Advanced Programming 2025

Yield Curve Forecasting with Machine Learning

A Comparative Study of Baseline, Random Forest, and XGBoost Models

Final Project Report

Anna Modolo
anna.modolo@unil.ch
Student ID: 25434150

January 4, 2026

Abstract

This report studies short horizon forecasting of euro area government bond yields using a benchmark and two machine learning models. Monthly end-of-period zero coupon spot rates at 2-, 5-, and 10-year maturities are taken from the European Central Bank and span multiple interest rate regimes. Forecasts are produced in a rolling window setting that preserves the time ordering of the data and avoids bias.

We compare a naïve persistence benchmark with two tree-based regressors: Random Forest and XGBoost. Performance is assessed with standard regression metrics (RMSE and MAE) and, crucially, with predicted versus actual plots to diagnose turning point behavior and regime sensitivity.

The results show incremental gains from machine learning, especially at the 5- and 10-year maturities where yield dynamics are smoother and cross maturity information is more explanatory. Random Forest produces stable, smoothed forecasts but tends to lag during rapid tightening phases. XGBoost is more responsive to emerging trends. All models struggle during abrupt policy driven regime shifts, highlighting the limits of purely backward-looking approaches.

Keywords: yield curve, bond yields, forecasting, machine learning, Random Forest, XGBoost, ECB

Contents

1 Introduction **3**

2 Literature Review **3**

3 Methodology **3**

3.1 Data Description . . . 3

3.2 Forecasting Framework and Models . . . 4

3.3 Evaluation Strategy . . . 5

4 Codebase and Reproducibility **5**

5 Results and Model Comparison **5**

5.1 Predicted vs Actual Paths . . . 8

5.1.1 2-year maturity . . . 8

5.1.2 5-year maturity . . . 8

5.1.3 10-year maturity . . . 9

6 Discussion **9**

7 Limitations **9**

8 Conclusion and Future Work **10**

References **11**

A AI Tools Used **12**

B Reproducibility Notes **12**

1 Introduction

Forecasting government bond yields is a central problem in financial economics, given the crucial role of the yield curve in monetary policy transmission, asset allocation, risk management, and the valuation of interest rate sensitive securities. Yield curves summarize market expectations about future interest rates, inflation, and macroeconomic conditions, which makes them an essential tool for policy makers, investors, and financial institutions. However, accurate yield forecasting remains challenging due to the strong persistence of interest rates, nonlinear dynamics, and abrupt regime shifts driven by changes in monetary policies and economic conditions. Traditional econometric approaches to yield forecasting rely on explicit structural assumptions, such as linear relationships or stable factor structures. While these models provide interpretability and economic intuition, they may struggle to capture complex nonlinear interactions across maturities or adapt quickly to evolving policy regimes. Machine learning methods, by contrast, offer a better flexibility by relaxing many of these assumptions, improving predictive performance, althought often at the cost of reducing transparency and interpretability. This project investigates the short horizon forecasting performance of three distinct models applied to euro area government bond yields: a naïve baseline model, a Random Forest regressor, and an XGBoost regressor. The baseline model serves as a benchmark reflecting the strong persistence observed in interest rate series, while the machine learning models are designed to capture nonlinear patterns and cross-maturity interactions that simpler approaches may overlook. The empirical analysis uses monthly end-of-period zero coupon spot rates at 2-year, 5-year, and 10year maturities from the European Central Bank Statistical Data Warehouse, representing short-, medium-, and long-term segments of the yield curve. Forecasts are generated within a rollingwindow framework that respects the temporal structure of the data and avoids bias. Model performance is evaluated using standard regression metrics and visual diagnostics, allowing for a detailed assessment of how each approach behaves across different maturities and market regimes. Overall, the analysis highlights the trade offs involved in applying machine learning techniques to macro financial time series: while flexible models can improve forecast accuracy in stable environments, they may struggle to adjust rapidly to sudden policy driven shifts, underscoring the importance of combining data-driven methods with economic insight.

2 Literature Review

Yield curve forecasting has a long tradition in both finance and macroeconomics. Classic empirical work documents the strong persistence of interest rates and the difficulty of outperforming simple benchmarks at short horizons. Term structure models and factor based representations (e.g., level, slope, curvature) provide economically interpretable summaries of yield movements, while macro finance approaches connect yields to inflation and monetary policy expectations. More recently, machine learning methods have been applied to the term structure to capture nonlinearities and interactions that may be missed by linear specifications.

In this project, the goal is not to propose a new forecasting method, but to assess how a simple benchmark compares to two widely used tree based machine learning approaches when applied to euro area yields across maturities and regimes.

3 Methodology

3.1 Data Description

As anticipated before, the dataset consists of euro area government bond spot rates at 2-year, 5-year, and 10-year maturities, sourced from the European Central Bank. The original data are daily, but for the purpose of this analysis they are aggregated to monthly end-of-period observations in order to reduce noise and align the analysis with medium-term macro financial dynamics. The sample period begins in 2004 and spans multiple, clearly distinct interest rate regimes. It includes the preGlobal Financial Crisis period, characterized by relatively stable growth and conventional monetary policy; the prolonged low interest rate and negative rate environment that followed the crisis, marked by unconventional monetary policy measures such as quantitative easing; and the sharp post 2021 tightening cycle, during which inflationary pressures prompted an abrupt reversal of accommodative policies. These regime shifts introduce substantial variation in yield dynamics, volatility, and persistence.

This diversity makes the dataset particularly suitable for evaluating model robustness across regimes.

3.2 Forecasting Framework and Models

Forecasting Framework

All models are evaluated using a rolling window forecasting strategy, ensuring that only past information is used at each prediction step. This approach mirrors realistic forecasting conditions and avoids look ahead bias.

Baseline Model

The baseline model generates forecasts by setting the predicted yield at time t+1 equal to the most recently observed yield at time t. Despite its simplicity, this naïve approach represents a strong and widely used benchmark in the context of interest rate forecasting, as government bond yields typically exhibit high persistence and slow moving dynamics. In practice, such a model closely resembles a random walk without drift and often proves difficult to outperform at short horizons. The primary role of the baseline model in this project isnt to provide clear predictions, but to establish a reference point from which we will evaluate more complex methods. Any model that fails to outperform this benchmark may offer limited practical value. However, while the baseline performs reasonably well during stable regimes, it is unable to anticipate turning points, structural breaks, or changes in volatility, as it relies entirely on the most recent observation.

Random Forest

The Random Forest model is an ensemble learning method that builds a large number of decision trees using different random subsets of the training data and averages their predictions. Using a random subset of features at each split reduces correlation across trees and improves generalization performance. This structure allows the model to capture nonlinear relationships and complex interactions between predictors without imposing strong parametric assumptions. In the context of yield forecasting, Random Forest is particularly well suited to exploit crossmaturity information and lagged yield dynamics. By averaging across many trees, the model tends to produce smooth forecasts that filter out short-term noise while adapting to local patterns in the data. This smoothing property can be advantageous during stable macroeconomic environments, where yield dynamics evolve gradually over time. However this can also limit the models ability to respond quickly to abrupt changes in the underlying data generating process. As a result, Random Forest models may lag during periods of rapid regime shifts, such as sudden monetary policy changes, and may underestimate some sharp yield movements.

XGBoost

XGBoost is a gradient boosting framework that builds decision trees sequentially, with each new tree trained to correct the prediction errors of the previous ensemble. Unlike Random Forest, which relies on averaging independent trees, XGBoost focuses on reducing residual errors, which allows it to capture more complex patterns in the data. XGBoost is generally more flexible than Random Forest and can adapt more rapidly to changes in the dynamics of yields. This makes it better in order to capture sharper transitions or nonlinear responses to macroeconomic shocks. In yield curve forecasting, this flexibility can translate into improved performance during periods of changing monetary policy expectations. At the same time, since the model is very sensitive, it may react strongly to short-lived fluctuations, particularly in small samples or during volatile periods. Moreover, its complexity reduces interpretability relative to simpler benchmarks and even to Random Forest models. Consequently, while XGBoost may offer superior predictive power in some regimes, its performance can be less stable across different market environments.

3.3 Evaluation Strategy

Model performance is evaluated using standard regression metrics, including the root mean squared error (RMSE) and mean absolute error (MAE), which are summarized in a comparative results table. While these quantitative measures provide a useful assessment of average forecast accuracy, they offer only a partial view of model performance in the presence of strongly autocorrelated time series such as government bond yields. For this reason, visual inspection of predicted versus actual yield paths plays a central role in the analysis. Time-series plots allow for a more nuanced evaluation of how each model tracks yield dynamics over time, captures turning points, and responds to changes in volatility and market regimes. By combining numerical metrics with graphical analysis, the evaluation framework provides a more comprehensive understanding of the strengths and limitations of each forecasting approach.

4 Codebase and Reproducibility

All experiments were implemented in pyhton using standard scientific libraries (NumPy, pandas, scikitlearn, XGBoost, matplotlib). The code is organized in a modular structure and can be executed by running the main script, which reproduces all figures and evaluation results. Fixed random seeds are used where applicable to ensure reproducibility.

5 Results and Model Comparison

I. Two-Year Yield

Baseline Model

The baseline model tracks the actual 2-year yield closely during relatively stable periods, reflecting the strong persistence and near random-walk behavior that characterize short-term interest rates. In quiet regimes, this simple specification performs reasonably well, as changes in short-term yields tend to be gradual and largely driven by incremental adjustments in policy expectations. However, the baseline model reacts slowly to turning points, as it relies entirely on past information and lacks any mechanism to anticipate changes in the direction or intensity of yield movements. This limitation becomes particularly evident during the sharp upward shift associated with the ECB monetary tightening cycle. In this phase, the baseline systematically underestimates both the speed and the magnitude of the increase in the 2-year yield, failing to adjust promptly to rapidly evolving policy expectations. As a result, forecast errors widen precisely during periods of heightened policy uncertainty, highlighting the inability of purely persistence-based models to cope with abrupt regime changes in short-term interest rates.

Random Forest

The Random Forest model generates noticeably smoother predictions than the baseline, effectively filtering out short-term fluctuations and reducing high-frequency noise in the yield series. During relatively calm market conditions, this smoothing behavior enhances forecast stability and leads to more regular prediction paths, which can be advantageous in environments characterized by gradual adjustments in interest rates. By averaging across a large number of decision trees, the model captures persistent patterns in the data while dampening idiosyncratic movements. However, this same smoothing mechanism becomes a drawback during periods of abrupt regime change. In the post-2021 monetary tightening phase, the Random Forest model exhibits a systematic lag relative to the actual yield, adjusting only gradually to the rapid upward shift in interest rates. This delayed response indicates a limited ability to extrapolate sudden policy-driven changes beyond the historical patterns observed in the training window. As a result, forecast errors increase precisely during periods when timely adaptation is most critical, highlighting the trade-off between stability and responsiveness inherent in ensemble-based smoothing methods.

XGBoost

XGBoost exhibits a higher degree of responsiveness than the Random Forest model, allowing it to adjust more rapidly to changes in the direction and intensity of yield movements. By sequentially correcting previous prediction errors, the boosting framework enables the model to capture portions of the sharp upward shift in yields earlier than the other approaches, particularly during the initial phase of the post-2021 tightening cycle. This increased adaptability allows XGBoost to better track evolving trends when market conditions begin to change. However, the greater flexibility of XGBoost also introduces higher short-term variability in its forecasts. In periods of heightened volatility, the model occasionally overshoots or undershoots the actual yield path, reflecting sensitivity to transient fluctuations in the training data. These deviations suggest that while XGBoost is more reactive to emerging signals, it may also amplify noise when structural changes are rapid and persistent. Consequently, the model faces a trade-off between improved responsiveness and forecast stability, performing well in capturing directional shifts but at the cost of less smooth prediction paths relative to Random Forest.

Comparison

For the 2-year maturity, all models encounter significant difficulties in responding to sudden, policy-driven changes in interest rates, reflecting the strong influence of central bank decisions on short-term yields. Among the three approaches, XGBoost emerges as the most reactive model, adjusting more quickly to shifts in the direction of yields and partially capturing the onset of abrupt tightening phases. However, this increased responsiveness is accompanied by higher short-term volatility in forecasts. In contrast, the Random Forest model produces the smoothest prediction paths, prioritizing stability and noise reduction over rapid adjustment. While this behavior leads to reliable performance during stable periods, it results in pronounced lag during episodes of rapid policy normalization. The baseline model, despite its simplicity, remains surprisingly competitive in tranquil environments, where short-term yields exhibit strong persistence and limited directional change. This comparison highlights that, at short maturities, the forecasting challenge is dominated by policy uncertainty rather than model complexity, limiting the scope for substantial gains from more sophisticated methods.

II. Five-Year Yield

Baseline Model

At the 5-year maturity, the baseline model performs noticeably better than in the short-term case, reflecting the reduced influence of immediate policy announcements and the more gradual evolution of medium-term yields. The model is able to track broad trends reasonably well during extended periods of stability, as medium term rates tend to adjust more smoothly to changes in macroeconomic expectations. Nevertheless, the baseline continues to detect turning points with delay, particularly during phases of rapid adjustment, indicating that persistence alone remains insufficient to capture shifts in the slope and level of the yield curve.

Random Forest

The Random Forest model exhibits a clear improvement in performance at the 5-year maturity. Predictions closely follow the actual yield path during both declining and rising phases, with substantially reduced lag relative to the 2-year case. This improvement suggests that medium-term yields contain more stable and exploitable patterns, which tree-based models are better able to learn from historical data. The smoothing properties of Random Forest remain evident, leading to stable and well-behaved forecasts that successfully filter out short-term noise without excessively delaying adjustment to evolving trends.

However, during periods of sharp acceleration, such as the initial phase of the post-2021 tightening cycle, some lag persists. This indicates that while Random Forest adapts more effectively at medium maturities, it still prioritizes stability over rapid responsiveness.

XGBoost

XGBoost provides the closest visual fit among the three models for the 5-year yield. It successfully captures both the prolonged downward trend observed during the negative-rate environment and the subsequent sharp upward adjustment during monetary tightening. Compared to Random Forest, XGBoost reacts more quickly to changes in yield direction, allowing it to track emerging trends with greater precision. Nonetheless, during the steepest segments of the tightening phase, XGBoost still exhibits some delay and mild overshooting, reflecting the inherent difficulty of forecasting rapid structural changes using historical patterns alone. Overall, the model achieves a favorable balance between responsiveness and accuracy at this maturity.

Comparison

At the 5-year maturity, the advantage of machine learning models over the baseline becomes more pronounced. XGBoost generally outperforms Random Forest in terms of responsiveness, particularly during transitional phases, while Random Forest delivers smoother and more stable forecasts. The baseline model is less competitive than at shorter maturities, as persistence alone fails to capture the richer dynamics present in medium-term yields. These results suggest that machine learning methods are especially well suited to forecasting yields in this segment of the curve, where expectations evolve more gradually and are less dominated by immediate policy actions.

III.Ten-Year Yield

Baseline Model

For long-term yields, the baseline model performs remarkably well, reflecting the strong inertia and persistence characteristic of long-maturity rates. The model effectively tracks long-run trends and exhibits relatively small forecast errors during stable periods. However, it consistently underestimates the magnitude of large upward shifts, particularly during episodes of rapidly changing inflation expectations or long-term policy outlooks. This limitation highlights the inability of purely backward-looking models to capture shifts in long-term risk premia.

Random Forest

The Random Forest model closely tracks the dynamics of the 10-year yield, striking a balance between smoothness and adaptability. Deviations from the actual series are generally small, and the model captures the overall shape and direction of long-term yield movements effectively. Compared to shorter maturities, the smoothing behavior of Random Forest is less problematic here, as longterm yields adjust more gradually and are less sensitive to high-frequency shocks. As a result, Random Forest delivers robust and stable forecasts at this maturity, with limited lag even during tightening phases.

XGBoost

XGBoost again produces the most reactive forecasts for the 10-year yield, closely following the upward trend observed during periods of monetary tightening. The model adapts more rapidly than both the baseline and Random Forest, allowing it to capture changes in long-term yield expectations with minimal delay. While minor overshooting is occasionally observed, these deviations are relatively limited and do not materially detract from overall forecast accuracy. The strong performance of XGBoost at this maturity suggests that its flexibility is particularly well suited to modeling long-term yield dynamics, where structural changes tend to unfold more gradually.

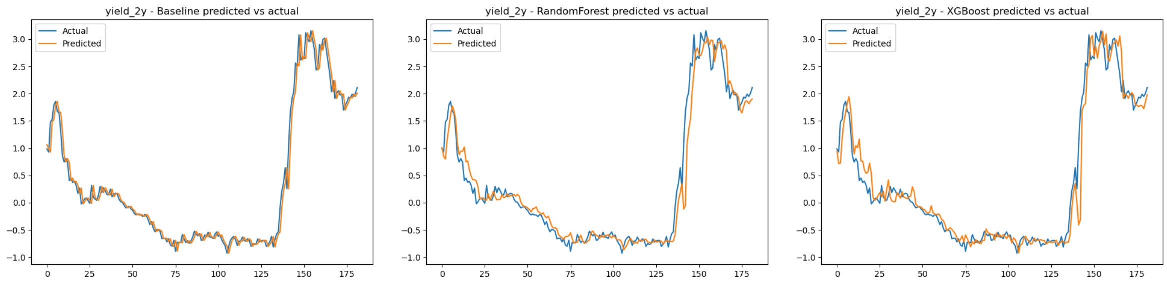
Comparison

At the 10-year maturity, differences between models narrow considerably. All approaches perform better than at shorter maturities, highlighting the increased predictability of long-term yields. While XGBoost remains the most responsive and Random Forest the most stable, the baseline model also performs competitively due to the high persistence of long-term rates. These findings indicate that, at long horizons, the forecasting challenge is less dominated by abrupt policy shifts and more driven by slow-moving macroeconomic fundamentals, reducing the relative advantage of highly flexible models.

5.1 Predicted vs Actual Paths

The following figures summarize the qualitative behavior of each model across maturities (baseline, Random Forest, and XGBoost). They complement the numerical metrics by highlighting turning-point behavior, lag during regime shifts, and short-term volatility.

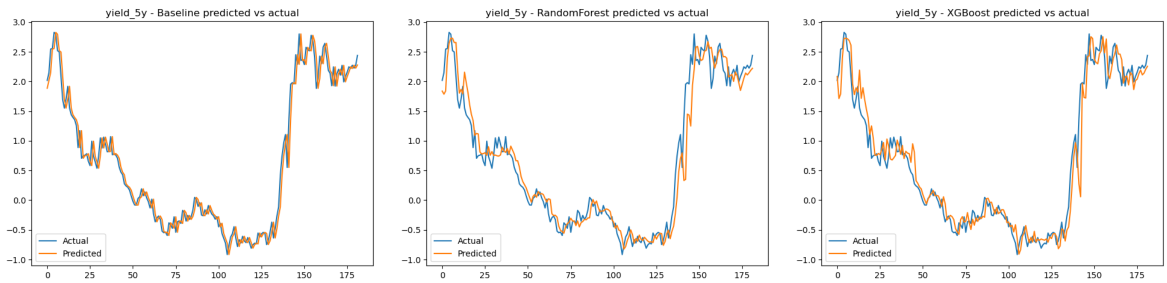
5.1.1 2-year maturity



(a) Baseline (b) Random Forest (c) XGBoost

Figure 1: Predicted vs actual 2-year yield.

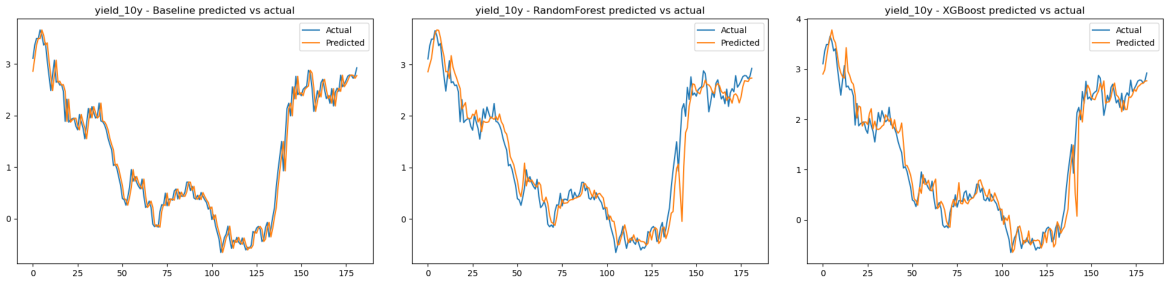
5.1.2 5-year maturity



(a) Baseline (b) Random Forest (c) XGBoost

Figure 2: Predicted vs actual 5-year yield.

5.1.3 10-year maturity



(a) Baseline (b) Random Forest (c) XGBoost

Figure 3: Predicted vs actual 10-year yield.

6 Discussion

Across all maturities, a clear and consistent pattern emerges from the empirical analysis. Machine learning models tend to outperform the naïve baseline primarily at medium- and long-term maturities, where yield dynamics evolve more smoothly and are less directly anchored to immediate monetary policy decisions. In these segments of the yield curve, expectations about long-run inflation, growth, and risk premia play a more prominent role, creating an environment in which nonlinear models can better exploit historical patterns and cross-maturity information. Nevertheless, the results also underscore an important limitation shared by all approaches. None of the models is able to fully anticipate abrupt structural breaks, particularly those associated with

sudden shifts in monetary policy regimes. During such episodes, forecast errors increase markedly, highlighting the intrinsic difficulty of predicting policy-driven markets using purely historical yield information. This finding reinforces the notion that even flexible machine learning models remain fundamentally backward-looking and may lag when confronted with rapid changes in the policy environment. Differences in model behavior further illustrate the trade-offs involved in yield forecasting. Random Forest consistently prioritizes smoothness and stability, producing regular and low-variance prediction paths that perform well in environments characterized by gradual adjustments. This makes it particularly suitable during extended periods of macroeconomic stability, albeit at the cost of slower adaptation to regime shifts. XGBoost, by contrast, offers greater responsiveness to emerging trends, allowing it to react more quickly to changes in yield dynamics. However, this increased flexibility may introduce higher short-term volatility and occasional overshooting. The baseline model, despite its simplicity, remains a strong benchmark, especially during tranquil regimes where yield persistence dominates. Together, these results suggest that forecasting performance reflects a balance between stability and adaptability, rather than sheer model complexity alone.

7 Limitations

This study relies only on historical yield information and intentionally excludes macroeconomic variables and explicit policy indicators. While this choice allows the analysis to focus on the information contained in the yield curve, it also creates important limitations. In particular, the models are unable to anticipate exogenous shocks, such as unexpected macroeconomic releases, geopolitical events, or central bank policy announcements, which can lead to abrupt and significant yield adjustments. As a consequence, forecast performance deteriorates during periods when new information arrives that is not reflected in past yield dynamics. Moreover, the use of monthly end-of-period aggregation, although effective in reducing highfrequency noise and enhancing model stability, inevitably smooths out intra-month dynamics and short-lived fluctuations. Important information contained in the timing and sequencing of policy communications or market reactions within the month may therefore be lost. This aggregation choice may reduce the models ability to detect early signs of regime change and helps explain the forecast lag observed during rapid tightening or easing phases. Together, these limitations highlight the trade-off between noise reduction and informational richness, and suggest that incorporating higher-frequency data or macroeconomic covariates could improve model responsiveness and robustness in future extensions.

8 Conclusion and Future Work

This project shows that machine learning models can improve yield forecasts compared to simple benchmarks, especially at medium and long term maturities where yield movements are smoother and less directly influenced by short-term monetary policy. In these segments of the yield curve, flexible models such as Random Forest and XGBoost are better able to exploit nonlinear patterns and cross-maturity relationships embedded in historical data. However, the empirical gains achieved by these methods are incremental rather than transformative, and simple persistence-based benchmarks remain difficult to outperform in highly autocorrelated financial time series. The analysis also highlights the fundamental challenges that are part of forecasting interest rates. Even the most flexible machine learning models struggle to anticipate abrupt structural breaks associated with sudden changes in monetary policy regimes or macroeconomic conditions. This

limitation highlights that data-driven models rely heavily on past information and that forecast improvements are limited by what historical yields can explain. Future research could extend the present framework by incorporating macroeconomic and policy related covariates, such as inflation measures, output indicators, or central bank communication variables, to enhance model responsiveness during periods of regime change. Additionally, the integration of regime-switching mechanisms or hybrid models that combine machine learning with structural economic insights may offer a promising avenue for improving robustness and interpretability in yield curve forecasting.

References

1. European Central Bank. Statistical Data Warehouse: Euro area yield curve (zero-coupon spot rates). Dataset accessed via ECB SDW.

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

3. Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance*, 57(1), 405–443.

4. Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, 50(4), 745–787.

A AI Tools Used

ChatGPT was used as writing and programming assistance. In particular, it was used to (i) refine report phrasing and improve clarity, (ii) suggest repository structure and reproducible pipeline organization, and (iii) help debug environment and dependency issues. All methodological decisions, model choices, and interpretation of results were made by the author, and the final code was tested locally to ensure that `python main.py` runs end-to-end. Github Copilot was used for inline suggestions while writing code in VS code.

All AI tools were used in this project only as a support instrument to improve the structure and technical efficiency of the project, during the development process. Specifically, AI assistance was used to help with code debugging, which is about identifying mistakes in file paths, environment setup and module imports. AI also helped me review codes that I had already written in order to be more clear. In other cases, Copilot completed obvioud boilerplate.

B Reproducibility Notes

The project is designed to be reproducible: data preprocessing and monthly aggregation are deterministic, and random seeds are fixed in the machine learning models and evaluation routines. To reproduce the results, create the specified environment, ensure the ECB CSV files are located under `data/raw/`, and run `python main.py`. Figures and metrics are written to the `results/` directory.