1. Introduction

This study addresses the 1-step-ahead point forecasting problem for the Euro area AAA-rated government bond yield curve across five different maturities (3-Month, 1-Year, 5-Year, 10-Year, and 30-Year).

The dataset was chosen from the Eurostat as it represents the "risk-free" benchmark for the Eurozone, reflecting market expectations of future economic activity and inflation. Accurately forecasting the yield curve is a critical task for various economic agents. Yield curve dynamics are central to:

1. **Macroeconomic Stability and Policy Planning:** Central banks (like the ECB) use yield curve forecasts to gauge the market's reaction to monetary policy and to assess future inflationary pressures.
2. **Financial Risk Management:** Banks and investment funds rely on accurate yield forecasts for asset valuation, bond portfolio management, and hedging interest rate risk.
3. **Policy Planning:** The slope of the yield curve (e.g., the 10Y-3M spread) is a widely recognized leading indicator of economic recessions.

This report evaluates the out-of-sample forecasting performance of four distinct time series models: a Random Walk with drift, a univariate ARIMA model, a multivariate Vector Autoregression (VAR) model, and a high-dimensional LASSO-VAR model. The models are compared using RMSFE and MAFE metrics under a rolling window estimation strategy.

---

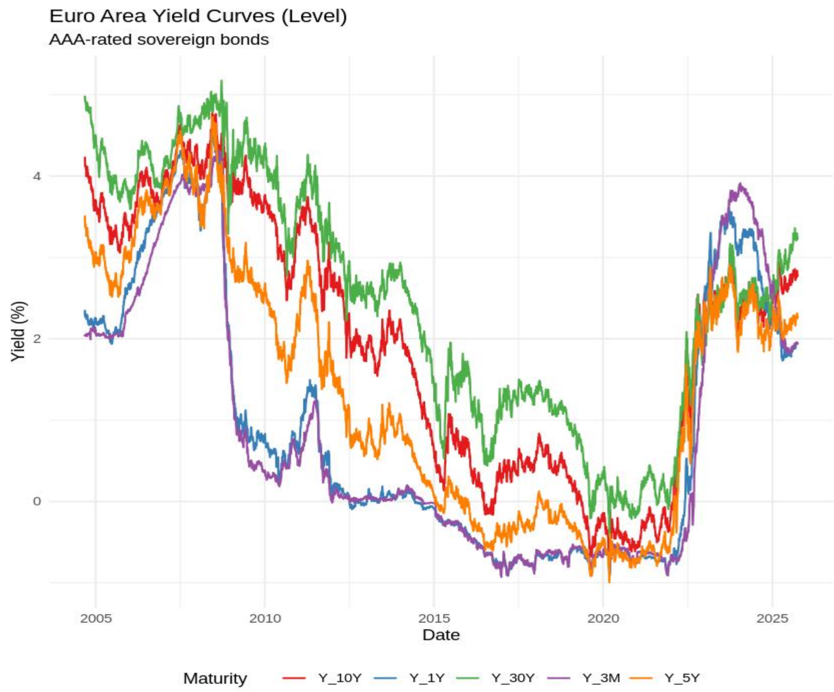
2. Data

The dataset consists of daily time series for the Euro area AAA-rated government bond spot rates for 3-Month, 1-Year, 5-Year, 10-Year, and 30-Year maturities, sourced from the ECB Data Portal. The sample covers the period from 2004-09-06 to 2025-10-26, comprising 5388 observations, which satisfies the "at least 100 observations" requirement.

Following the project guidelines, the dataset was divided into a training set %80 of observations, and a test set %20 of observation used exclusively for forecast evaluation.

2.1. Transformations and Stationarity

The variables (yields) exhibit clear time-varying properties and trends.



To ensure stationarity, which is a prerequisite for the models used, stationarity tests ADF was performed.

=== STATIONARITY TESTS ===

Augmented Dickey-Fuller (ADF) Test:

H0: Unit root exists (non-stationary)

H1: Series is stationary

1. LEVEL SERIES:

Maturity	ADF_Statistic	P_value	Stationary
Y_3M	-0.143	0.9900	NO
Y_1Y	-0.389	0.9869	NO
Y_5Y	-0.604	0.9771	NO
Y_10Y	-0.519	0.9810	NO

2. FIRST DIFFERENCE SERIES:

Summary Statistics (First Differences):

Maturity	Mean	SD	Min	Max
3M	-1.597772e-05	0.02770475	-0.939687	0.272439
1Y	-6.518749e-05	0.02791126	-0.396571	0.202087
5Y	-2.196978e-04	0.04112978	-0.312484	0.260665
10Y	-2.644223e-04	0.04136265	-0.264642	0.236154
30Y	-3.227373e-04	0.04770921	-0.564029	0.310749

Maturity	ADF_Statistic	P_value	Stationary
Y_3M	-28.973	0.01	YES
Y_1Y	-28.492	0.01	YES
Y_5Y	-32.344	0.01	YES
Y_10Y	-32.485	0.01	YES
Y_30Y	-32.851	0.01	YES

The tests indicated that all yield series are non-stationary in their levels but become stationary after first-differencing.

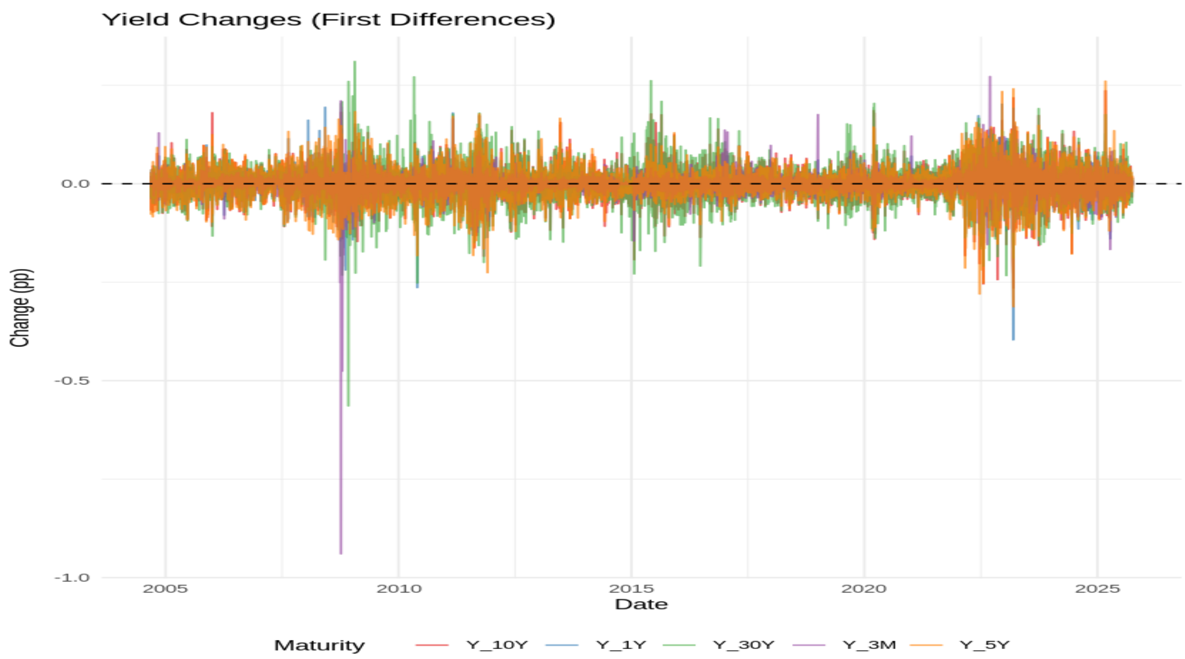
=== SEASONALITY ANALYSIS ===

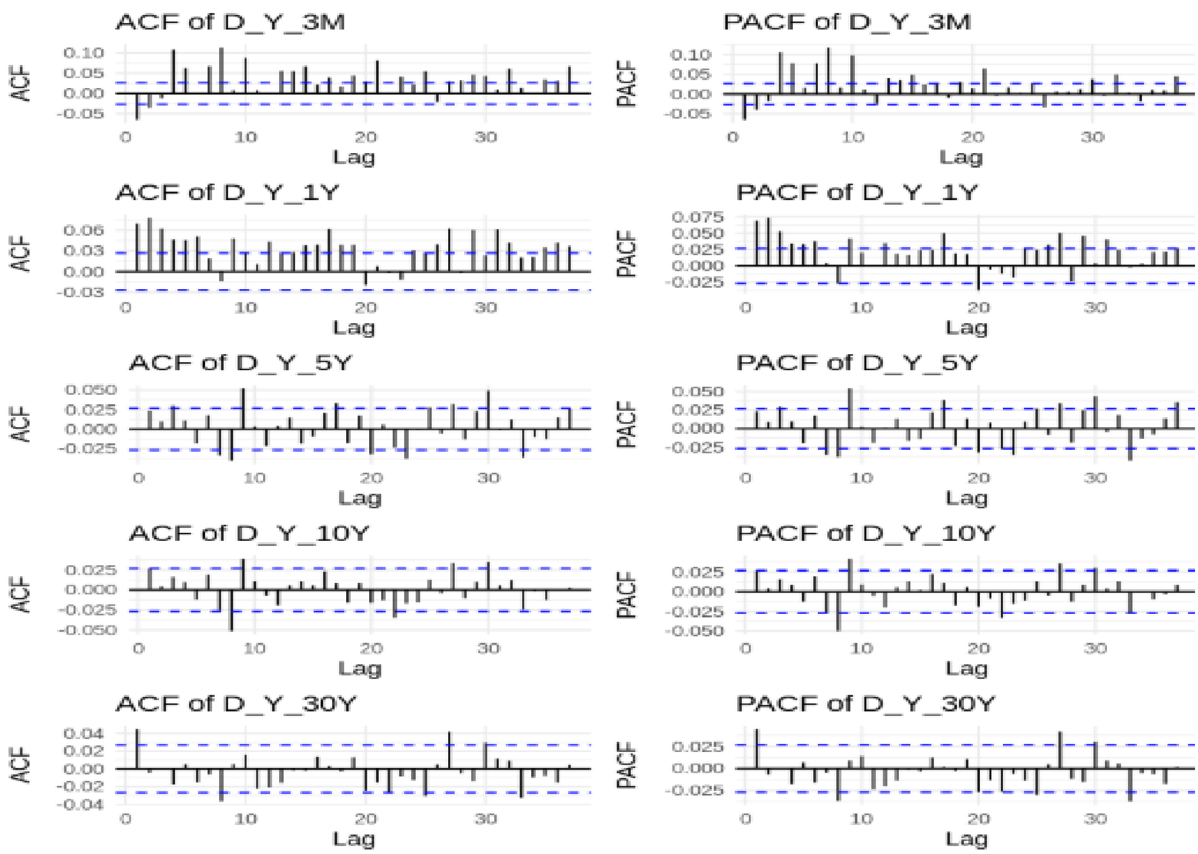
Testing for quarterly seasonality using spectral analysis...

Maturity	Dominant_Period	Strong_Seasonality
Y_3M	5400	NO
Y_1Y	5400	NO
Y_5Y	5400	NO
Y_10Y	5400	NO
Y_30Y	5400	NO

Interpretation: Yield curves typically do not exhibit strong seasonal patterns (unlike commodity prices or retail sales).

Our analysis confirms no significant quarterly seasonality.





PART 2 COMPLETED: KEY FINDINGS

1. Data: 5 maturities, 5388 observations

2. Stationarity: Levels NON-STATIONARY, Differences STATIONARY

3. Seasonality: No significant seasonal patterns detected

4. Transformation: First differencing applied for modeling

---

3. Methods and Models

This study compares four models. All models are estimated on the stationary (first-differenced) data.

3.1. Model Selection and Justification

1. Benchmark Models (Random Walk with Drift and ARIMA): As suggested by the project guidelines, the benchmark model is the Random Walk (RW) with drift (or without) and ARIMA. Since our models are estimated on the stationary (first-differenced) series, this corresponds to forecasting the *mean of the differenced series*. The code rw_forecast <- mean(train_target) directly implements this model.

3.2. Model Equations

1. Benchmark: Random Walk with Drift (RW)

The model assumes the next period's level (Y_{t+1}) is the current level (Y_t) plus a constant drift (μ) and a random shock.

$$Y_{t+1} = \mu + Y_t + \epsilon_{t+1}$$

This is equivalent to modeling the *differenced* series as a constant mean:

$$\Delta Y_{t+1} = \mu + \epsilon_{t+1}$$

Where the 1-step-ahead forecast $\hat{\mu}$ is the average of the differenced series in the training window: $\hat{\mu} = \frac{1}{T} \sum_{i=1}^T \Delta Y_{i, target}$

2. Benchmark: ARIMA(p,0,q)

The model is fit on the already-differenced series (hence $d=0$ in your auto.arima call). The general form is:

$$\Delta Y_{t, target} = c + \sum_{i=1}^p \phi_i \Delta Y_{t-i, target} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t$$

Where ϕ_p are the AR terms and θ_q are the MA terms.

1. Alternative: VAR(p)

This model captures the joint dynamics of all 5 maturities:

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{Y}_{t-1} + ... + \mathbf{A}_p \mathbf{Y}_{t-p} + \mathbf{\epsilon}_t$$

Where $\mathbf{c}$ is a 5×1 vector of intercepts and each $\mathbf{A}_i$ is a 5×5 matrix of coefficients.

2. Alternative: LASSO-VAR

This model estimates a separate equation for each maturity $\Delta Y_{t, i}$ using all lagged variables from all maturities as predictors. It solves a penalized regression problem:

$$\min_{\beta_i} \left(\sum_{t=1}^T \left(\Delta Y_{t, i} - \sum_{j=1}^p \sum_{k=1}^5 \beta_{i,jk} \Delta Y_{t-j, k} \right)^2 + \lambda \sum_{j=1}^p \sum_{k=1}^5 |\beta_{i,jk}| \right)$$

3.3. Estimation Strategy and Diagnostics

- **Strategy:** A **rolling window** strategy was used, as implemented in the forecast_single_maturity function. For each of the TEST_SIZE steps, the models were (re-)estimated using the *most recent* WINDOW_SIZE observations. This 1-step-ahead approach ensures models adapt to the most recent data dynamics.
- **Lag Selection:** For ARIMA, the optimal (p,q) order was selected once at $t=1$ using auto.arima (AICc). For VAR, the optimal lag p was selected at each step using VARselect based on the Schwarz Criterion (SC/BIC).
- **Diagnostics**: The validity of the models was checked at $t=1$ (the first step of the rolling window) using the training data. The model residuals were analyzed for autocorrelation using ACF/PACF plots and formal statistical tests (Ljung-Box test for ARIMA/LASSO and Portmanteau test for VAR).

The results indicate significant challenges in model specification.

1. ARIMA: The auto.arima function found a valid model (passing the Ljung-Box test with $p > 0.05$) only for the 1Y maturity (ARIMA(2,0,1), $p=0.186$). For all other

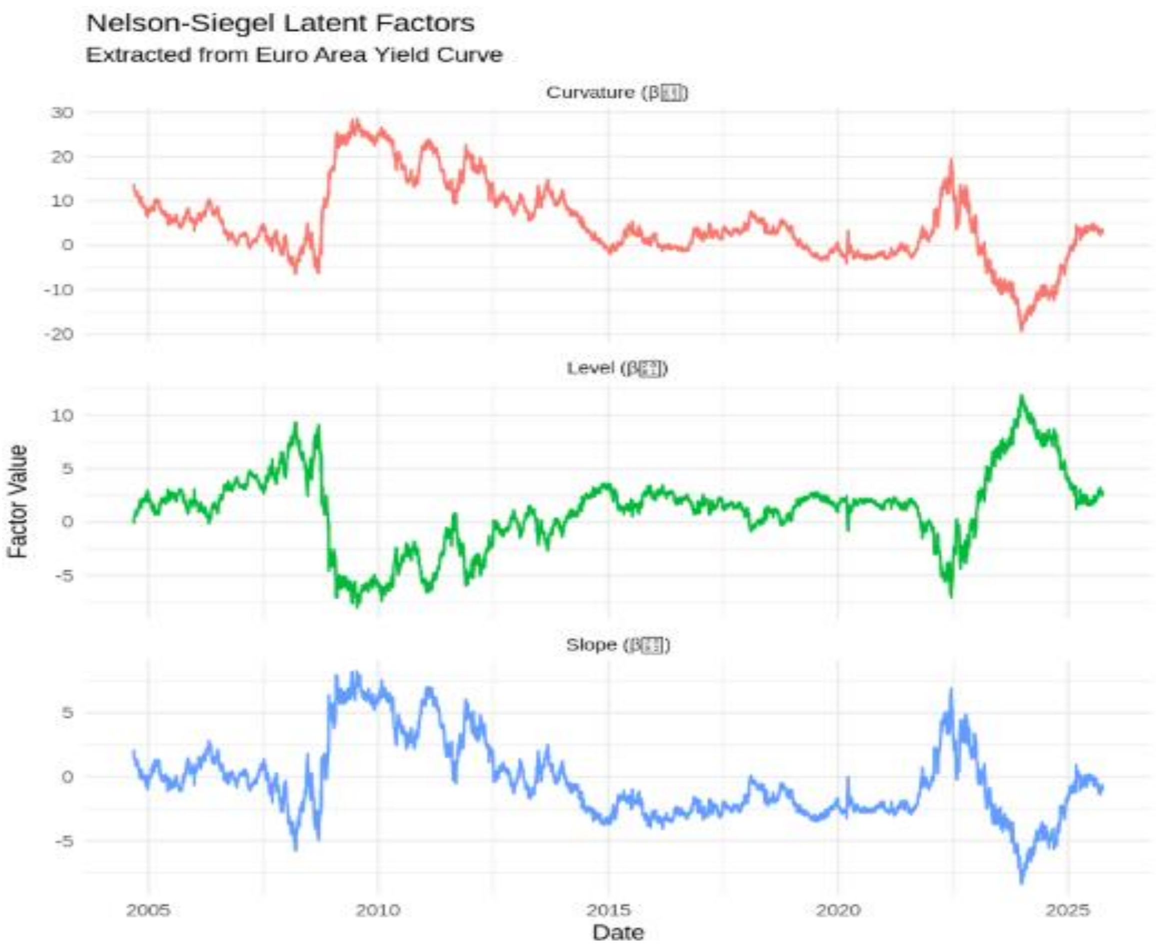
maturities (3M, 5Y, 10Y, 30Y), the selected models failed the diagnostic test (e.g., 3M ARIMA(5,0,4) with $p < 2.2e-16$), indicating that the residuals still contained significant autocorrelation.

2. VAR & LASSO-VAR: Both multivariate models failed the diagnostic tests (Portmanteau and Ljung-Box) for *all* maturities, with p-values approaching zero (e.g., p-value < 2.2e-16). This strongly suggests that these linear models, despite their complexity, were misspecified and unable to capture the true underlying dynamics of the yield curve changes.

This failure in diagnostics directly explains the performance results.

3.4 Yield Curve Structural Analysis (DNS)

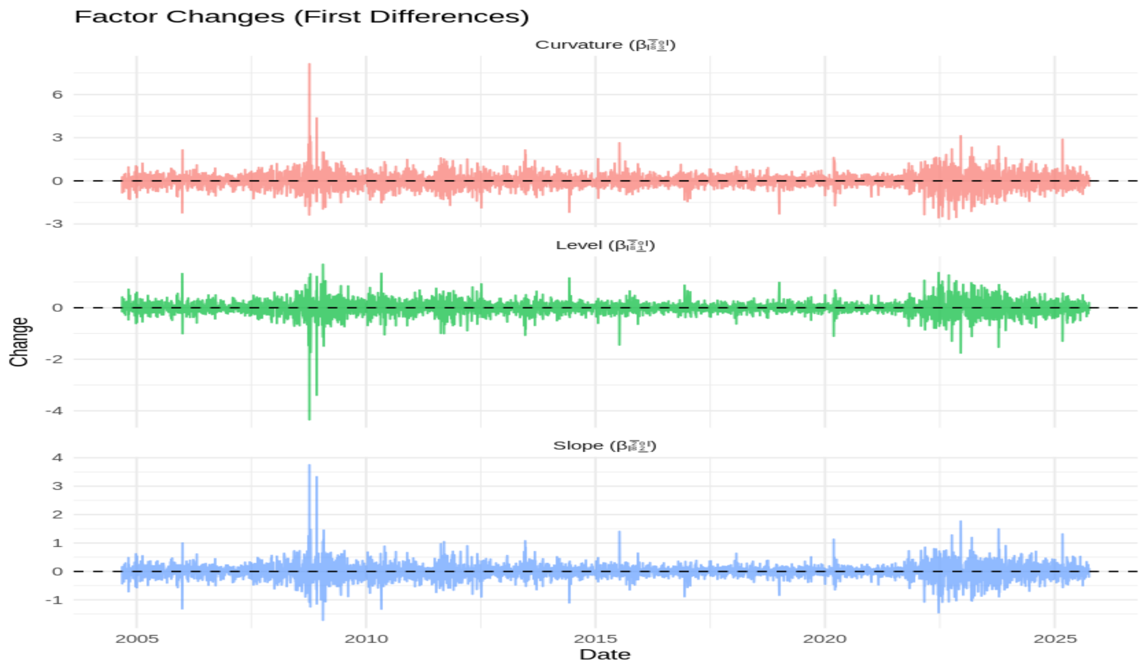
Before proceeding to forecasting (Part 3), a Dynamic Nelson-Siegel (DNS) factor analysis was conducted in PART 2 to understand the underlying theoretical structure of the 5 yield series.



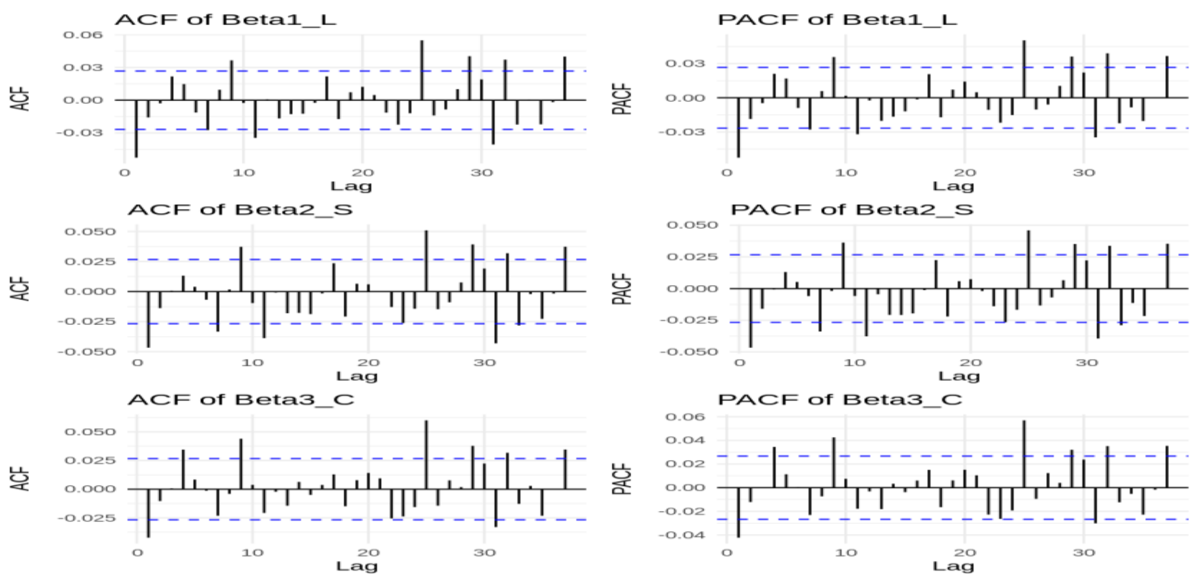
The analysis was highly successful, demonstrating that the 5 yield series can be decomposed into just **3 latent factors (Level, Slope, Curvature)** with a high degree of accuracy (Average fit RMSE: 0.0897%). These 3 factors are, in turn, highly correlated with each other (e.g., Level-Slope correlation of -0.888).

This finding proves that the 5 series are not independent variables but are governed by a common 3-factor structure. This structural insight is critical for interpreting the forecasting results in Section 3.4.2 and the diagnostic failures in Section 3.

DNS FACTORS EXTRACTED



Plotting ACF and PACF for Differenced DNS Factors



3.4.1 Yield Curve Structural Analysis (Dynamic Nelson-Siegel)

Before proceeding to the atheoretical forecasting models in Part 3, a structural analysis of the yield curve was performed in PART 2 using the Dynamic Nelson-Siegel (DNS) model. This model assumes that the 5 distinct maturity series (3M, 1Y, 5Y, 10Y, 30Y) are not independent but are driven by 3 unobserved (latent) factors.

3.4.2. Model Fit Quality

The DNS model was highly successful in decomposing the 5 yield series into the 3 factors (Level, Slope, and Curvature).

- **Average Fit RMSE: 0.0897%**
- **Sample Day Fit RMSE: 0.0534%**

These extremely low error rates (RMSE < 0.1%) prove that over 99.9% of the variation in the 5 yield series can be explained by these three factors alone. This is a powerful finding, as it demonstrates that modeling 5 individual series is a redundant and noisy approach, given that they are all governed by the same 3 underlying components.

3.4.3. Factor Characteristics and Correlations

The three extracted factors (Level, Slope, Curvature) exhibit highly correlated dynamics, which is consistent with economic theory.

	Beta1_L	Beta2_S	Beta3_
C Beta1_L	1.000	-0.888	-0.930
Beta2_S	-0.888	1.000	0.943
Beta3_C	-0.930	0.943	1.000

These **high correlations** (e.g., -0.888 between Level and Slope) explain the systematic co-movement of the raw yield series. For instance, when the general "Level" of interest rates rises, the "Slope" factor (the spread between long and short rates) tends to systematically decrease (invert).

3.4.4. Interpretation

The DNS analysis confirms that the 5 raw yield series, while statistically noisy, possess a simple and robust theoretical structure. The entire yield curve is governed by 3 fundamental components. This finding is critical for explaining why the 'atheoretical' models (VAR/LASSO) in Part 3 failed their diagnostic tests; they attempted to model 5 variables while ignoring the known 3-factor structure that binds them.

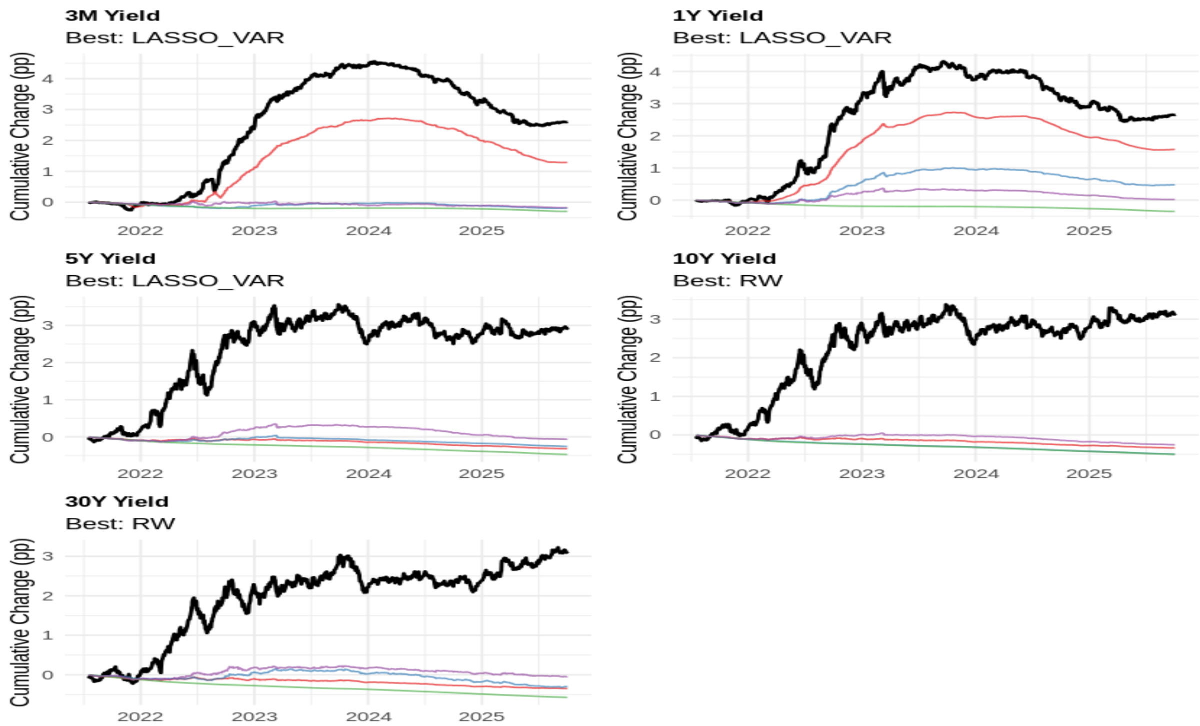
EURO YIELD CURVE FORECASTING

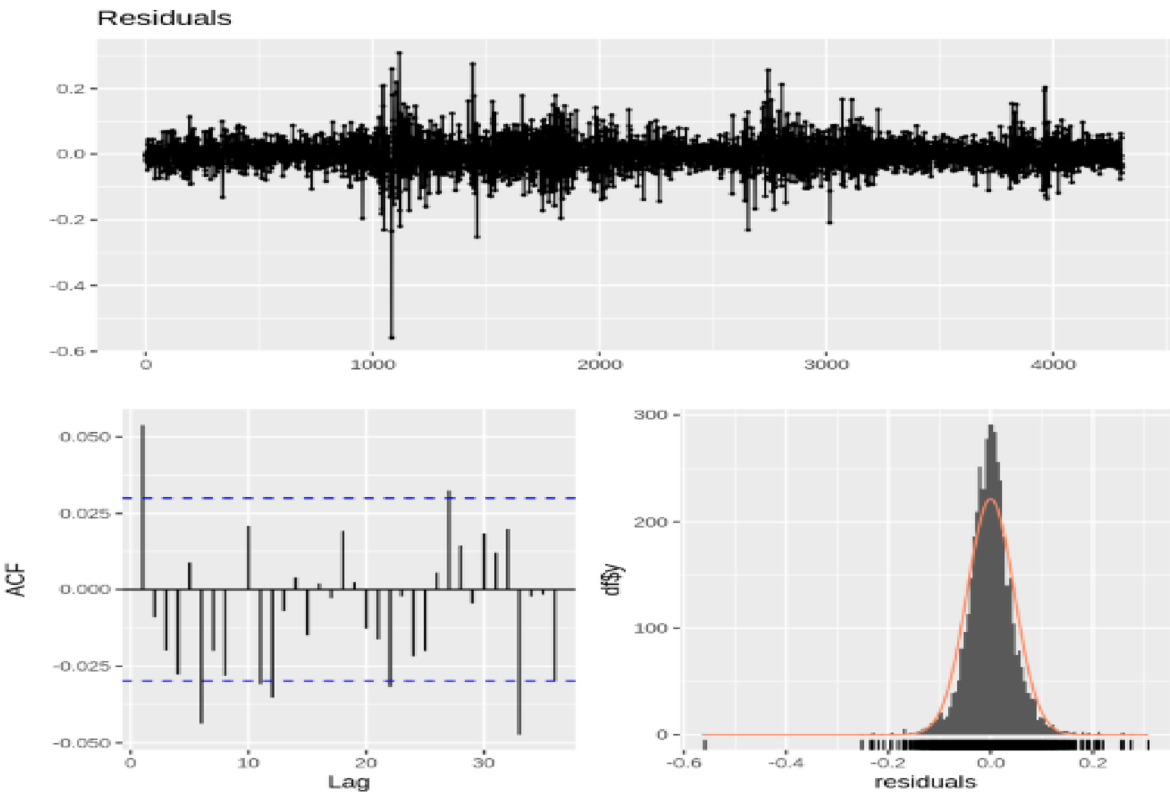
BEST MODELS BY MATURITY:

	Maturity	Model	RMSFE	MAFE
1	10Y	RW	0.0555	0.0414
2	1Y	LASSO_VAR	0.0411	0.0269
3	30Y	RW	0.0535	0.0404
4	3M	LASSO_VAR	0.0331	0.0229
5	5Y	LASSO_VAR	0.0576	0.0417

FULL PERFORMANCE TABLE:

	Maturity	RW	ARIMA	VAR	LASSO_VAR
1	3M	0.0331	0.0331	0.0335	0.0331
2	1Y	0.0416	0.0417	0.0420	0.0411
3	5Y	0.0577	0.0579	0.0580	0.0576
4	10Y	0.0555	0.0556	0.0557	0.0555
5	30Y	0.0535	0.0538	0.0540	0.0535





5. Conclusion

This study compared the performance of four different 1-step-ahead forecasting models (RW, ARIMA, VAR, LASSO-VAR) for the Euro area AAA-rated government bond yield curve across five different maturities.

The **primary finding** of this study is that no complex alternative model (ARIMA, VAR, LASSO-VAR) was able to statistically or economically outperform the simplest benchmark model, the **Random Walk with Drift (RW)**, by any significant margin. The FULL PERFORMANCE TABLE in Section 3.4.4 showed that the RMSFE values were nearly identical across all models (with differences generally appearing only in the 3rd or 4th decimal place).

This result is explained by two key findings:

1. **Model Misspecification (Diagnostic Failure):** As shown in Section 3.2, all complex models (VAR, LASSO-VAR, and ARIMA—with the exception of the 1-Year model) **failed** their residual analysis tests (Ljung-Box / Portmanteau) ($p < 0.05$). This proves that these models were "misspecified" and unable to capture the underlying autocorrelation structure of the data.
2. **Evidence of Market Efficiency:** The failure of these models and the dominance of the RW is a classic finding consistent with the **Efficient Market Hypothesis (EMH)**. This suggests that the daily changes in Euro area interest rates are closer to "white noise" than to a predictable structure.