Ellen Waara Ankarstrand
In Collaboration with SEK Svensk Exportkredit

Yield Curve Forecasting With Nelson-Siegel Models

For Risk Assessment

Table of contents

01

Introduction

Context and research questions

02

Methodology

Data, models, and testing

03

Results

Predictions intervals, accuracy, statistical properties

04

Discussion

Key findings, limitations, and future work

06

Questions

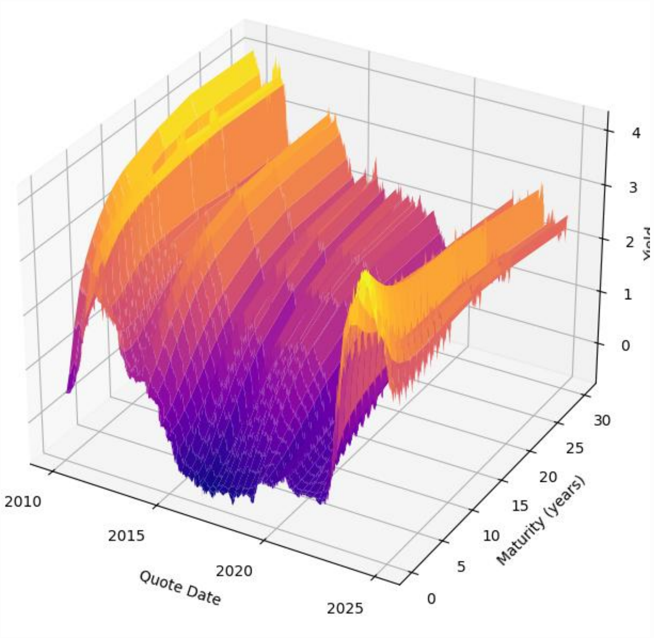
Open for audience questions

01

Introduction

What are yield curves?

- Describe the relationship between time to maturity and interest rate for bonds (usually zero-coupon).
- Evolve over time and across maturities, forming a **three-dimensional surface**
- Not directly observed – estimated from reference rates and market instruments (e.g., bonds, swaps, futures).

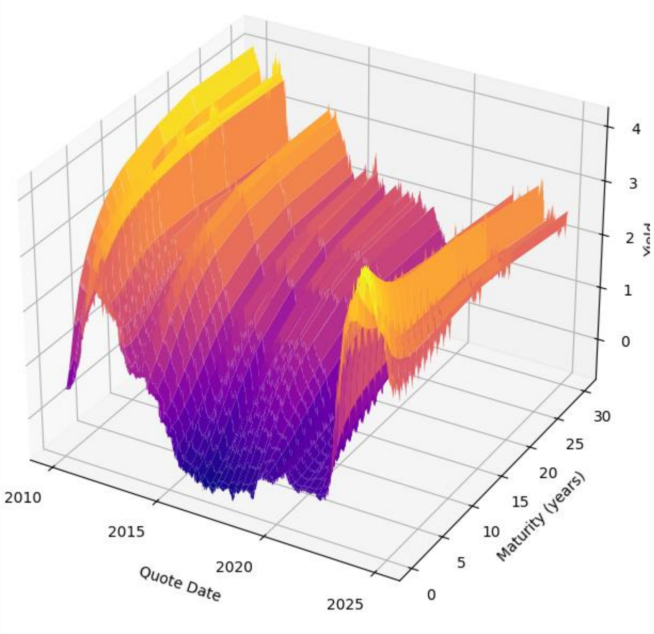


Surface plot of Stibor 2010-2025.

Why do we care about yield curves?

- Reflect market pricing of the **time value of money**, expectations of future interest rates, and monetary policy.

- Reliable modeling is critical for understanding economic conditions and managing financial risks.



Surface plot of Stibor 2010-2025.

Research Questions

Dynamic Nelson-Siegel Models

How do different Dynamic Nelson-Siegel (DNS)-based models compare in terms of **forecast accuracy** and their ability to **replicate statistical properties** of yield curves?

Conformal Prediction

Is conformal prediction useful for uncertainty quantification in yield curve forecasting?

Different Markets

How do these methods perform across the Swedish, European, and US market.

02

Methodology

Methodology Overview

Data

Stibor
Euribor
US Federal Funds

Models

Two-Step DNS
One-Step DNS
Bootstrap DNS

Testing

Prediction Intervals:
Conformal Prediction

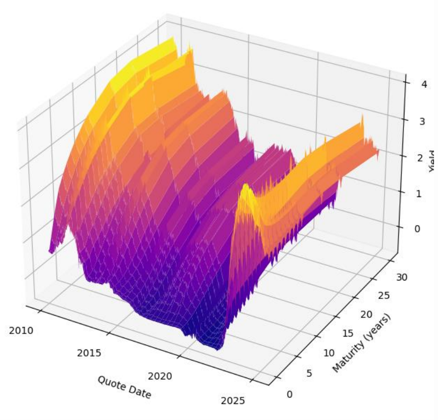
MSE
PCC

Statistical Tests

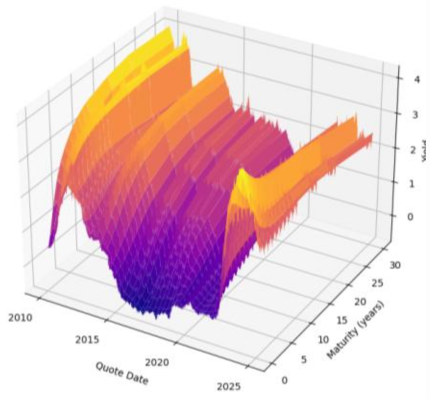
Data

Data

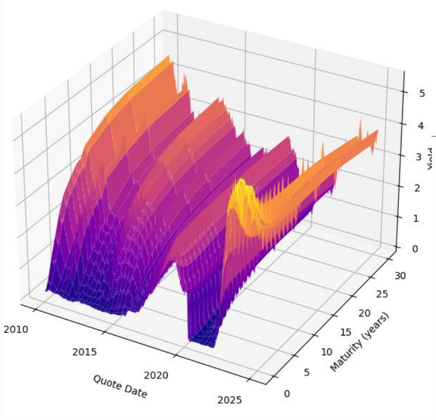
- Daily yield curve data (2010–2024)
- Provided by SEK
- Curves are derived from Stibor 3M, Euribor 3M, and US federal funds using government bonds, swaps, and futures
- Maturities: 3M, 6M, 1–10Y, 15Y, 20Y, 30Y



Stibor



Euribor



US Federal Funds

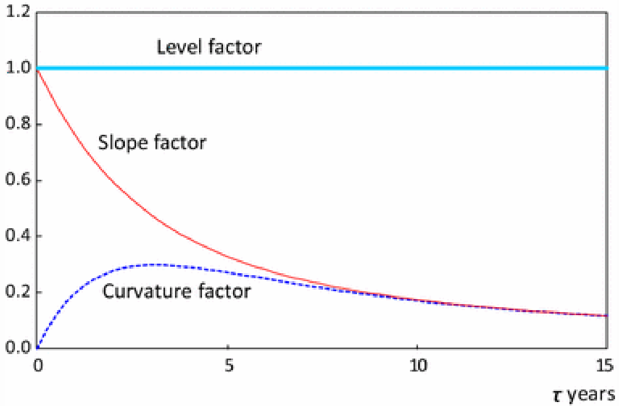
Models

The Nelson-Siegel Model

- Parametric model for yield curves (Nelson et al. 1987)
- Fits diverse curve shapes with just three latent factors
- Formula:

$$y(\tau) = \beta_1 + \beta_2\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_3\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$$

- β1: Level, β2: Slope, β3: Curvature, λ: Decay



Dynamic Nelson Siegel Model (DNS)

- Models factors as time-series → better for forecasting

Two-step DNS with VAR

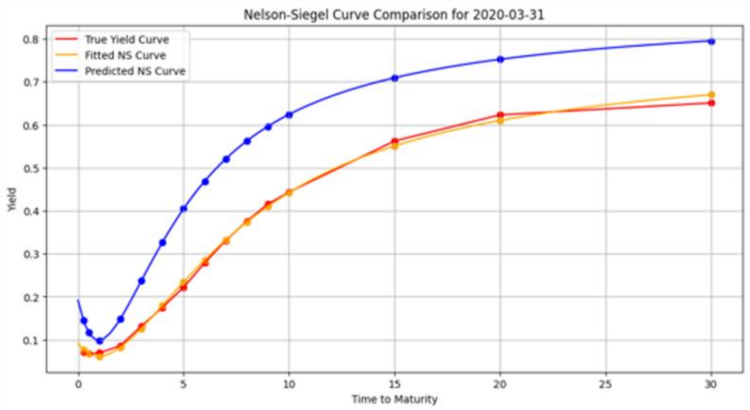
Standard method proposed by Diebold and Li (2006)

Step 1:

- Fix λ
- Fit the factors → time-series

Step 2:

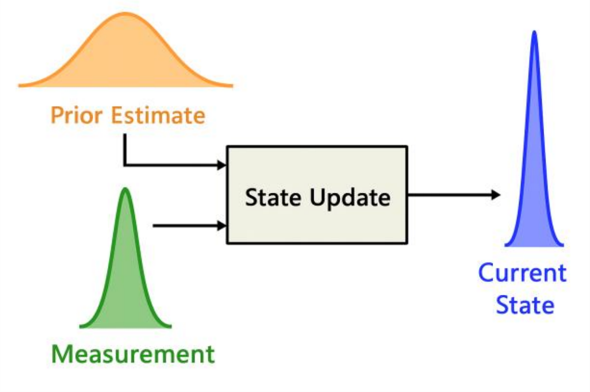
- Forecast the time-series using VAR(1)

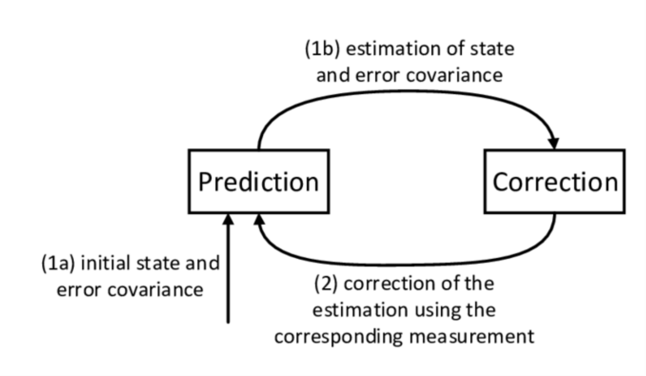


True, fitted, and forecasted curve (14 days) for US federal funds, 2020-03-31

One-step DNS with Kalman Filter

- Using the state-space formulation of DNS
- Latent factors and yields estimated simultaneously
- Kalman filter recursively updates estimates





Bootstrap DNS

- DNS Factors are fitted to historical data using OLS
- Resamples past factor changes using a moving block bootstrap
- Distribution-free, avoids parametric modelling

Observed yields

$\big\downarrow$ OLS

Estimated factors

$$p_i \text{ for } i \in (t-N, t)$$

$$\Delta p_i = p_i - p_{i-1}$$

$\big\downarrow$

Sample blocks of changes

$$\{\Delta p_{t^i_{start}}, \ldots, \Delta p_{t^i_{end}}\}$$

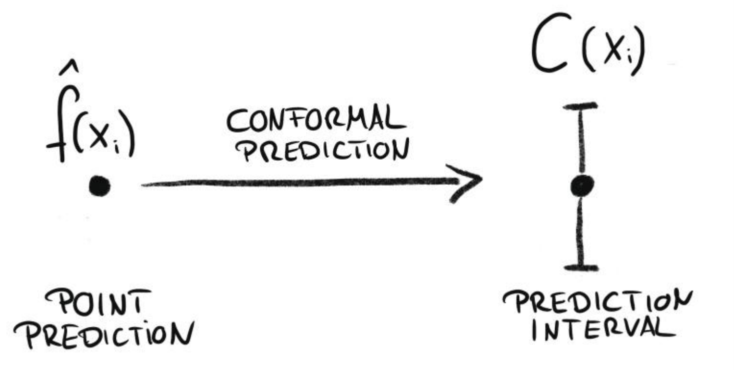
$\big\downarrow$

Sum to create forecast

$$p(t+\Delta t) = p(t) + \sum_{j \leq \Delta t} \Delta p_{t_j}$$

Testing

Quantifying Uncertainty with Conformal Prediction



- Provides guaranteed coverage under mild assumptions for any model
- Computed using the residuals of the fitted model

- Assumes exchangeable data → for time-series we need to apply weighted quantiles
- Weights are chosen to emphasize recent data

- Used split conformal prediction method

Evaluation – MSE and PCC

Mean Squared Error (MSE)

- Measures the average squared difference between predicted and actual yields.

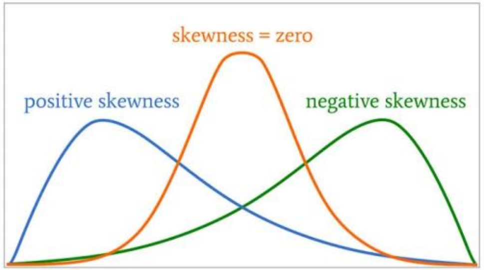
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y_i})^2,$$

Pearson Correlation Coefficient (PCC)

- Measures the strength and direction of a linear relationship.
- Covariance divided by standard deviation
- - 1: perfect negative correlation, 1: perfect positive correlation

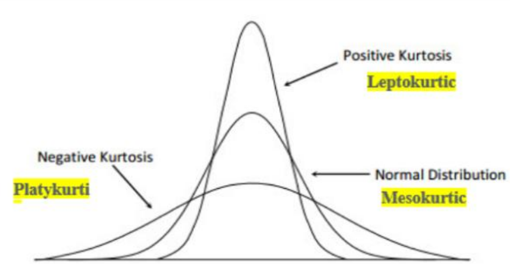
Evaluation – Autocorr and Moments

Comparing statistical properties of observed vs forecasted yields, tests if models **preserve underlying dynamics**



Autocorrelation

- Measures the correlation of a time series with its own lagged values



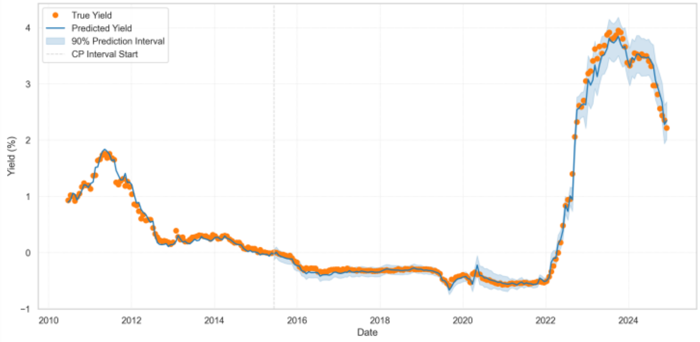
Moments

- Mean, variance, skewness, kurtosis
- **Skewness:** asymmetry, important for identifying directional risk.
- **Kurtosis:** tail heaviness, high kurtosis = fat tails and extreme risks.

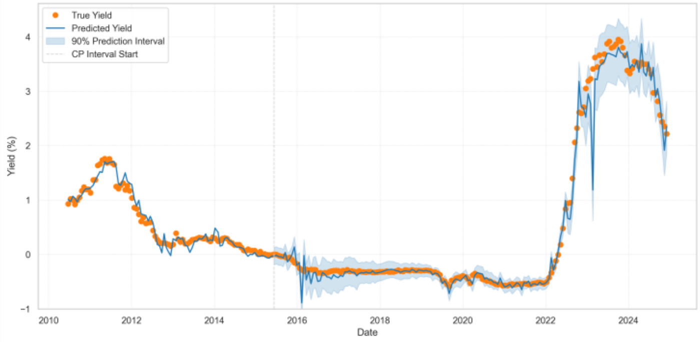
03

Results

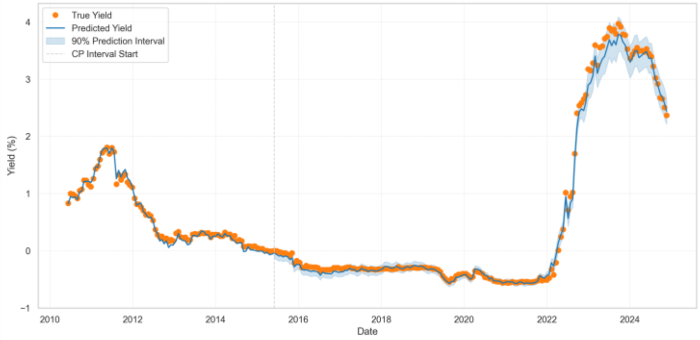
Conformal Prediction — 1 year



Two-step DNS with VAR(1)



One-step DNS with Kalman Filter



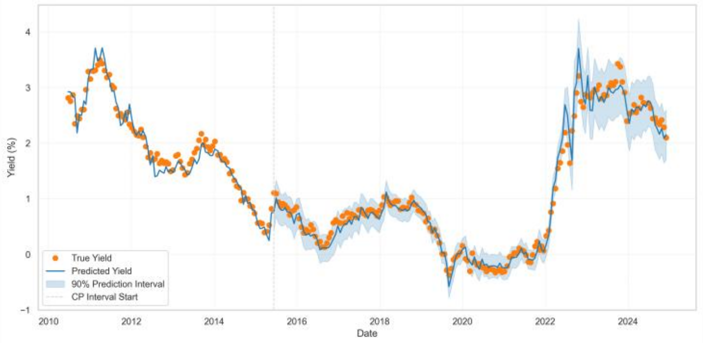
Bootstrap DNS

Euribor, maturity 1 year, coverage 90%

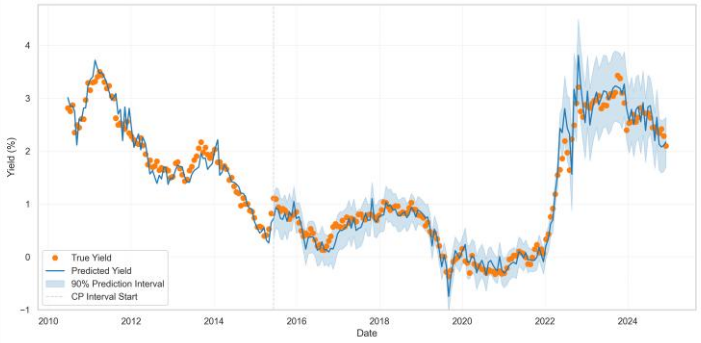
14-day forecast-horizon, 360-day rolling window, 1000 bootstrap samples

Conformal prediction calibration window: 5 years

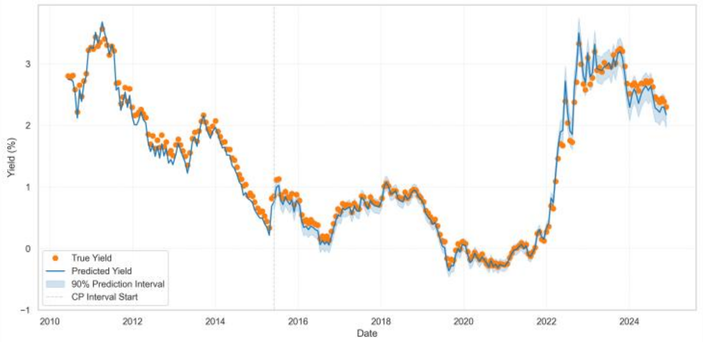
Conformal Prediction — 10 years



Two-step DNS with VAR(1)



One-step DNS with Kalman Filter



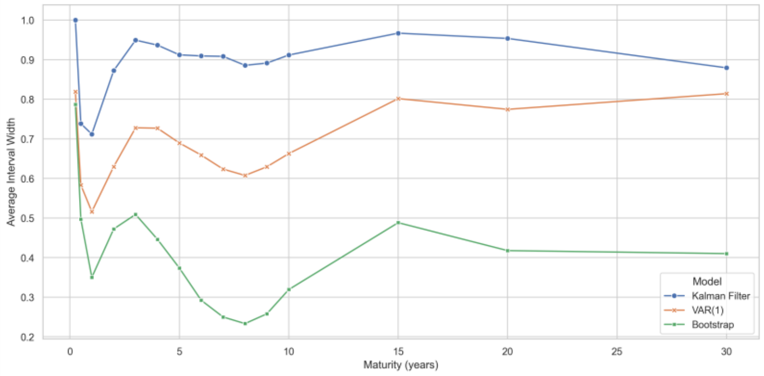
Bootstrap DNS

Stibor, maturity 10 years, coverage 90%

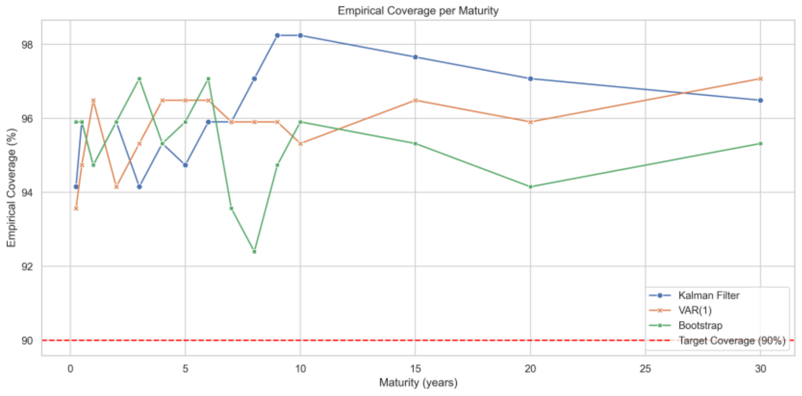
14-day forecast-horizon, 360-day rolling window, 1000 bootstrap samples

Conformal prediction calibration window: 5 years

Conformal Prediction — Coverage

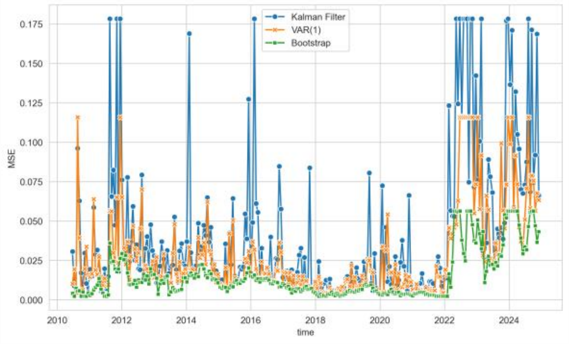


Average conformal prediction interval width over time by maturity and model, coverage level 90%

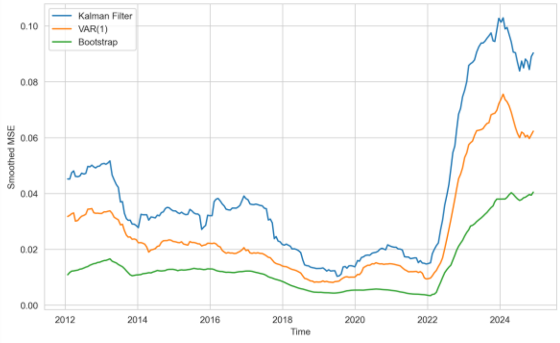


Empirical coverage of the conformal prediction interval width over time by maturity and model, coverage level 90%

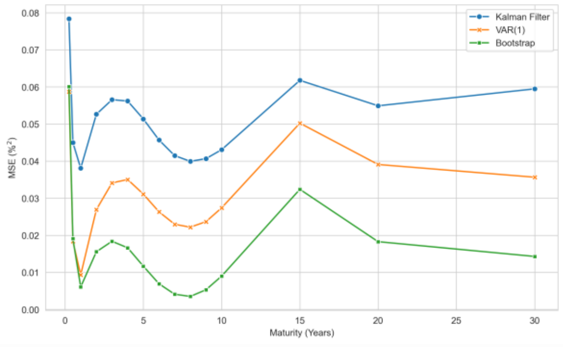
Mean Squared Error (MSE)



(a) MSE over time

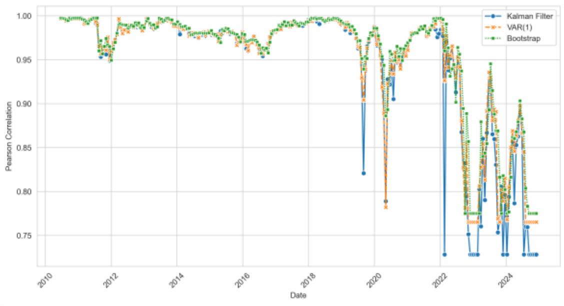


(b) 30-day moving average of MSE

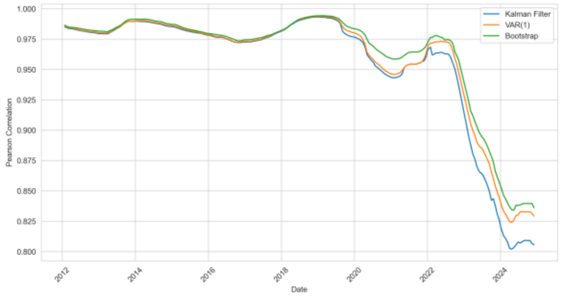


MSE per maturity

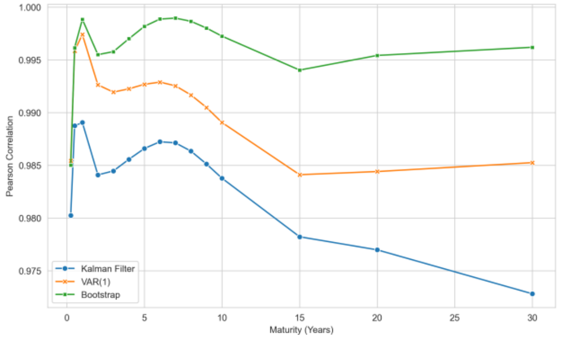
Pearson Correlation Coefficient (PCC)



(a) PCC over time

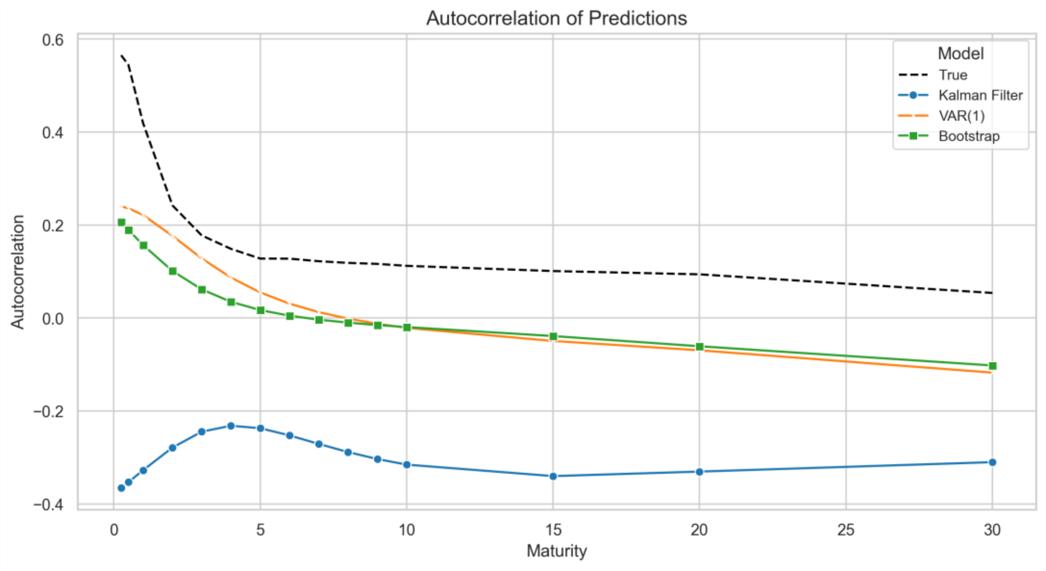


(b) 30-day moving average of PCC



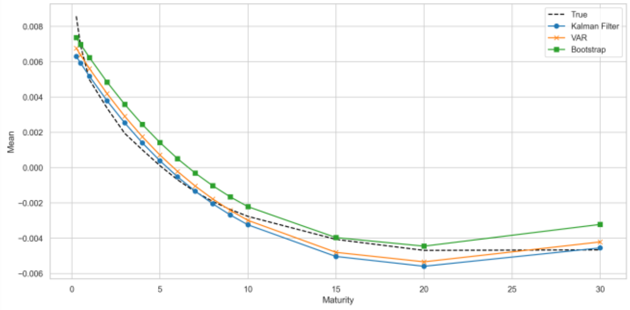
PCC per maturity

Statistical Properties — Autocorrelation

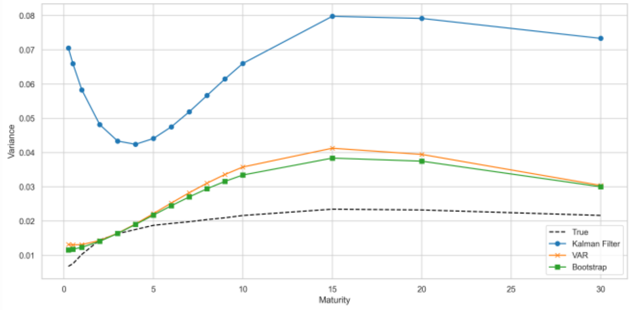


Autocorrelation of 14-day yield differences for Euribor, averaged over time.

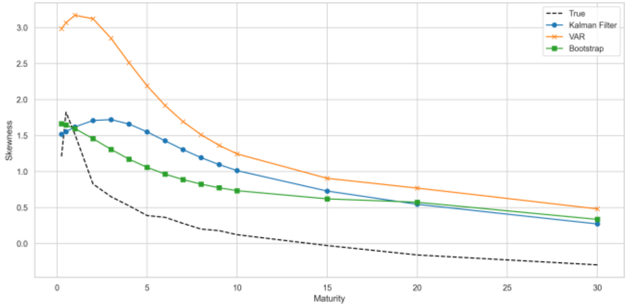
Statistical Properties — Moments



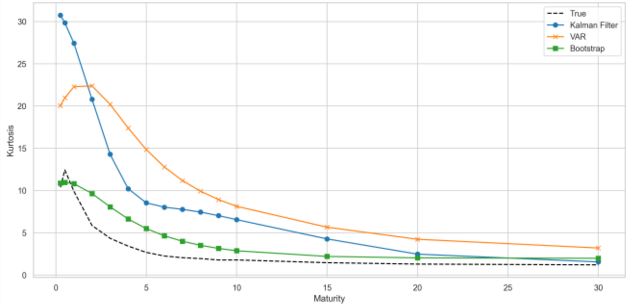
Mean



Variance



Skewness



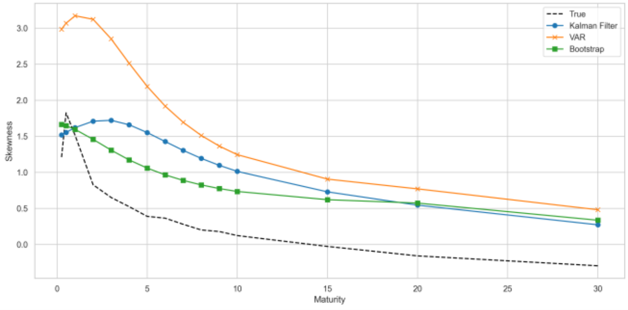
Kurtosis

04

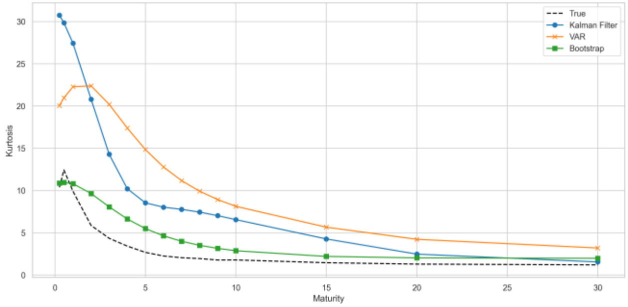
Discussion

Implications for Risk — Tail Risk

- Risk managers focus on extreme events (tail risk), not average yield changes.
- Skewness and kurtosis capture tail behavior.
- The bootstrap model generally matched moments better.
- One- and two-step DNS models tended to overestimste skewness and kurtosis.



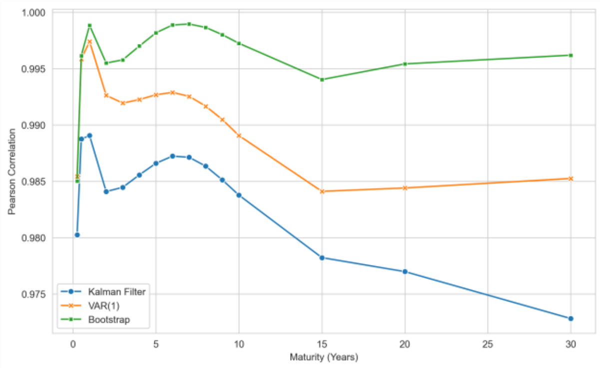
Skewness



Kurtosis

Implications for Risk — Cross-Maturity Performance

- Financial exposures typically span multiple maturities, models must replicate yield changes across the entire curve
- Bootstrap performed best and most consistently across the maturity spectrum for all metrics



PCC per maturity

Implications for Risk — Market Stress

- Model performance dropped under stress for all models
- Bootstrap was most robust, maintaining narrower intervals and more stable predictions.
- One-step DNS performed worst during stress, with large errors and wide intervals.

Image source

Why do we see these differences?

Bootstrap method:

- Non-parametric approach, resampling past changes without imposing strong distributional assumptions.

✓ May help it replicate empirical distribution and serial dependencies

One-step DNS (Kalman filter)

- Estimates the DNS system as a linear-Gaussian state-space model

Two-step DNS (VAR(1))

- Assumes linear factor dynamics and normally distributed innovations

✗ Limit ability to capture complex, non-linear dependencies

Why do we see these differences? Part Two

One-step DNS (Kalman filter)

- One-step DNS with Kalman filter is complex and sensitive to model specification and initial settings.
- Performance may be compromised by model misspecification or suboptimal initialization of matrices.
- While theoretically appealing, its effectiveness depends heavily on correct specification and setup

Conclusions

- Bootstrap DNS shows strong overall performance, providing accurate forecasts, narrower intervals, and replication of statistical properties.
- Two-step DNS is simpler and moderately effective, though it failed to replicate skewness and kurtosis.
- One-step DNS faces challenges under stress. Failed to replicate autocorrelations and variance.

- Conformal prediction provided valid coverage but was overly-conservative.

05

Questions?

00

Additional Slides

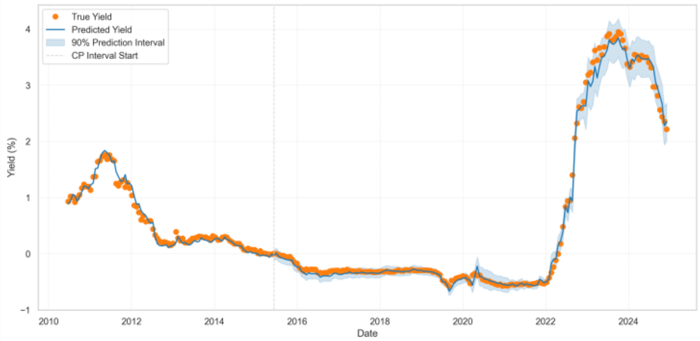
Limitations

Market Stress	All models are vulnerable to market stress, including policy changes or geopolitical events.
Choice of λ	Decay parameter λ was not optimized for this dataset
Initialization	Kalman filter may need better initialization for more reliable performance.
Historical Data	The bootstrap approach depends on historical data, which might not fully capture new market conditions.

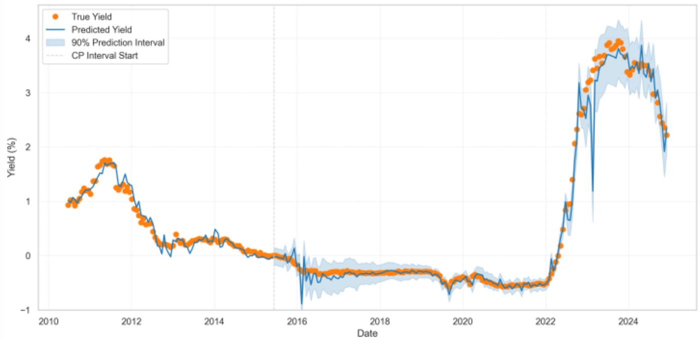
Future Work

- Try different values for the decay parameter λ, such as empirical calibration or a time-varying λ.
- Explore full conformal prediction, which may provide better performance than the split method used here.
- Consider extending the DNS model with the Nelson-Siegel-Svensson to add extra curvature and improve fit.
- Refine the bootstrap approach, potentially using a conditional bootstrap that incorporates macroeconomic data and policy rates.

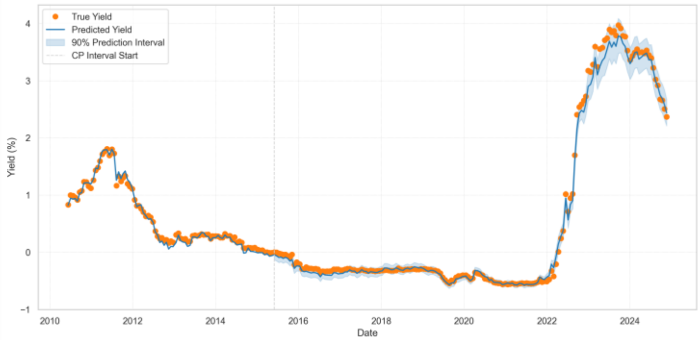
Conformal Prediction — 1 year



Two-step DNS with VAR(1)



One-step DNS with Kalman Filter



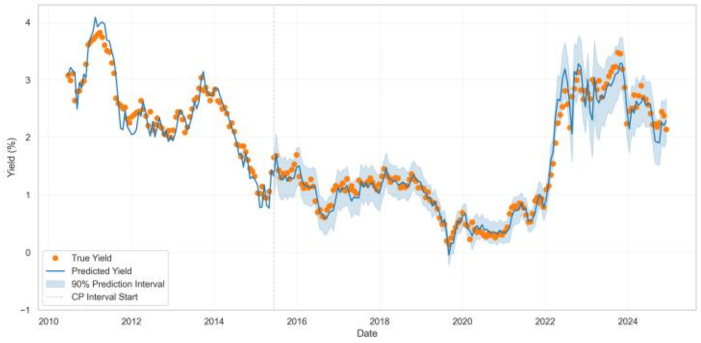
Bootstrap DNS

Stibor, maturity 1 year, coverage 90%

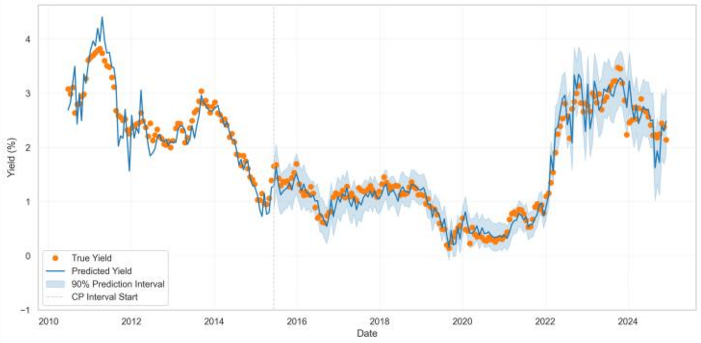
14-day forecast-horizon, 360-day rolling window, 1000 bootstrap samples

Conformal prediction calibration window: 5 years

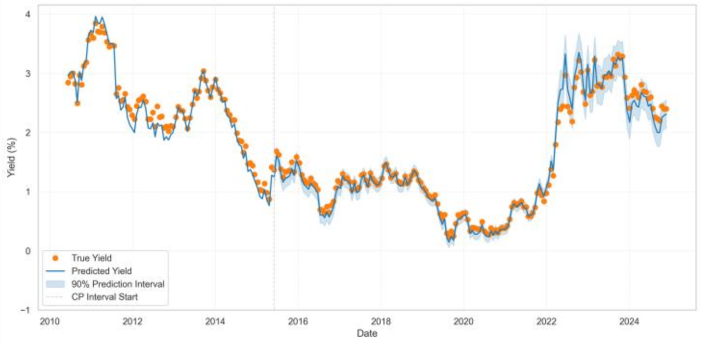
Conformal Prediction — 10 years



Two-step DNS with VAR(1)



One-step DNS with Kalman Filter



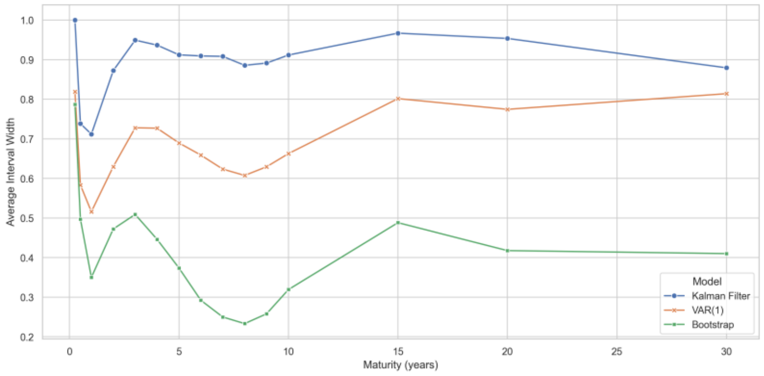
Bootstrap DNS

Stibor, maturity 10 years, coverage 90%

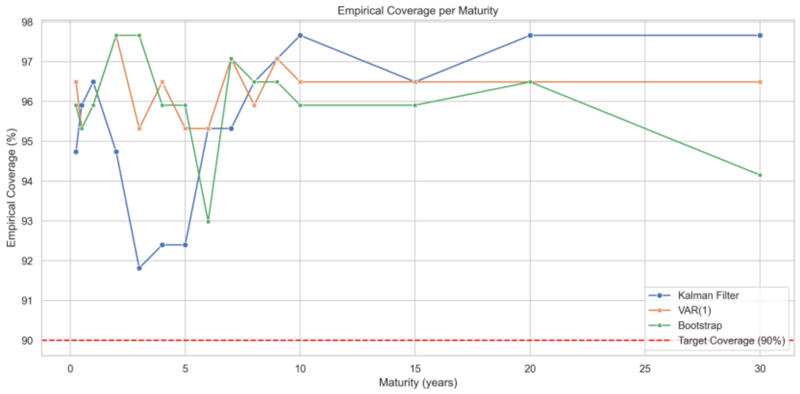
14-day forecast-horizon, 360-day rolling window, 1000 bootstrap samples

Conformal prediction calibration window: 5 years

Conformal Prediction — Coverage

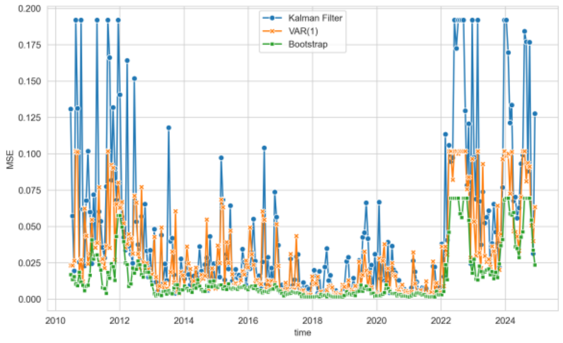


Average conformal prediction interval width over time by maturity and model, coverage level 90%

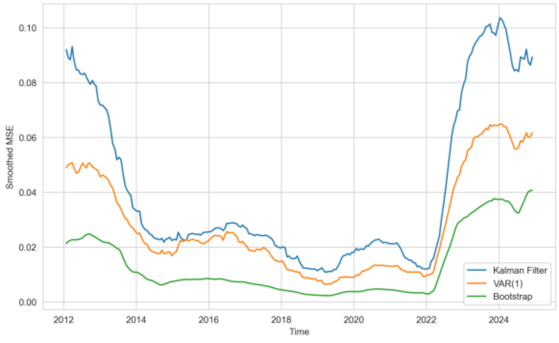


Empirical coverage of the conformal prediction interval width over time by maturity and model, coverage level 90%

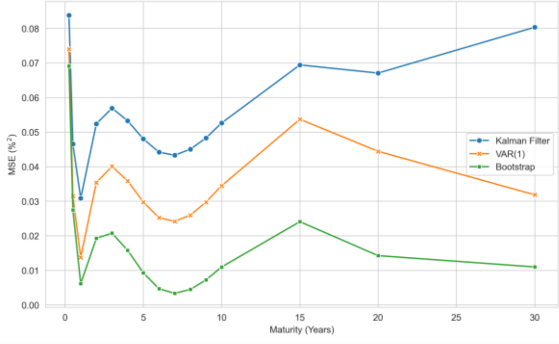
Mean Squared Error (MSE)



(a) MSE over time

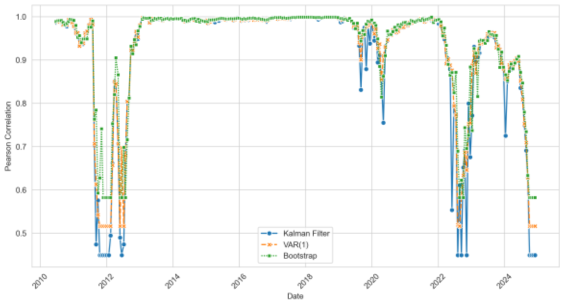


(b) 30-day moving average of MSE

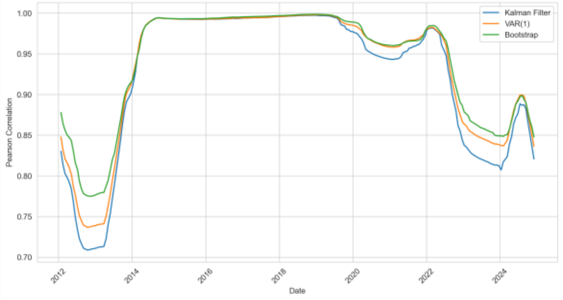


MSE per maturity

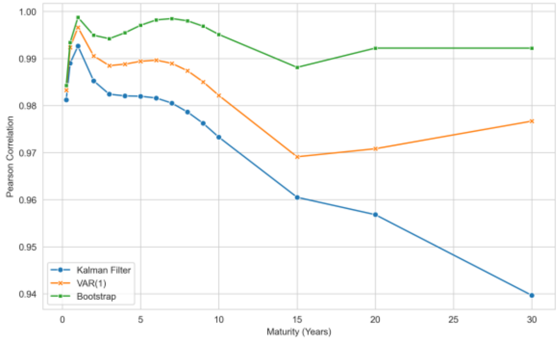
Pearson Correlation Coefficient (PCC)



(a) PCC over time

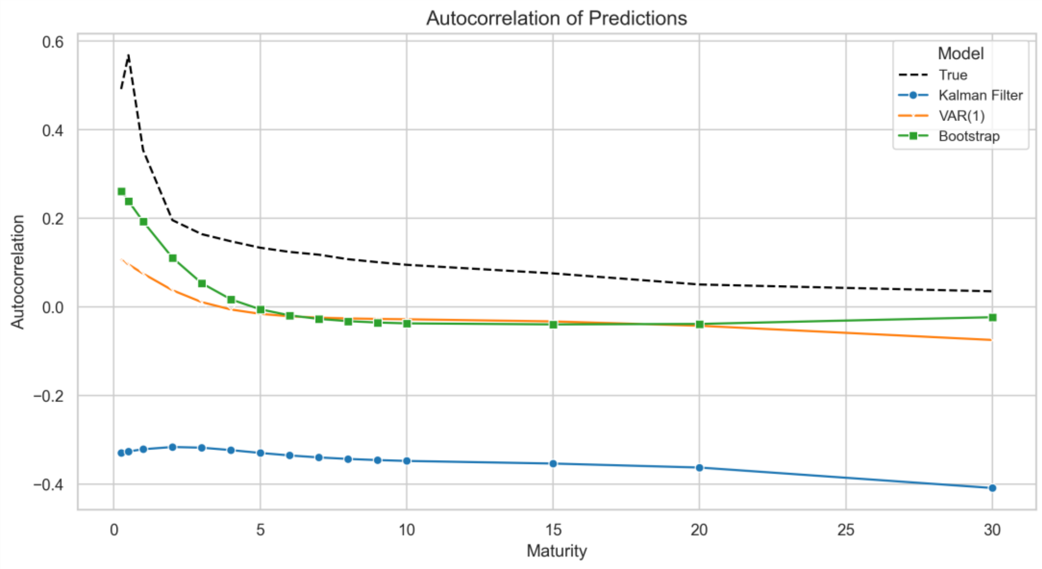


(b) 30-day moving average of PCC



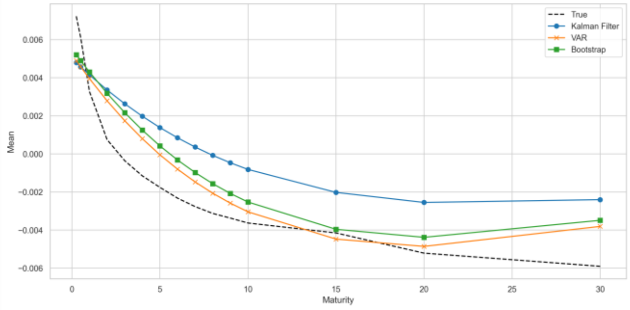
PCC per maturity

Statistical Properties — Autocorrelation

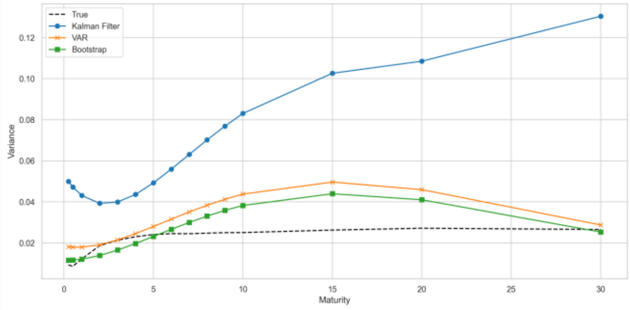


Autocorrelation of 14-day yield differences for Stibor, averaged over time.

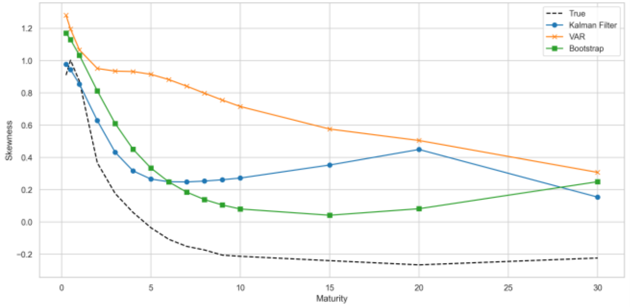
Statistical Properties — Moments, Stibor



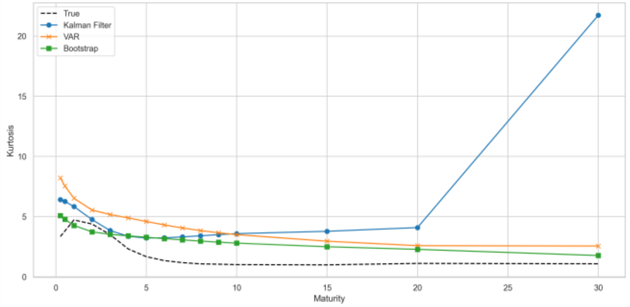
Mean



Variance



Skewness



Kurtosis