as the dominant modes of variation. However, the Nelson and Siegel (1987) approach imposes a specific functional structure on how these factors load across maturities, which may be restrictive if the true factor loadings differ from the assumed exponential decay pattern.

2.2 Statistical Dimensionality Reduction: Principal Component Analysis

Litterman and Scheinkman (1991) took a different approach to dimensionality reduction by applying principal component analysis (PCA) to a large dataset of U.S. Treasury yields. Rather than imposing a parametric structure, they let the data reveal the dominant factors driving yield curve variation. Their empirical analysis demonstrated that the first three principal components explain over 95% of the total variance in yield changes across maturities. Moreover, the factor loadings from PCA exhibit patterns strikingly similar to the Nelson and Siegel (1987) factors: the first principal component loads roughly equally across all maturities (level), the second principal component loads with opposite signs on short and long maturities (slope), and the third principal component loads most heavily on intermediate maturities (curvature).

This finding has profound implications for yield curve modeling. It suggests that despite the high dimensionality of observed yields, the effective dimensionality of yield curve movements is remarkably low. The PCA approach offers two advantages over parametric methods like Nelson and Siegel (1987). First, it requires no assumption about the functional form of factor loadings; the loadings are estimated directly from the data. Second, it provides a natural measure of how much variance each factor explains, allowing researchers to determine the appropriate number of factors empirically rather than assuming three factors a priori. The dominance of three factors in explaining yield variation has been confirmed across numerous markets and time periods, establishing factor models as the standard framework for yield curve analysis.

2.3 Dynamic Factor Forecasting: The Diebold-Li Framework

While Nelson and Siegel (1987) and Litterman and Scheinkman (1991) established that yield curves can be effectively represented by three factors, Diebold and Li (2006) made the crucial step of creating a forecasting framework by treating these factors as time series to be predicted. Diebold and Li (2006) combined the Nelson and Siegel (1987) factor structure with vector autoregression (VAR) models to forecast the entire yield curve. Their two-stage approach works as follows. First, at each point in time, they estimate the three Nelson and Siegel (1987) factors $\beta_{1,t}$, $\beta_{2,t}$, and $\beta_{3,t}$ by fitting equation (1) to observed yields. Second, they model the dynamics of these factors using a VAR(1) specification:

$$\begin{bmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \begin{bmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \\ \beta_{3,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix} \tag{2}$$

To generate forecasts h periods ahead, they iterate the VAR forward and then reconstruct the yield curve at each horizon using the Nelson and Siegel (1987) formula 1 with the forecasted factor values. This approach transforms a high-dimensional forecasting problem (predicting yields at many maturities) into a low-dimensional one (predicting three factors).

The key empirical finding of Diebold and Li (2006) is that this factor-based VAR approach significantly outperforms a simple random walk benchmark in out-of-sample forecasts, particularly at medium to long horizons. The random walk model, had proven difficult to beat in earlier research, making this result notable. The success of the Diebold and Li (2006) framework stems from exploiting the cross-sectional restrictions implied by the factor structure: rather than forecasting each maturity independently, the model recognizes that yields across maturities move together in predictable ways captured by the three factors. This framework has become the benchmark against which subsequent yield curve forecasting methods are evaluated.

However, the Diebold and Li (2006) approach has an important limitation: the VAR specification assumes that the relationship between past and future factor values is linear. While this assumption simplifies estimation and interpretation, it may be restrictive if yield curve dynamics exhibit regime changes, threshold effects, or other nonlinearities. Moreover, the model relies exclusively on the historical behavior of the factors themselves, ignoring potentially valuable information from other market variables that might help predict future movements.

2.4 Neural Networks for Yield Curve Analysis

The application of neural networks to yield curve modeling has taken several distinct forms, addressing different aspects of the curve analysis problem. We review three key contributions that collectively motivate this approach: one demonstrating the ability of neural networks to capture nonlinear dynamics in interest rate stress testing, one proposing autoencoders as a flexible alternative for dimensionality reduction, and one applying neural networks with market variables to bond return forecasting.

2.4.1 Capturing Nonlinear Dynamics: Kondratyev (2018)

Kondratyev (2018) demonstrated the potential of artificial neural networks to capture complex nonlinear relationships in interest rate markets through an application to yield curve stress testing. While stress testing differs from forecasting, it involves generating plausible adverse scenarios rather than predicting most likely outcomes, Kondratyev (2018)'s work provides important evidence that ANNs can model the nonlinear dynamics of yield curves effectively. Using a feedforward neural network architecture, Kondratyev (2018) showed that ANNs could learn the joint distribution of yield curve changes across maturities, including tail dependencies and non-Gaussian features that are difficult to capture with linear models.

The key insight from this work is that yield curve movements exhibit nonlinear patterns that vary with the market regime, level of rates, and volatility environment. During periods of monetary policy transitions or financial stress, the relationship between factors may differ substantially from normal periods. Neural networks, with their ability to approximate arbitrary nonlinear functions through hidden layers, offer a flexible framework for capturing these regime-dependent dynamics. While Kondratyev (2018)'s application focused on scenario generation rather than forecasting, the demonstrated capability of ANNs to model complex yield curve dynamics motivates their use in this forecasting context.

2.4.2 Nonlinear Dimensionality Reduction: Suimon et al. (2020)

Suimon et al. (2020) proposed using autoencoders, as an alternative to both parametric models like Nelson and Siegel (1987) and linear methods like PCA demonstrated by Litterman and Scheinkman (1991) for representing yield curve shapes. An autoencoder consists of an encoder network that compresses the input (yields at all maturities) into a low-dimensional representation (hidden layer), and a decoder network that reconstructs the original yields from this compressed representation. By training the network to minimize reconstruction error, the hidden layer learns a compact representation of the yield curve.

The architecture of their three-factor autoencoder can be expressed as:

$$\mathbf{y}_t \rightarrow \text{Encoder}(\mathbf{y}_t) = \mathbf{f}_t \rightarrow \text{Decoder}(\mathbf{f}_t) = \hat{\mathbf{y}}_t \tag{3}$$

where $\mathbf{y}_t$ is the vector of observed yields, $\mathbf{f}_t$ is the three-dimensional hidden layer representation, and $\hat{\mathbf{y}}_t$ is the reconstructed yield curve. The key finding is that the three hidden layer nodes learn to represent factors corresponding to level, slope, and curvature, similar to PCA Litterman and Scheinkman (1991) and Nelson and Siegel (1987), but through a nonlinear transformation. This demonstrates that the three-factor structure is robust across different dimensionality reduction methods.

However, it is crucial to recognize what Suimon et al. (2020)'s approach does and does not accomplish. The autoencoder provides a flexible way to represent or fit yield curves at a point in time, it addresses the dimensionality reduction component of the yield curve modeling problem. But it does not address forecasting: the autoencoder takes today's curve as input and outputs a compressed representation of today's curve. Extending this to forecasting would require a separate model to predict the evolution of the hidden layer factors over time. In this framework, we use PCA demonstrated by Litterman and Scheinkman (1991) rather than autoencoders for the dimensionality reduction step, as PCA is simpler, more interpretable, and computationally efficient, while we deploy neural networks for the forecasting step where their ability to capture nonlinear dynamics is most valuable.

2.4.3 Neural Network Forecasting with Market Variables: Verner et al. (2021)

Verner et al. (2021) applied nonlinear autoregressive neural networks (NAR and NARX architectures) to forecast long-term bond prices, representing one of the first attempts to incorporate external market variables into neural network-based fixed-income forecasting. Their approach involved forecasting individual bond prices (rather than yield curves) using two types of models: a pure autoregressive network using only past bond prices, and an augmented network incorporating 50-year EUR interest rate swaps and the VIX volatility index as external inputs.

The NARX (Nonlinear Autoregressive with eXogenous inputs) architecture they employed can be represented as:

$$P_t = f(P_{t-1}, P_{t-2}, ..., P_{t-d}, X_t, X_{t-1}, ..., X_{t-d}) + \varepsilon_t \tag{4}$$

where P_t is the bond price, X_t represents external variables (swap rates and VIX), and $f(\cdot)$ is approximated by a feedforward neural network with multiple hidden layers. Their results showed that both the Levenberg-Marquardt and Scaled Conjugate Gradient training algorithms achieved high in-sample fit ($R^2 > 95\%$), demonstrating the capability of neural networks to model bond price dynamics.

However, their key finding regarding external variables was somewhat negative: incorporating VIX and swap rates provided only marginal improvement over models using price history alone. This result has several potential explanations. First, their focus on individual bond prices rather than the entire yield curve structure means they did not exploit the cross-sectional information that factor models capture. Second, VIX represents equity market volatility and may not be the most relevant measure for fixed-income forecasting; curve-specific measures like swaption implied volatility might be more informative. Third, positioning or order flow data, which can capture informed trading activity, was not included in their analysis.

This research differs from Verner et al. (2021) in several important respects. First, we forecast the entire yield curve structure through factors rather than individual bond prices, exploiting the systematic co-movement of yields across maturities. Second, we use curve-specific implied volatility measures (butterfly volatility from interest rate options) rather than equity market volatility, which should be more directly relevant to yield curve dynamics. Third, we incorporate futures positioning data, which captures informed trader expectations and potential supply-demand imbalances not reflected in price history alone. Finally, we conduct rigorous out-of-sample testing across different forecast horizons and market regimes to evaluate whether these augmentations provide genuine predictive improvements.

3 Data

3.1 Yield Curve Data

The zero-coupon yield curve data for Germany is obtained from IBOXX historical bond data, which have been used to obtain the zero coupon yields. The final dataset spans from 4 January 1999 to 28 January 2026, comprising 6957 observations (weekends excluded) across 15 maturity points ranging from 1 to 30 years.

Table 1 presents summary statistics for the zero-coupon yields across all maturities. The sample period captures multiple interest rate regimes, including the pre-financial crisis era, the European sovereign debt crisis, a prolonged low-interest-rate environment, and the recent monetary policy normalization.

Figure 1 visualizes the evolution of the German zero-coupon yield curve over the sample period.

Table 1: Summary Statistics of German Zero-Coupon Yields

Maturity	Mean	Std. Dev.	Min	Q25	Median	Q75	Max	N
1Y	0.015	0.018	−0.009	−0.002	0.015	0.030	0.051	6,957
2Y	0.016	0.018	−0.010	−0.002	0.017	0.030	0.052	6,957
3Y	0.017	0.018	−0.010	−0.002	0.019	0.030	0.052	6,957
4Y	0.018	0.018	−0.010	−0.001	0.021	0.033	0.052	6,957
5Y	0.019	0.018	−0.010	0.000	0.022	0.035	0.053	6,957
6Y	0.020	0.018	−0.010	0.002	0.023	0.036	0.053	6,957
7Y	0.021	0.018	−0.010	0.003	0.023	0.037	0.055	6,957
8Y	0.022	0.018	−0.009	0.004	0.024	0.039	0.055	6,957
9Y	0.023	0.018	−0.009	0.005	0.025	0.039	0.056	6,957
10Y	0.024	0.018	−0.009	0.006	0.025	0.040	0.056	6,957
12Y	0.026	0.018	−0.008	0.008	0.027	0.041	0.057	6,957
15Y	0.027	0.017	−0.008	0.010	0.028	0.042	0.058	6,957
20Y	0.029	0.018	−0.007	0.011	0.030	0.045	0.059	6,957
25Y	0.031	0.019	−0.006	0.013	0.030	0.047	0.068	6,957
30Y	0.030	0.018	−0.005	0.014	0.030	0.045	0.063	6,957

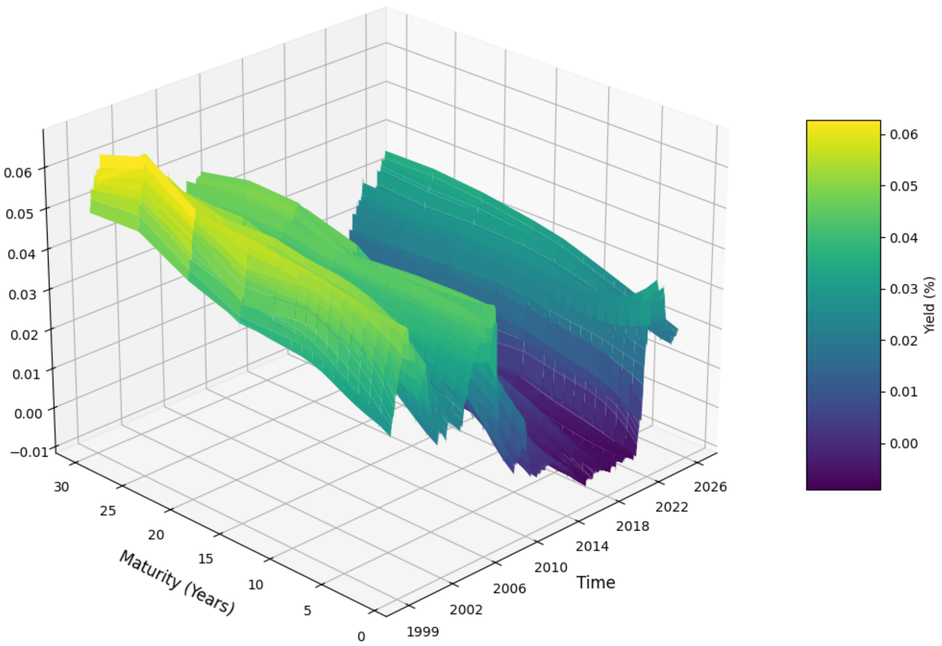


Figure 1: Evolution of the German Zero-Coupon Yield Curve

3.2 Swaption Implied Volatility Data

The implied volatility data is derived from 3-month swaption prices obtained from the statistics database within the ECB. It spans the same sample period as the yield curve data, from January 4, 1999 to January 28, 2026. For dates with missing implied volatility observations (96 dates), values are interpolated using linear interpolation to ensure a complete time series.

Linear interpolation estimates missing values IV_t at time t based on the nearest available observations before and after the missing date:

$$IV_t = IV_{t-k} + \frac{t - (t-k)}{(t+j) - (t-k)} \left(IV_{t+j} - IV_{t-k} \right) \tag{5}$$

where IV_{t-k} and IV_{t+j} are the nearest observed values before and after time t, respectively, with k and j denoting the number of days from t to these observations.

Table 2 presents summary statistics for the 3-month swaption implied volatility.

Figure 2 illustrates the time series evolution of the 3-month swaption implied volatility over the sample period.

Table 2: Summary Statistics of 3-Month Swaption Implied Volatility

Variable	Mean	Std. Dev.	Min	Q25	Median	Q75	Max	N
3M IV	66.632	25.103	26.340	49.580	62.310	79.500	180.310	6,957

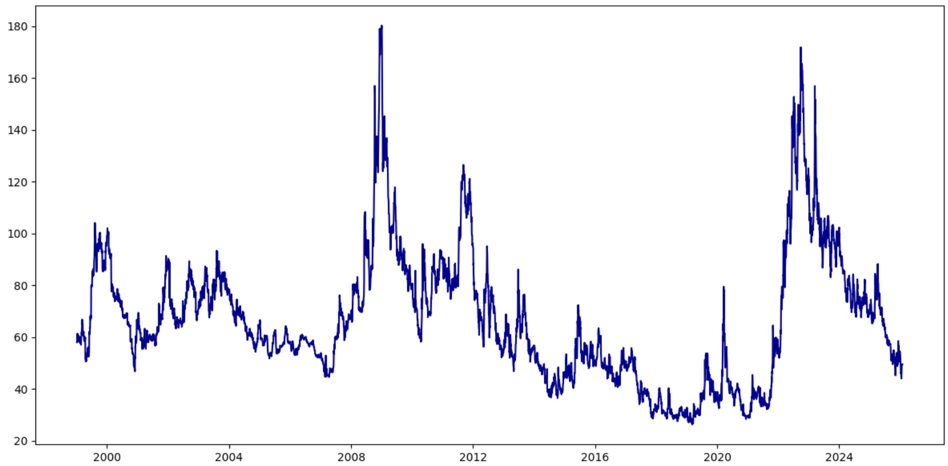


Figure 2: Evolution of 3-Month Swaption Implied Volatility

4 Yield Curve Fitting

4.1 Motivation

Forecasting the entire term structure of interest rates presents a high-dimensional challenge: predicting 15 distinct maturities simultaneously requires modeling complex dependencies. We address this by employing dimensionality reduction techniques that exploit the strong co-movement of yields across maturities. We evaluate three approaches: Principal Component Analysis on yield levels (Level-PCA), Principal Component Analysis on yield changes (Changes-PCA), and the Nelson-Siegel (NS) parametric model. All methods compress the yield curve into a lower-dimensional representation, enabling more efficient forecasting while preserving essential term structure dynamics.

4.2 Principal Component Analysis Approach

4.2.1 Mathematical Framework

Let $\mathbf{y}_t \in \mathbb{R}^{15}$ denote the vector of zero-coupon yields at time t. For Level-PCA, we apply PCA directly to yield levels; for Changes-PCA, we apply PCA to yield changes $\Delta\mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$. In both cases, the covariance matrix eigenvalue decomposition yields:

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{\Lambda}\mathbf{L}^T \tag{6}$$

where $\mathbf{L} \in \mathbb{R}^{15\times k}$ contains the loadings for $k = 4$ factors. For each yield curve (or change vector), we extract factors via:

$$\mathbf{f}_t = \mathbf{y}_t\mathbf{L} \quad \text{or} \quad \mathbf{f}_t = \Delta\mathbf{y}_t\mathbf{L} \tag{7}$$

and reconstruct as:

$$\hat{\mathbf{y}}_t = \mathbf{f}_t\mathbf{L}^T \quad \text{or} \quad \widehat{\Delta\mathbf{y}}_t = \mathbf{f}_t\mathbf{L}^T \tag{8}$$

For forecasting, we estimate loadings $\mathbf{L}_t$ using a rolling window of $\tau = 756$ trading days (3 years) and make the key assumption:

$$\mathbf{L}_{t+h} \approx \mathbf{L}_t \tag{9}$$

where $h = 21$ trading days (1 month). This enables a clean separation: (1) extract factors at time t using $\mathbf{L}_t$, (2) forecast factors $\hat{\mathbf{f}}_{t+h}$, and (3) reconstruct the yield curve using the same loadings $\mathbf{L}_t$.

4.2.2 Empirical Validation

We test assumption (9) using the German zero-coupon yield data described in Section 2, yielding 6,183 observations for Level-PCA and 6,182 for Changes-PCA. Tables 3 and 4 show reconstruction errors when using "stale" loadings from time t to decompose yields at time $t + 21$.

Table 3: Level-PCA Reconstruction Errors (Future Yields, Stale Loadings)

Statistic	RMSE	MAE	TAE
Mean	0.001091	0.000851	0.012767
Median	0.000646	0.000472	0.007084
Std Dev	0.000881	0.000716	0.010740
Min	0.000117	0.000094	0.001405
Max	0.003933	0.003332	0.049973

Table 4: Changes-PCA Reconstruction Errors (Future Changes, Stale Loadings)

Statistic	RMSE	MAE	TAE
Mean	0.000029	0.000022	0.000333
Median	0.000021	0.000016	0.000246
Std Dev	0.000038	0.000025	0.000371
Min	0.000000	0.000000	0.000000
Max	0.001847	0.000930	0.013946

To directly assess loading stability, we compute the ratio of reconstruction errors using stale versus fresh loadings. Tables 5 and 6 present these ratios.

Table 5: Level-PCA Loading Stability Ratios (Stale/Fresh)

Statistic	RMSE Ratio	MAE Ratio	TAE Ratio
Mean	1.031	1.032	1.032
Median	1.008	1.016	1.016
Std Dev	0.192	0.193	0.193
25th Percentile	0.950	0.950	0.950
75th Percentile	1.086	1.090	1.090

The median ratios indicate that using one-month-old loadings degrades reconstruction quality by 0.8% for Level-PCA and only 0.3% for Changes-PCA, providing strong support for assumption (9). Changes-PCA exhibits superior loading stability, suggesting that yield curve dynamics are more stable when measured in first differences.

Table 6: Changes-PCA Loading Stability Ratios (Stale/Fresh)

Statistic	RMSE Ratio	MAE Ratio	TAE Ratio
Mean	1.023	1.019	1.019
Median	1.003	1.003	1.003
Std Dev	0.191	0.139	0.139
25th Percentile	0.994	0.993	0.993
75th Percentile	1.019	1.020	1.020

4.3 Nelson-Siegel Approach

4.3.1 Mathematical Framework

The Nelson-Siegel model (Nelson and Siegel, 1987) parametrically represents the yield curve with three factors (level, slope, curvature) and a decay parameter λ:

$$y(\tau) = \beta_0 + \beta_1 \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \tag{10}$$

where τ denotes maturity. In matrix form, yields are reconstructed as:

$$\mathbf{y}_t = \mathbf{L}(\lambda_t)\boldsymbol{\beta}_t \tag{11}$$

where $\mathbf{L}(\lambda_t) \in \mathbb{R}^{15\times 3}$ contains the maturity-dependent loadings determined by λ_t, and $\boldsymbol{\beta}_t = [\beta_0, \beta_1, \beta_2]^T$ are the factor weights.

For forecasting, we estimate the optimal λ_t by minimizing reconstruction error over the same 3-year rolling window used for PCA. The key assumption is:

$$\lambda_{t+h} \approx \lambda_t \tag{12}$$

Given stable λ, we can refit the betas $\boldsymbol{\beta}_{t+h}$ at the forecast horizon while maintaining consistent yield curve structure.

4.3.2 Empirical Validation

We test assumption (12) using the same data, yielding 6,183 observations. Table 7 shows reconstruction errors when using stale λ from time t to fit yields at time $t + 21$.

Table 7: Nelson-Siegel Reconstruction Errors (Future Yields, Stale Lambda)

Statistic	RMSE	MAE	TAE
Mean	0.000642	0.000504	0.007561
Median	0.000535	0.000428	0.006419
Std Dev	0.000374	0.000290	0.004350
Min	0.000112	0.000076	0.001140
Max	0.002317	0.001599	0.023983

Table 8 presents the stability ratios.

The median ratio of 1.001 indicates that using one-month-old λ degrades reconstruction quality by only 0.1%, providing strong support for assumption (12).

Table 8: Nelson-Siegel Lambda Stability Ratios (Stale/Fresh)

Statistic	RMSE Ratio	MAE Ratio	TAE Ratio
Mean	1.018	1.019	1.019
Median	1.001	1.002	1.002
Std Dev	0.108	0.104	0.104
25th Percentile	1.000	1.000	1.000
75th Percentile	1.011	1.013	1.013

4.4 Comparative Performance

While both methods achieve low reconstruction errors individually, their relative performance varies across different yield curve environments. Table 9 presents the distribution of PCA-to-NS error ratios.

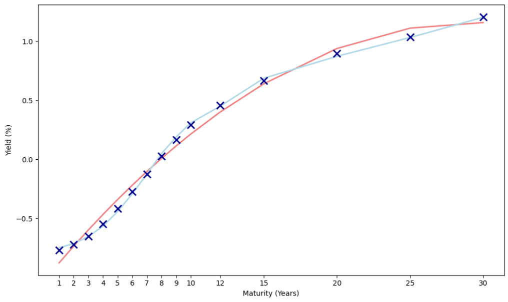
Table 9: PCA vs Nelson-Siegel Performance Ratios

Statistic	RMSE Ratio	MAE Ratio	TAE Ratio
Mean	1.078	1.039	1.039
Median	0.915	0.883	0.883
Std Dev	0.546	0.539	0.539
Min	0.281	0.265	0.265
Max	3.778	3.754	3.754
PCA Wins (%)		56.1%	

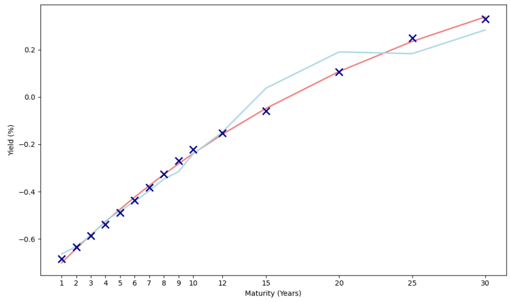
The median ratio of 0.915 indicates PCA typically outperforms NS by 8.5%, and PCA achieves lower errors in 56.1% of observations. However, the wide distribution (min 0.281, max 3.778) reveals substantial variation in relative performance.

Figure 3 illustrates four representative cases where the light blue line is the PCA fit, the light red line the NS fit and the crosses the actual yields. Panel (a) shows PCA's strongest performance relative to NS (RMSE ratio 0.281): the yield curve exhibits irregular short-end behavior that NS struggles to capture with its smooth parametric form. Panel (b) shows NS's strongest performance relative to PCA (RMSE ratio 3.778): the yield curve follows a textbook smooth shape perfectly suited to NS's functional form. Panels (c) and (d) show the worst individual fits for each method, revealing that even poor fits achieve acceptable accuracy (RMSE < 0.15).

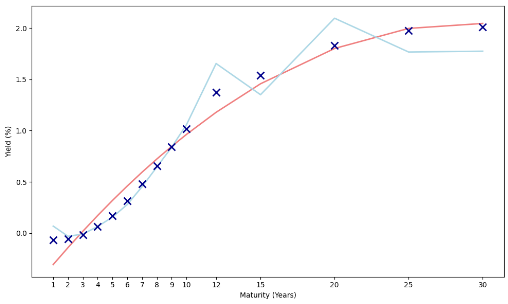
The comparative analysis reveals complementary strengths: NS excels when yield curves conform to smooth, well-behaved shapes characteristic of stable market conditions, while PCA's data-driven flexibility better accommodates irregular patterns and structural breaks.



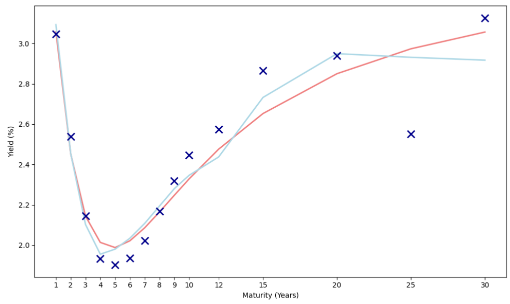
(a) Best PCA Performance Compared to NS



(b) Best NS Performance Compared to PCA



(c) Worst PCA Fit



(d) Worst NS Fit

Figure 3: Yield Curve Fits

5 Training ANN to forecast the Yield Curve

6 Conclusion

References

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.

Kondratyev, A. (2018). Learning curve dynamics with artificial neural networks. *SSRN Electronic Journal*, (3041232).

Litterman, R. and Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, 1(1):54–61.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of business (Chicago, Ill.)*, 60(4):473–489.

Suimon, Y., Sakaji, H., Izumi, K., and Matsushima, H. (2020). Autoencoder-based three-factor model for the yield curve of japanese government bonds and a trading strategy. *Journal of Risk and Financial Management*, 13(4).

Verner, R., Tkáč, M., and Tkáč, M. (2021). Improving quality of long-term bond price prediction using artificial neural networks. *Kvalita inovácia prosperita*, 25(1):103–123.