Technical Documentation

Version 1.0

Status

Reviewed		
Approved		

PROJECT IDENTITY

2024/HT, Finance project
Linköping University, Dept. of Mathematics (MAI)

Group members

Name	Responsibility	Email
Martin Ladan	Project Leader	marla223@student.liu.se
Maximiliano Lenz	Responsible for testing and validation	maxle066@student.liu.se
Michael Birtman	Responsible for the derivation of Kalman filtering	micbi949@student.liu.se
Axel Karlsson	Responsible for the interpolation	axeka624@student.liu.se
Ludvig Ottosson	Responsible for the documentation	ludot931@student.liu.se
Carl Jungerth	Responsible for the implementation of Kalman filtering	carju678@student.liu.se
Nils Åsman	Responsible for OIS and PCA	nilas587@student.liu.se

Email list for the whole group: marla223@student.liu.se
Customer: Mathias Henningsson, IEI, Linköpings universitet, 581 83 Linköping
Customer contact: Mathias Henningsson, 013-28 21 82, mathias.henningsson@liu.se
Course leader: Elina Rönnberg, 013-281645, elina.ronnberg@liu.se
Tutor: Jörgen Blomvall, 013-281406, jorgen.blomvall@liu.se

Nomenclature

AD	Automatic Differentiation
EM	Expectation Maximization
KKT	Karush-Kuhn-Tucker
MSE	Mean Squared Error
OIS	Overnight Index Swap
ON	Overnight
PCA	Principal Component Analysis
Q-Q	Quantile-Quantile

Contents

Document history

Version	Date	Changes	Sign	Reviewed
0.1	2024-12-14	First draft.	All	
1.0	2024-01-27	Approved draft.	All	

1 Introduction

This document presents the technical documentation for the Swap Squad system and is intended to serve as a comprehensive guide to the entire system. It is part of the project work for the TATA62 course, Project in Applied Mathematics, at Linköping University, undertaken in the fall of 2024.

1.1 Background

In a world where markets move faster than ever and economic decisions have global significant consequences, the ability to foresee future trends is more valuable than ever. Understanding market trends and economic signals is crucial for making valid financial decisions, and one of the most powerful tools for this is the forward curve. The forward curve provides important insights into market expectations for future interest rates and therefore also gives hints for the expected coming inflation and economic situation. A significant factor embedded in the forward curve is the expected outcome of monetary policy meetings, where the decision to hike or cut the rate is made.

The ability to accurately interpret the forward curve can be the difference between making well-informed financial decisions and being caught off guard by sudden market shifts. By understanding the forward curve, investors, policymakers, and corporations can anticipate the economic environment and adjust their strategies accordingly. This makes the forward curve not just a tool for financial analysis, but a vital part of decision-making in a market landscape.

1.2 Objective and Aim

In this project, the focus will be on constructing a model that can accurately depict the implied forward curve over a 10-year period, from the OIS yield curve. The aim is to have a model that factors in outcomes of central bank meetings and accounts for potential impact of stale or incorrect data.

By doing so, the model seeks to provide a more reliable representation of market expectations for future interest rates and the broader economic outlook. This document outlines the approach, methodology, and validation process required to develop and test such a model.

1.3 Application and Market Importance

The continuous forward curve model developed in this project will serve as a vital tool for pricing interest rate-based derivatives and making investment decisions, providing valuable insights across the financial and economic landscape. By ensuring accurate pricing of derivatives, the model contributes to the goal of maintaining an arbitrage-free market, which is essential for maintaining market stability and further emphasizes its importance.

2 Valuation of OIS-instruments

This section will describe the theoretical pricing of the interest rates derivatives used in the project.

An OIS is an interest rate swap where the floating leg depends on the ON interest rates. This means that a fixed interest rate is exchanged for a geometric average of the overnight interest rates for a fixed period. The OIS rates is regarded as a proxy for the risk free rate (Hull 2017). The reasoning for this is that the probability of a financial institution defaulting on an overnight loan is seen as very slim.

The floating interest rate r_i^o is exchanged for a fixed interest rate y^{OIS} and payments is executed at time points $\tilde{T}_k \in \{\tilde{T}_1, \ldots, \tilde{T}_K\}$. For a time period between $\bar{T}_i$ to $\bar{T}_j$ with payment at time $\tilde{T}_k$ we can define the set $(i,j,k) \in \mathcal{P}$. We can approximate the compounded risk-free rate $f(t)$ with the daily compounded interest r_i^o as

$$\prod_{i=1}^{n} \left(1 + r_i^o \Delta \bar{T}_i^o\right) - 1 \approx \exp\left\{\int_{\bar{T}_i}^{\bar{T}_j} f(s)\, ds\right\} - 1, \tag{2.1}$$

where n is the number of business days between the dates $\tilde{T}_i$ and $\tilde{T}_j$ and $i \in \{1, \ldots, n\}$ denotes the i:th business day.

The initial value of an OIS-contract should be zero, to avoid any arbitrage opportunities. And thus

$$P_{\text{OIS}} = P_{\text{OIS}}^{\text{fix}} - P_{\text{OIS}}^{\text{fl}} = 0. \tag{2.2}$$

We can define the payout from the fixed leg at time $\tilde{T}_k$ as

$$\Phi_{t_k}^{\text{fix}}\left(\tilde{T}_k, y^{\text{OIS}}\right) = N y^{\text{OIS}} \Delta \tilde{T}_k, \tag{2.3}$$

where N is the nominal amount (Färm and Hammarbäck 2024). We can thus value the fixed leg by summarizing the payout at each payment time and discounting each of the payments with the correct discount rate as follows

$$P_{\text{OIS}}^{\text{fix}} = N y^{\text{OIS}} \sum_{k=1}^{K} \Delta \tilde{T}_k \frac{P_{t,\tilde{T}_k}}{P_{t,T_0}}. \tag{2.4}$$

The price, P_{t,T_i} of a zero-coupon bond is equal to a discount rate for deterministic interest rates. The floating leg is approximated as

$$P_{\text{OIS}}^{\text{fl}} = \sum_{(i,j,k)\in\mathcal{P}} N \left(\frac{P_{t,\tilde{T}_i}}{P_{t,\tilde{T}_j}} - 1\right) \frac{P_{t,\tilde{T}_k}}{P_{t,T_0}}. \tag{2.5}$$

The zero-coupon bonds are calculated as

$$P_{t,T_i} = \mathbb{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{T_i} f(t,s)\,ds}\right]. \tag{2.6}$$

By discretizing the risk-free forward rate $f(t)$ with daily time interval as $\Delta t = \frac{1}{365}$

$$\int_{T_j}^{T_i} f(s)\,ds \approx \sum_{t=T_j}^{\frac{T_n}{\Delta T}-1} f(t)\Delta t. \tag{2.7}$$

Furthermore, we can calculate the yield for an OIS contract, y^{OIS} as

$$y^{\text{OIS}} = \frac{P_{t,T_0} - P_{t,T_K}}{\sum_{i=1}^{K} \Delta T_i P_{t,T_i}}, \tag{2.8}$$

where ΔT_i is the year fraction between $\tilde{T}_{i-1}$ and $\tilde{T}_i$ (Rydholm and Villwock 2022).

3 Principal Component Analysis

This section will describe how principal component analysis, PCA, will be used for dimensionality reduction.

We want to express the forward curve in terms of risk factors which is done by eigen decomposition of the covariance matrix of the forward curve, $\Sigma_{f_t} = \text{cov}(f_t)$. The matrix f_t consist of m different discretized forward curves covering a period of n days. The forward curve will cover the next 10 years. However the daily discretized forward curve will vary in length due to calender effects. Thus the data has to be processed before we can calculate the covariance function. Hence $f_t \in \mathbb{R}^{n\times m}$ and $\Sigma_{f_t} \in \mathbb{R}^{n\times n}$.

$$\Sigma_{f_t} = Q\Lambda Q^T, \tag{3.1}$$

where Q is the matrix containing the eigenvectors and Λ is the diagonal matrix containing the eigenvalues. We can use a subset of these eigenvectors to approximate the covariance matrix as

$$\Sigma_{f_t} \approx Q_K \Lambda_K Q_K^T, \tag{3.2}$$

where Q_K and Λ_K represents the first K eigenvectors and eigenvalues. We can thus represent the risk factors ξ_K as the approximation

$$\xi_t \approx Q_K^T f_t, \tag{3.3}$$

which follows that

$$f_t \approx Q_K \xi_t, \tag{3.4}$$

where $f_t \in \mathbb{R}^{n\times 1}$ is a vector of the daily discretized forward curve. This follows since Q_K is a orthogonal matrix. This representation of the forward curve will allow us to represent the forward curve with the risk factors $\xi_t \in \mathbb{R}^{K\times 1}$ which will be used as state variables and thus reduce the dimensionality of problem substantially. This means that we only need K variables to make a good approximation of the forward curve.

4 Optimization

The optimization system will estimate optimal states for the defined model. The states will describe the forward curve using risk factors from the PCA and central bank interest changes on policy meeting dates. The model will also estimate unsystematic risk factors, i.e. pricing errors of the different instruments.

4.1 Pricing Functions

The pricing function is used as a constraint

$$z_t = g_t(x_t^s) + I_t^u x_t^u + v_t, \quad t = 1, \dots, T, \tag{4.1}$$

where z_t is a vector of OIS yields from the market and x_t^s is state variables representing the eigenvalues and interest rate decisions on monetary policy meeting dates. The variable x_t^u represents pricing errors in the OIS instruments represented in terms of yields and v_t is student-t distributed noise variable.

We can write $g_t(x_t^s)$ as a function of discount factors, and the discount factors can be described using the forward curve. The forward curve will consist of risk factors from PCA and jumps in the forward curve made on policy meeting days. This gives us the forward yield curve

$$f_t = Q_K x_t^p + C x_t^c, \tag{4.2}$$

where $x_t^p = \xi_K$. We will use

$$o_j = \begin{pmatrix} -\Delta t 1_{i<T_j}^T Q_K & -\Delta t 1_{i<T_j}^T C \end{pmatrix}, \tag{4.3}$$

where the matrix C is a matrix which is 1 for the days after a policy meeting day and $C \in \mathbb{R}^{T \times c}$ where T is the number of days of the discretized forward curve and c is the number of policy meeting days. The vector $1_{i<T}^T$ is 1 for all elements with position smaller than T which we use to summarize the forward elements in the forward curve, since we use a discretization of the forward rate with daily time interval. We will write the state variables as $x_t^s = (x_t^p; x_t^c)$ and can thus express the discount factors as

$$d(t, T) = e^{o_T x_t^s}. \tag{4.4}$$

The yield can thus be described as

$$g(x_t^s) = \frac{e^{o_0 x_t^s} - e^{o_K x_t^s}}{\sum_{j=1}^K \Delta T_j e^{o_j x_t^s}}. \tag{4.5}$$

We can differentiate this function w.r.t x_t^s and by the quotient rule we get

$$\nabla g(x_t^s) = \frac{o_0 e^{o_0 x_t^s} - o_K e^{o_K x_t^s}}{\sum_{j=1}^K \Delta T_j e^{o_j x_t^s}} - \frac{\left(e^{o_0 x_t^s} - e^{o_K x_t^s}\right) \sum_{j=1}^K \Delta T_j o_j e^{o_j x_t^s}}{\left(\sum_{j=1}^K \Delta T_j e^{o_j x_t^s}\right)^2}. \tag{4.6}$$

4.1.1 Derived Pricing Function

The pricing equation outlined in Equation (4.25) is the standard pricing function derived from the original problem. However, given a solution, as described in Section 4.3, we can derive a modified version of the constraint

$$
\begin{aligned}
z_t &= g_t(x_t^s) + I_t^z x_t^u + v_t, \text{ with } x_t^s = \bar{x}_t^s + \Delta x_t^s, x_t^u = \bar{x}_t^u + \Delta x_t^u, v_t = \bar{v}_t + \Delta v_t && (4.7)\\
&= g_t(\bar{x}_t^s + \Delta x_t^s) + I_t^z(\bar{x}_t^u + \Delta x_t^u) + \bar{v}_t + \Delta v_t, && (4.8)
\end{aligned}
$$

and with one rearrangement, we get

$$\Delta x_t^s = x_t^s - \bar{x}_t^s, \tag{4.9}$$

where $\bar{x}_t^s$ is the prior state variable. Using this rearrangement, we can linearize the non-linear function $g_t(\bar{x}_t^s + \Delta x_t^s)$ by a Taylor expression of order 1, where we disregard the ordo-term as

$$g_t(\bar{x}_t^s + \Delta x_t^s) \approx g_t(\bar{x}_t^s) + \nabla g_t(\bar{x}_t^s)\Delta x_t^s \tag{4.10}$$

By inserting this into equation (4.8) we can express

$$
\begin{aligned}
z_t &= g_t(\bar{x}_t^s) + \nabla g_t(\bar{x}_t^s)\Delta x_t^s + I_t^z(\bar{x}_t^u + \Delta x_t^u) + \bar{v}_t + \Delta v_t\\
&= g_t(\bar{x}_t^s) + I_t^z \bar{x}_t^u + \nabla g_t(\bar{x}_t^s)\Delta x_t^s + I_t^z \Delta x_t^u + \bar{v}_t + \Delta v_t, \text{ with } \bar{\omega}_t = g_t(\bar{x}_t^s) + I_t^z \bar{x}_t^u\\
&= \bar{\omega}_t + \nabla g_t(\bar{x}_t^s)\Delta x_t^s + I_t^z \Delta x_t^u + \bar{v}_t + \Delta v_t\\
0 &= \nabla g_t(\bar{x}_t^s)\Delta x_t^s + I_t^z \Delta x_t^u + \bar{v}_t + \Delta v_t - z_t + \bar{\omega}_t.
\end{aligned}
$$

We can use

$$
\begin{aligned}
\Delta x_t &= [\Delta x_t^s; \Delta x_t^u], && (4.11)\\
H_t &= [\nabla g_t(\bar{x}_t^s) \quad I_t^z], && (4.12)
\end{aligned}
$$

to simplify the expression. Thus by inserting into equation (4.11) we can express it as

$$
\begin{aligned}
0 &= H_t \Delta x_t + \bar{v}_t + \Delta v_t - z_t + \bar{\omega}_t && (4.13)\\
&= H_t \Delta x_t + \Delta v_t + \delta_t^v, && (4.14)
\end{aligned}
$$

where $\delta_t^v = \bar{v}_t - z_t + \bar{\omega}_t$.

4.2 Derived State Change Functions

The state change equation outlined in Equation (4.24) is the standard state function derived from the original problem. However, given a solution, as described in Section 4.3, we derive a modified version of this constraint to the optimization model. We express this as

$$
\begin{aligned}
x_t &= F_t x_{t-1} + w_t, \text{ with } w_t = \bar{w}_t + \Delta w_t, x_t = \bar{x} + \Delta x_t && (4.15)\\
\bar{x}_t + \Delta x_t &= F_t(\bar{x}_{t-1} + \Delta x_{t-1}) + \bar{w}_t + \Delta w_t, && (4.16)
\end{aligned}
$$

and by reordering the terms we get

$$\Delta x_t = F_t \Delta x_{t-1} + \Delta w_t + \delta_t^w, \tag{4.17}$$

with $\delta_t^w = (F_t \bar{x}_{t-1} + \bar{w}_t - \bar{x}_t)$.

4.3 Derived Optimal Objective Function

The state space, $x_t = (x_t^s; x_t^u) \in \mathbb{R}^{n_t^x \times 1}$ is split into the systematic risk factor states $x_t^s \in \mathbb{R}^{n_t^s \times 1}$, which define the interest rate curve, and the unsystematic risk factor states $x_t^u \in \mathbb{R}^{n_t^u \times 1}$, which define price deviations for each instrument. Furthermore, the systematic risk factors $x_t^s = (x_t^p; x_t^c)$ are split into the principal components, $x_t^p \in \mathbb{R}^{n_t^p \times 1}$, from PCA analysis of the interest rate curves, and central bank interest rate changes $x_t^c \in \mathbb{R}^{n_t^c \times 1}$ (steps in the interest rate curve). States evolve as

$$x_t = F_t x_{t-1} + w_t, \tag{4.18}$$

with

$$F_t = \begin{pmatrix} I_t^x & 0 \\ 0 & D_t \end{pmatrix}, \tag{4.19}$$

where $D_t \in \mathbb{R}^{n_t^u \times n_{t-1}^u}$ is a diagonal matrix with positive values smaller than (or equal to) one, which (can) drive x_t^u towards zero, and $w_t \in \mathbb{R}^{n_t^x \times 1}$ is the noise. Note that for dates when a step is removed or introduced, $I_t^x \in \mathbb{R}^{n_t^s \times n_{t-1}^s}$ will not be the identity matrix. This means that the matrix will essentially change when a new policy meeting date leaves or enters the range of the forward curve. Observations of the market prices,

$$z_t = g_t(x_t^s) + I_t^z x_t^u + v_t, \tag{4.20}$$

with dimensionality $z_t \in \mathbb{R}^{n_t^z \times 1}$, has a different number of observations, n_t^z over time. The nonlinear function $g_t : \mathbb{R}^{n_t^s \times 1} \to \mathbb{R}^{n_t^z \times 1}$ determines the quoted rates (e.g., interest rates) implied by the forward curve described by the state variable x_t^s. $I_t^z \in \mathbb{R}^{n_t^z \times n_t^u}$ contains ones for the instruments that have an observation at time t, and $v_t \in \mathbb{R}^{n_t^z \times 1}$ is noise (Blomvall 2024b).

We estimate the parameters by maximizing the likelihoods of w_t, v_t and x_t. The log-likelihood objective function is given by,

$$\max \prod_{i=0}^{T} f_t^w(w_i) f_t^v(v_i) f_t^x(x_i) \iff \max \ln \left(\prod_{i=0}^{T} f_t^w(w_i) f_t^v(v_i) f_t^x(x_i) \right) =$$
$$= \max \ln f_0^x(x_0) + \sum_{i=1}^{T} (\ln f_t^w(w_i) + \ln f_t^v(v_i) + \ln f_t^x(x_i))$$

where f_t^w, f_t^v, f_t^x are the multivariate probability density functions of the variables w_t, v_t and x_t. These pdf:s depend on the distribution set for the variables. The variable v_t will most likely be set as a multivariate Student's t-distribution, since it represents the noise of the market quotes. We get the pdf

$$f_t^v(v_t; \mu, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu + n_t^z}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \nu^{n_t^z/2} \pi^{n_t^z/2} |\Sigma|^{1/2}} \left[1 + \frac{1}{\nu} (v_t - \mu)^\top \Sigma^{-1} (v_t - \mu) \right]^{-\frac{\nu + n_t^z}{2}}, \tag{4.21}$$

where Σ is a diagonal matrix if the elements in v_t are independent and with the function

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt, \quad \text{for } x > 0. \tag{4.22}$$

The pdf of w_t and x_t will of course be similar if they follow a multivariate Student's t-distribution. However if we assume that e.g. w_t follows a multivariate normal distribution we get the pdf

$$f(w_t; \mu, \Sigma) = \frac{1}{(2\pi)^{n_t^x/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(w_t - \mu)^\top \Sigma^{-1}(w_t - \mu)\right), \tag{4.23}$$

where Σ is a diagonal matrix if the elements in w_t are independent. We can write our defined state space function and pricing function as its conditions,

$$\text{s.t.} \quad x_t = F_t x_{t-1} + w_t, \qquad t = 1, \ldots, T, \tag{4.24}$$

$$z_t = g_t(x_t^s) + I_t^s x_t^u + v_t, \qquad t = 1, \ldots, T. \tag{4.25}$$

where the likelihood functions are defined as $f_t^w : \mathbb{R}^{n_t^x \times 1} \to \mathbb{R}$, $f_t^v : \mathbb{R}^{n_t^z \times \mathbb{R}} \to \mathbb{R}$, and $f_t^x : \mathbb{R}^{n_t^x \times 1} \to \mathbb{R}$ (Blomvall 2024a). Since the function $g_t(x_t^s)$ is linear, $\ln f_0^x(x_0)$ is arbitrarily set to zero, and both f_t^w and f_t^v are multivariate normal distributions, we see that Rauch-Tung-Striebel gives the optimal solution to the optimization problem.

Based on the nature of the optimization problem, we can apply this two-pass algorithm using Riccati recursion, to find the optimal solution to the following optimization model for $t = 1, ..., T$,

$$\max \ln f_t^w(w_t) + \ln f_t^v(v_t) + \ln f_t^x(x_t), \tag{4.26}$$

$$\text{s.t.} \quad x_t = F_t x_{t-1} + w_t, \tag{4.27}$$

$$z_t = g_t(x_t^s) + I_t^s x_t^u + v_t, \tag{4.28}$$

for each $t = 1, \ldots, T$. When at least one assumption do not hold, then quadratic sub-problems can be solved iteratively to find a local optima (Rauch, Tung, and Striebel 1965).

The Taylor expansion of a general function is given by the expression,

$$-\ln y \approx -\ln f(\tilde{y}) + \nabla \ln f(\tilde{y})^T (y - \tilde{y}) + \frac{1}{2}(y - \tilde{y})^T \nabla^2 \ln f(\tilde{y})(y - \tilde{y}). \tag{4.29}$$

To Taylor expand the natural logarithm of 4.21 we define

$$C(n_t^z; \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu + n_t^z}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \nu^{n_t^z/2} \pi^{n_t^z/2} |\Sigma|^{1/2}} \tag{4.30}$$

as it only depends on the dimensionality of the random variable and hyperparameters and we can express the log-likelihood as

$$
\begin{aligned}
\ln f_t^v(v_t;\mu,\Sigma,\nu) &= \ln(C(n_t^z;\Sigma,\nu)) + \ln\left(1+\frac{1}{\nu}(v_t-\mu)^\top\Sigma^{-1}(v_t-\mu)\right)^{-\frac{\nu+n_t^z}{2}} \\
&= \ln(C(n_t^z;\Sigma,\nu)) - \frac{\nu+n_t^z}{2}\ln\left(1+\frac{1}{\nu}(v_t-\mu)^\top\Sigma^{-1}(v_t-\mu)\right) \qquad (4.31)
\end{aligned}
$$

We differentiate w.r.t. v_t and get

$$
\nabla_{v_t}\ln f_t^v(v_t;\mu,\Sigma,\nu) = -\frac{\nu+n_t^z}{2}\left(\frac{\frac{2}{\nu}\Sigma^{-1}(v_t-\mu)}{1+\frac{1}{\nu}(v_t-\mu)^\top\Sigma^{-1}(v_t-\mu)}\right) \tag{4.32}
$$

which we can differentiate once again with the quotient rule and we get

$$
\begin{aligned}
\nabla_{v_t}^2\ln f_t^v(v_t;\mu,\Sigma,\nu) = -\frac{\nu+n_t^z}{2}\Bigg(&\frac{\frac{2}{\nu}\Sigma^{-1}}{1+\frac{1}{\nu}(v_t-\mu)^\top\Sigma^{-1}(v_t-\mu)} \\
&-\frac{\frac{4}{\nu^2}\Sigma^{-1}(v_t-\mu)(v_t-\mu)^\top\Sigma^{-1}}{\left(1+\frac{1}{\nu}(v_t-\mu)^\top\Sigma^{-1}(v_t-\mu)\right)^2}\Bigg). \qquad (4.33)
\end{aligned}
$$

However, if the variable follows a multivariate normal distribution, we have to Taylor expand the natural logarithm of 4.23 where we can use

$$
D(n_t^x;\Sigma) = \frac{1}{(2\pi)^{n_t^x/2}|\Sigma|^{1/2}}, \tag{4.34}
$$

and thus we can write the natural logarithm of the pdf of w_t as

$$
\ln f_t^w(w_t;\mu,\Sigma) = \ln D(n_t^z;\Sigma) + \left(-\frac{1}{2}(w_t-\mu)^T\Sigma^{-1}(w_t-\mu)\right). \tag{4.35}
$$

We can differentiate this function w.r.t. w_t and we get the simple expression

$$
\nabla_{w_t}\ln f_t^w(w_t;\mu,\Sigma) = -\Sigma^{-1}(w_t-\mu), \tag{4.36}
$$

which we can differentiate one more time and get

$$
\nabla_{w_t}^2\ln f_t^w(w_t;\mu,\Sigma) = -\Sigma^{-1} \tag{4.37}
$$

By Taylor expanding the terms multiplied by -1 in the objective function we get the following equations

$$\begin{cases} -\ln f_t^w(w_t) \approx -c_w - a_t^T \Delta w_t + \frac{1}{2}\Delta w_t^T A_t \Delta w_t \\ -\ln f_t^v(v_t) \approx -c_v - b_t^T \Delta v_t + \frac{1}{2}\Delta v_t^T B_t \Delta v_t \\ -\ln f_t^x(x_t) \approx -c_x - e_t^T \Delta x_t + \frac{1}{2}\Delta x_t^T E_t \Delta x_t \end{cases},$$

where the switch of the sign is to transform the maximization problem to a minimization problem. We can use the first and second order expansions of the natural logarithm of the pdf for the Student's t-distribution and the normal distribution to describe the vectors and matrices used in the quadratic Taylor approximations. Thus our assumption of the probability distribution of the variables affect the different Taylor approximations of the log-likelihood function of the variables w_t, v_t and x_t. Given a solution $\tilde{x}_t, \tilde{w}_t, \tilde{v}_t$, we can define

$$\begin{cases} \Delta w_t = w_t - \tilde{w}_t \\ \Delta v_t = v_t - \tilde{v}_t \\ \Delta x_t = x_t - \tilde{x}_t \end{cases}.$$

Finally, with Section 4.2 and Section 4.1.1 about the derived pricing function and derived state change function, we get the optimization problem

$$\min_{\Delta x, \Delta w, \Delta v} \sum_{t=1}^{T} \Big(c_t + a_t^T \Delta w_t + \frac{1}{2}\Delta w_t^T A_t \Delta w_t + b_t^T \Delta v_t + \frac{1}{2}\Delta v_t^T B_t \Delta v_t \tag{4.38}$$

$$+ e_t^T \Delta x_t + \frac{1}{2}\Delta x_t^T E_t \Delta x_t \Big) + c_0 + e_0^T \Delta x_0 + \frac{1}{2}\Delta x_0^T E_0 \Delta x_0, \tag{4.39}$$

$$\textbf{s.t.} \Delta x_t = F_t \Delta x_{t-1} + \Delta w_t + \delta_t^w, \tag{4.40}$$

$$0 = H_t \Delta x_t + \Delta v_t + \delta_t^v. \tag{4.41}$$

where we define $c_w + c_v + c_x = c_t, \ \forall t \neq 0$. Furthermore, assume inductively that for $t+1$ the value function can be written as

$$V_{t+1}(\Delta x_t) = d_{t+1} + q_{t+1}^T \Delta x_t + \frac{1}{2}\Delta x_t^T Q_{t+1} \Delta x_t, \tag{4.42}$$

where $V_{t+1}(\Delta x_t)$ represents the minimal cost from t to T given the state Δx_t. Also, d_t represents the accumulated costs independent of Δx_{t-1} and q_t represents the linear dependence on the initial state of x_{t-1}. Q_{t+1} is a symmetric matrix explaining the quadratic interaction between elements of Δx_t. So, we therefore get

$$V_t(\Delta x_{t-1}) = \min \sum_{\tau=t}^{T} \left(c_\tau + a_\tau^T \Delta w_\tau + \frac{1}{2}\Delta w_\tau^T A_\tau \Delta w_\tau + b_\tau^T \Delta v_\tau + \frac{1}{2}\Delta v_\tau^T B_\tau \Delta v_\tau \right) +$$

$$+ \sum_{\tau=t}^{T} \left(e_\tau^T \Delta x_\tau + \frac{1}{2}\Delta x_\tau^T E_\tau \Delta x_\tau \right), \tag{4.43}$$

$$\text{s.t.} \quad \Delta x_\tau = F_\tau \Delta x_{\tau-1} + \Delta w_\tau + \delta_\tau^w, \quad \tau = t, \ldots, T, \tag{4.44}$$

$$0 = H_\tau \Delta x_\tau + \Delta v_\tau + \delta_\tau^v, \quad \tau = t, \ldots, T. \tag{4.45}$$

Now, the value function can be divided into immediate and future cost.

Therefore, Equation (4.43) can be reformulated to

$$\begin{aligned} \min\ & c_t + a_t^T \Delta w_t + \frac{1}{2}\Delta w_t^T A_t \Delta w_t + b_t^T \Delta v_t + \frac{1}{2}\Delta v_t^T B_t \Delta v_t \\ & + e_t^T \Delta x_t + \frac{1}{2}\Delta x_t^T E_t \Delta x_t + V_{t+1}(\Delta x_t), \qquad (4.46) \\ \text{s.t.}\quad & \Delta x_t = F_t \Delta x_{t-1} + \Delta w_t + \delta_t^w, \qquad (4.47) \\ & 0 = H_t \Delta x_t + \Delta v_t + \delta_t^v. \qquad (4.48) \end{aligned}$$

Next, the noises Δw_t and Δv_t are defined by

$$\begin{cases} \Delta w_t = \Delta x_t - F_t \Delta x_{t-1} - \delta_t^w \\ \Delta v_t = -H_t \Delta x_t - \delta_t^v \end{cases}, \qquad (4.49)$$

where F_t and H_t are transition matrices for Δx_{t-1} and Δx_t, respectively, and δ_t^w and δ_t^v represent additional deviations.

The function f_t at time t is expressed as

$$\begin{aligned} f_t =\, & c_t + a_t^T \Delta w_t + \frac{1}{2}\Delta w_t^T A_t \Delta w_t + b_t^T \Delta v_t + \frac{1}{2}\Delta v_t^T B_t \Delta v_t \qquad (4.50) \\ & + e_t^T \Delta x_t + \frac{1}{2}\Delta x_t^T E_t \Delta x_t + V_{t+1}(\Delta x_t) \qquad (4.51) \\ =\, & c_t + a_t^T \Delta w_t + \frac{1}{2}\Delta w_t^T A_t \Delta w_t + b_t^T \Delta v_t + \frac{1}{2}\Delta v_t^T B_t \Delta v_t \qquad (4.52) \\ & + e_t^T \Delta x_t + \frac{1}{2}\Delta x_t^T E_t \Delta x_t + d_{t+1} + q_{t+1}^T \Delta x_t + \frac{1}{2}\Delta x_t^T Q_{t+1} \Delta x_t, \qquad (4.53) \end{aligned}$$

where c_t is a constant term, a_t and b_t are vectors defining the linear relationships with Δw_t and Δv_t, respectively, and A_t and B_t are symmetric matrices defining the quadratic relationships of Δw_t and Δv_t. The vector e_t and matrix E_t characterize the linear and quadratic terms of Δx_t. By using the expressions for Δw_t and Δv_t, we can expand the function f_t, where the terms in the function f_t are handled one at the time to ensure its accuracy. Then we find that, for $a_t^T \Delta w_t$

$$a_t^T \Delta w_t = a_t^T(\Delta x_t - F_t \Delta x_{t-1} - \delta_t^w) = a_t^T \Delta x_t - a_t F_t \Delta x_{t-1} - a_t \delta_t^w, \qquad (4.54)$$

and we can describe $\frac{1}{2}\Delta w_t^T A_t \Delta w_t$ as

$$\begin{aligned} \frac{1}{2}\Delta w_t^T A_t \Delta w_t &= \frac{1}{2}(\Delta x_t - F_t \Delta x_{t-1} - \delta_t^w)^T A_t (\Delta x_t - F_t \Delta x_{t-1} - \delta_t^w) = \qquad (4.55) \\ &= \frac{1}{2}(\Delta x_t^T A_t \Delta x_t - \Delta x_t^T A_t F_t \Delta x_{t-1} - \Delta x_t^T A_t \delta_t^w - \Delta x_{t-1}^T F_t^T A_t \Delta x_t + \\ &\quad + \Delta x_{t-1}^T F_t^T A_t F_t \Delta x_{t-1} + \Delta x_{t-1}^T F_t^T A_t \delta_t^w - (\delta_t^w)^T A_t^T \Delta x_t + \\ &\quad (\delta_t^w)^T A_t F_t \Delta x_{t-1} + (\delta_t^w)^T A_t \delta_t^w), \qquad (4.56) \end{aligned}$$

and $b_t^T \Delta v_t$ as

$$b_t^T \Delta v_t = b_t^T(-H_t \Delta x_t - \delta_t^v) = -b_t^T H_t \Delta x_t - b_t^T \delta_t^v, \tag{4.57}$$

and lastly, for $\frac{1}{2}\Delta v_t^T B_t \Delta v_t$ we get

$$\frac{1}{2}\Delta v_t^T B_t \Delta v_t = \frac{1}{2}(-H_t \Delta x_t - \delta_t^v)^T B_t(-H_t \Delta x_t - \delta_t^v) = \tag{4.58}$$

$$\frac{1}{2}(\Delta x_t^T H_t^T B_t H_t \Delta x_t + \Delta x_t^T H_t^T B_t \delta_t^v + (\delta_t^v)^T B_t H_t \Delta x_t + (\delta_t^v)^T B_t \delta_t^v). \tag{4.59}$$

The constant, linear and quadratic components are then identified, by defining new coefficients we get

$$\textit{Constant: } \bar{c}_t = c_t - a_t^T \delta_t^w + \frac{1}{2}(\delta_t^w)^T A_t \delta_t^w - b_t^T \delta_t^v + \frac{1}{2}(\delta_t^v)^T B_t \delta_t^v + d_{t+1}, \tag{4.60}$$

$$\textit{Linear: } \begin{cases} \Delta x_t : \bar{a}_t = a_t - H_t^T b_t + e_t + q_{t+1} - A_t \delta_t^w + H_t^T B_t \delta_t^v \\ \Delta x_{t-1} : \bar{r}_t = -F_t^T a_t + F_t^T A_t \delta_t^w \end{cases}, \tag{4.61}$$

$$\textit{Quadratic: } \begin{cases} \Delta x_t^T \bar{A}_t \Delta x_t : \bar{A}_t = A_t + E_t + Q_{t+1} + H_t^T B_t H_t \\ \Delta x_{t-1}^T \bar{R} \Delta x_{t-1} : \bar{R}_t = F_t^T A_t F_t \\ \Delta x_t \bar{P}_t \Delta x_{t-1} : \bar{P}_t = F_t^T A_t \end{cases}. \tag{4.62}$$

By adjusting the function f_t we can express it as

$$f_t = \bar{c}_t + \bar{a}_t^T \Delta x_t + \frac{1}{2}\Delta x_t^T \bar{A}_t \Delta x_t + \bar{r}_t^T \Delta x_{t-1} + \frac{1}{2}\Delta x_{t-1}^T \bar{R}_t \Delta x_{t-1} - \Delta x_{t-1}^T \bar{P}_t \Delta x_t. \tag{4.63}$$

We then minimize the function f_t, with respect to Δx_t and find the optimal solution through the stationary point, which can be derived from the derivative of the function,

$$\min_{\Delta x_t} f_t = \bar{c}_t + \bar{a}_t^T \Delta x_t + \frac{1}{2}\Delta x_t^T \bar{A}_t \Delta x_t + \bar{r}_t^T \Delta x_{t-1} + \frac{1}{2}\Delta x_{t-1}^T \bar{R}_t \Delta x_{t-1} - \Delta x_{t-1}^T \bar{P}_t \Delta x_t. \tag{4.64}$$

The derivative of f_t, with respect to Δx_t, is set to zero

$$\nabla_{\Delta x_t} f_t = \bar{a}_t + \bar{A}_t \Delta x_t - \bar{P}_t^T \Delta x_{t-1} = 0 \tag{4.65}$$

which results in

$$\bar{A}_t \Delta x_t = \bar{P}_t^T \Delta x_{t-1} - \bar{a}_t \Rightarrow \Delta x_t^* = \bar{A}_t^{-1}(\bar{P}_t^T \Delta x_{t-1} - \bar{a}_t) = \bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} - \bar{A}_t^{-1} \bar{a}_t, \quad (4.66)$$

and giving us the optimal solution Δx_t^*

$$\Delta x_t^* = \bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} - \bar{A}_t^{-1} \bar{a}_t, \quad (4.67)$$
$$(4.68)$$

Now, the optimal objective function value f^* is written as

$$\begin{aligned} f^* &= \bar{c}_t + \bar{a}_t^T \Delta x_t^* + \frac{1}{2}(\Delta x_t^*)^T \bar{A}_t \Delta x_t^* + \bar{r}_t^T \Delta x_{t-1} + \frac{1}{2} \Delta x_{t-1}^T \bar{R}_t \Delta x_{t-1} - \Delta x_{t-1}^T \bar{P}_t \Delta x_t^* \\ &= \bar{c}_t + \bar{a}_t^T \left(\bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} - \bar{A}_t^{-1} \bar{a}_t \right) + \frac{1}{2} \left(\bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} - \bar{A}_t^{-1} \bar{a}_t \right)^T \bar{A}_t \left(\bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} - \bar{A}_t^{-1} \bar{a}_t \right) \\ &\quad + \bar{r}_t^T \Delta x_{t-1} + \frac{1}{2} \Delta x_{t-1}^T \bar{R}_t \Delta x_{t-1} - \Delta x_{t-1}^T \bar{P}_t \left(\bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} - \bar{A}_t^{-1} \bar{a}_t \right). \end{aligned}$$

By expanding the expression we get,

$$\begin{aligned} &= \bar{c}_t + \bar{a}_t^T \bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} - \bar{a}_t^T \bar{A}_t^{-1} \bar{a}_t + \frac{1}{2} \left(\Delta x_{t-1}^T \bar{P}_t \bar{A}_t^{-T} - \bar{a}_t^T \bar{A}_t^{-T} \right) \bar{A}_t \left(\bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} - \bar{A}_t^{-1} \bar{a}_t \right) \\ &\quad + \bar{r}_t^T \Delta x_{t-1} + \frac{1}{2} \Delta x_{t-1}^T \bar{R}_t \Delta x_{t-1} - \Delta x_{t-1}^T \bar{P}_t \bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} + \Delta x_{t-1}^T \bar{P}_t \bar{A}_t^{-1} \bar{a}_t \\ &= \bar{c}_t + \bar{a}_t^T \bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} - \bar{a}_t^T \bar{A}_t^{-1} \bar{a}_t + \\ &\frac{1}{2} \left(\Delta x_{t-1}^T \bar{P}_t \bar{A}_t^{-T} \bar{P}_t^T \Delta x_{t-1} - \Delta x_{t-1}^T \bar{P}_t \bar{A}_t^{-T} \bar{a}_t - \bar{a}_t^T \bar{A}_t^{-T} \bar{P}_t^T \Delta x_{t-1} + \bar{a}_t^T \bar{A}_t^{-T} \bar{a}_t \right) \\ &\quad + \bar{r}_t^T \Delta x_{t-1} + \frac{1}{2} \Delta x_{t-1}^T \bar{R}_t \Delta x_{t-1} - \Delta x_{t-1}^T \bar{P}_t \bar{A}_t^{-1} \bar{P}_t^T \Delta x_{t-1} + \Delta x_{t-1}^T \bar{P}_t \bar{A}_t^{-1} \bar{a}_t \\ &= \bar{c}_t - \bar{a}_t^T \bar{A}_t^{-1} \bar{a}_t + \frac{1}{2} \bar{a}_t^T \bar{A}_t^{-T} \bar{a}_t + \left(\bar{r}_t^T + \bar{a}_t^T \bar{A}_t^T \bar{P}_t^T - \bar{a}_t^T \bar{A}_t^{-1} \bar{P}_t^T + \bar{a}_t^T \bar{A}_t^{-T} \bar{P}_t \right) \Delta x_{t-1} + \\ &\quad \frac{1}{2} \Delta x_{t-1}^T \left(\bar{r}_t - \bar{P}_t \bar{A}_t^{-1} \bar{P}_t^T \right) \Delta x_{t-1} \\ &= \bar{c}_t - \frac{1}{2} \bar{a}_t^T \bar{A}_t^{-1} \bar{a}_t + \left(\bar{r}_t + \bar{P}_t \bar{A}_t^{-1} \bar{a}_t \right)^T \Delta x_{t-1} + \frac{1}{2} \Delta x_{t-1}^T \left(\bar{R}_t - \bar{P}_t \bar{A}_t^{-1} \bar{P}_t^T \right) \Delta x_{t-1} \end{aligned}$$

We can now introduce new variables and further simplify by using the symmetry of A,

$$\begin{cases} d_t = \bar{c}_t - \frac{1}{2} \bar{a}_t^T \bar{A}_t^{-1} \bar{a}_t \\ q_t = \bar{r}_t + \bar{P}_t \bar{A}_t^{-1} \bar{a}_t \\ Q_t = \bar{R}_t - \bar{P}_t \bar{A}_t^{-1} \bar{P}_t^T \end{cases} \quad (4.69)$$

and thus rewrite the optimal objective function value as

$$f^* = d_t + q_t^T \Delta x_{t-1} + \frac{1}{2} \Delta x_{t-1}^T Q_t \Delta x_{t-1}. \quad (4.70)$$

Finally, by iterating backwards from the final time $t = T$ and using recursive updates for d_t, q_t, and Q_t, we minimize f_0 and determine the initial state. At the final time $t = T$, we have the objection function,

$$f_T = c_T + a_t^T \Delta w_T + \frac{1}{2}\Delta w_T^T A_T \Delta w_T + b_T^T \Delta v_T + \frac{1}{2}\Delta v_T^T B_T v_T + e_T^T \Delta x_T + \frac{1}{2}\Delta x_T^T E_T \Delta x_T, \tag{4.71}$$

and by using the optimal solution at time $t = T$

$$\Delta x_t^* = \bar{A}_t^{-1}\bar{P}_t^T \Delta x_{t-1} - \bar{A}_t^{-1}\bar{a}_t, \tag{4.72}$$
$$\tag{4.73}$$

by solving for arbitrary t and with respect to Δx_{t-1} we get

$$f_t = \bar{c}_t + \bar{r}_t^T \Delta x_{t-1} + \frac{1}{2}\Delta x_{t-1}^T \bar{R}_t \Delta x_{t-1} - \frac{1}{2}(\bar{a}_t - \bar{P}_t^T \Delta x_{t-1})^T \bar{A}_t^{-1}(\bar{a}_t - \bar{P}_t^T \Delta x_{t-1}). \tag{4.74}$$

Then we update the value function

$$V_{t-1}(\Delta x_t) = \begin{cases} d_t = \bar{c}_t - \frac{1}{2}\bar{a}_t^T \bar{A}_t^{-1}\bar{a}_t \\ q_t = \bar{r}_t + \bar{a}_t^T \bar{A}_t^T \bar{P}_t^T \\ Q_t = \bar{R}_t - \bar{P}_t \bar{A}_t^{-1}\bar{P}_t^T \end{cases} \tag{4.75}$$

and at t = 0, we have

$$\min f_0 = c_0 + e_0^T \Delta x_0 + \frac{1}{2}\Delta x_0^T E_0 \Delta x_0 + V_1(\Delta x_0). \tag{4.76}$$

By substituting the value function at t = 0, which is

$$V_1(\Delta x_0) = d_1 + q_1^T \Delta x_0 + \frac{1}{2}\Delta x_0^T Q_1 \Delta x_0 \tag{4.77}$$

and this results in

$$f_0 = c_0 + \bar{a_0}^T \Delta x_0 + \frac{1}{2}\Delta x_0 \bar{A}_0 \Delta x_0 \tag{4.78}$$

where,

$$\begin{cases} \bar{c}_0 = c_0 + d_1, \\ \bar{a}_0 = e_0 + q_1, \\ \bar{A}_0 = E_0 + Q_1. \end{cases}$$

To get the initial state, we use the derivative and set it equal to zero as following

$$\nabla_{\Delta x_0} f_0 = \bar{a_0}^T + \frac{1}{2}\bar{A}_0 \Delta x_0 = 0 \iff \Delta x_0^* = -\bar{A_0}^{-1}\bar{a_0}^T. \tag{4.79}$$

4.4 Computing the covariance matrix

Computing a meaningful value of a covariance matrix can be tricky. Here, we're going to assume that x is student's t distributed, v is student's t distributed and w is normal distributed. Assume $w \sim N(0, Q_t)$, then we get that the negative log-likelihood is a quadratic function. Now we can compute the covariance $P^w = A^{-1}$. In this case, we can use the EM algoritm to compute the value of P^w, seen in Section 6.

To compute the covariance for μ, we can make a local approximation $B_t = R_t^{-1}$, thus for each iteration k, we get $B_t = (R_t(\tilde{v}_t))^{-1}$, for each $\tilde{v}_t$ at each k. Now we treat $v_t \sim N(0, R_{t,k})$ in the forward and backward pass. By simply evaluating the hessian at the current point v_t at iteration k we get the inverse covariance in a local gaussian approximation, in other words $R_{t,k} = (B_t(\tilde{v}_t))^{-1}$ at each iteration k. We make a similar approximation for x by computing the hessian at $\tilde{x}_t$ at iteration k.

In conclusion:

- Init x, v, w
- Compute hessians as the local inverse of the covariance
  - Compute B_t at iteration k
  - Compute A_t at iteration k
  - Compute E_t at iteration k
- Update x, v, w and redo j

4.5 Overview of the Algorithm

Given the problem formulation we get the log-likelihood optimization problem

$$
\begin{aligned}
&\max \ln f_0^x(x_0) + \sum_{t=1}^{T} \left(\ln f_t^w(w_t) + \ln f_t^v(v_t) + \ln f_t^x(x_t) \right), \qquad (4.80)\\
&\text{s.t. } x_t = F_t x_{t-1} + w_t, \quad t = 1, \ldots, T,\\
&z_t = g_t(x_t^s) + I_t^z x_t^u + v_t, \quad t = 1, \ldots, T,
\end{aligned}
$$

By assuming that the distributions of the given random variables are differentiable to the second order we can make a Taylor expansion thereby getting an approximation of the objective function at iteration k. By making the inductive assumption we can show that we can do a backwards recursion using the prior state, i.e. Δx_{t+1} to get the optimal objective function at time $t = 0$ which, according to the derivations becomes $f^* = \bar{c_0} + \bar{a_0}^T \Delta x_0^T + \frac{1}{2} \Delta x_0 \bar{A}_0 \Delta x_0^T$.

To summarize:

1. **PCA:**
   We use PCA for dimensionality reduction and by this method we will be able to approximate the forward curve with just K risk factors. This can be compared to the size of $f_t \in \mathbb{R}^{n \times 1}$ where $n \approx 3650$. However the optimal solution of the Kalman filter will be limited by the eigenvectors Q_K from the PCA, and thus we will optimize these hyperparameters later.

2. **Riccati Recursion**

   - Initalize terminal conditions such that $V_{T+1}(\Delta x_T) = 0$ which implies the coefficients $d_{T+1} = 0$ $q_{T+1} = 0$ $Q_{T+1} = 0$
   - For t = T, . . . , 1, compute:

$$\begin{cases} d_t = \bar{c}_t - \frac{1}{2}\bar{a}_t^T \bar{A}_t^{-1}\bar{a}_t \\ q_t = \bar{r}_t + \bar{P}_t\bar{A}_t^{-1}\bar{a}_t \\ Q_t = \bar{R}_t - \bar{P}_t\bar{A}_t^{-1}\bar{P}_t^T \end{cases} \tag{4.81}$$

   however, note that the coefficients are

$$\bar{c}_t = c_t - a_t^T\delta_t^w + \frac{1}{2}(\delta_t^w)^T A_t\delta_t^w - b_t^T\delta_t^v + \frac{1}{2}(\delta_t^v)^T B_t\delta_t^v + d_{t+1}, \tag{4.82}$$

$$\bar{a}_t = a_t - A_t\delta_t^w - H_t^T b_t + H_t^T B_t\delta_t^v + e_t + q_{t+1}, \tag{4.83}$$

$$\bar{r}_t = -F_t^T a_t + F_t^T A_t\delta_t^w, \tag{4.84}$$

$$\bar{A}_t = A_t + E_t + Q_{t+1} + H_t^T B_t H_t, \tag{4.85}$$

$$\bar{R}_t = F_t^T A_t F_t, \tag{4.86}$$

$$\bar{P}_t = A_t F_t. \tag{4.87}$$

   And finally we get

   $f_0 = c_0 + e_0^T + \frac{1}{2}\Delta x_0^T E_0\Delta x_0 + V_1(\Delta x_0) = c_0 + e_0^T + \frac{1}{2}\Delta x_0^T E_0\Delta x_0 + d_1 + q_1^T\Delta x_0 + \frac{1}{2}\Delta x_0^T Q_1\Delta x_0$

   - Initialize: $\Delta x_0^* = -\bar{A}_0^{-1}\bar{a}_0$, where $\bar{A}_0 = E_0 + Q_1$ and $\bar{a}_0 = e_0 + q_1$
   - Compute: $x_0^* = \bar{x} + \Delta x_0^*$

   - For $t = 1, \ldots, T$
     - Recursively compute: $\Delta x_t^* = \bar{A}_t^{-1}\bar{P}_t^T - \bar{A}_t^{-1}\bar{a}_t$
     - Update the state estimate: $x_t^* = \bar{x}_t + \lambda\Delta x_t^*$, where λ is the step size.
     - Compute and update:

$$\begin{aligned} \Delta w_t^* &= \Delta x_t^* - F_t\Delta x_{t-1}^* - \delta_t^w, \\ \Delta v_t^* &= -H_t\Delta x_t^* - \delta_t^v, \\ w_t^* &= \bar{w} + \Delta w_t^*, \\ v_t^* &= \bar{v}_t + \Delta v_t^*. \end{aligned}$$

     - Recompute the Taylor expansions around the updated estimate
     - Update the coefficients

3. Repeat

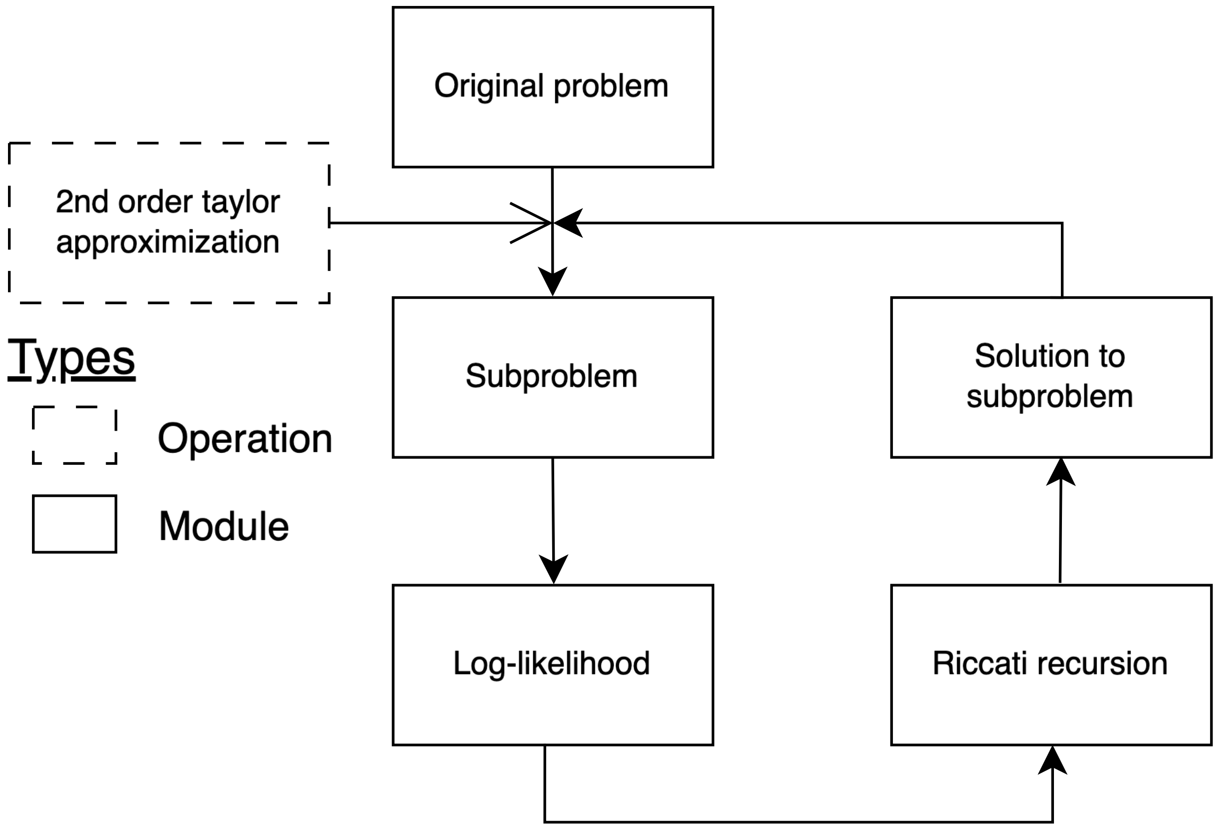

Figure 1: Overview of the psuedo-code for the solver.

5 Automatic Differentation

This section will describe AD, its two distinct modes, and how we implement this technique in optimizing calculations of PCA components.

5.1 Definition

AD is a collection of techniques that calculates numerical values for derivatives of a function with the same precision and processing capacity as determining the same function's values. AD calculates the derivatives by repeating the commonly known chain rule but applied to floating point numerical values (Bartholomew-Biggs et al. 2000).

5.2 Different Modes

There are different modes of AD that differ in efficiency depending on the dimensions of the selected function's input and output spaces (Baydin et al. 2018).

5.2.1 Forward Accumulation

Forward accumulation or forward mode is efficient when

$$f : \mathbb{R} \to \mathbb{R}^m,$$

we only have a one-dimensional input space, we can calculate its respective partial derivatives with only one forward pass (Baydin et al. 2018).

5.2.2 Reverse Accumulation

On the other hand, reverse accumulation and reverse mode, provides less costly computation with function with a large input space with respect to the output space, and thus when

$$f : \mathbb{R}^n \to \mathbb{R},$$

because this requires only one iteration for the reverse mode. In contrast, with a forward mode this would require $\mathcal{O}(n)$ computations (Baydin et al. 2018).

5.3 Implementation

Given the different properties of AD's modes, we will implement reverse accumulation. The reason being that the objective function of the optimization problem is of the form:

$$f : \mathbb{R}^n \to \mathbb{R}.$$

The implementation will be carried out in the programming language Julia with the package JuliaDiff/Zygote.jl.

6 Expectation Maximization

This section will delve into EM, its definition, and how we utilizes this method in parameter estimation of our noise and state distributions.

6.1 Definition

EM is a method that is used to compute maximum-likelihood estimations iteratively. The method's name is given by the fact that each step in the associated algorithm entails one expectation iteration and a subsequent maximization iteration (Dempster, Laird, and Rubin 1977).

6.2 Expectation Step (E-Step)

During the expectation step of the algorithm, we use the forward pass of the modified Kalman-filter to estimate the next state. This is followed by a smoothing done by the backward pass to refine the estimate over the entire sequence, refining the noises w_t and v_t (Einicke et al. n.d.).

6.3 Maximization Step (M-Step)

In the maximization step we update the parameters of Equation (4.26) to maximize the expected log-likelihood based on these refined state estimates (Einicke et al. n.d.).

6.4 Implementation

We will implement the EM method to update the parameters to w_t and v_t that is found in our optimization model, representing normally distributed and student-t distributed noise variables, respectively. Given a multivariate normal distribution, we can estimate an updated measurement noise covariance matrix by first implementing an expectation step, derived from the maximum likelihood calculation

$$\hat{R} = \frac{1}{T}\sum_{t=1}^{T} \left(z_t - \hat{z}_t\right)\left(z_t - \hat{z}_t\right)^T, \text{ where } \hat{z}_t = g_t(x_t^s) + I_t^z x_t^u, \tag{6.1}$$

and the estimation of the updated process noise covariance matrix is then

$$\hat{Q} = \frac{1}{T}\sum_{t=0}^{T-1} \left(\hat{x}_{t+1} - F_t\hat{x}_t\right)\left(\hat{x}_{t+1} - F_t\hat{x}_t\right)^T. \tag{6.2}$$

This will be implemented on the reverse-pass of the optimization model (Einicke et al. n.d.).

7 Interpolation Methods

This chapter contains various interpolation methods that will be used to plot the forward rates over time. Each method will generate a system of equations, which will be solved using the Newton-Raphson method and Newton's optimization technique. This section is based on Hagan and West (2006) and it will provide a detailed explanation of the mathematical principles of the interpolation methods. All calculation will be performed in MATLAB, using data collected from Eikon.

7.1 Data

Data sourced from Refinitiv Eikon will be utilized as discrete and fixed points to construct the interest rate yield curves that will be interpolated.

7.2 Linear on Discount Factors

Given $\tau_1, \tau_2, \ldots, \tau_n$ and $r_1, r_2, \ldots, r_n$, where $r_i := r(\tau_i)$, which represents a tenor and the corresponding yield. For a τ which is not equal to any of the τ_i, we will estimate the function $r(\tau)$ using only r and r_i. Simple interpolation methods can estimate forward rates by utilizing either discount factors d_i or zero rates r_i. Additionally, taking the logarithm of these values provides four distinct approaches in total.

Let $d(\tau) = \exp(-r(\tau)\tau)$ represent the discount function, where d_i and d_{i+1} are the discount factors at times i and $i+1$, respectively. Using this method, we obtain the following:

The discount factor can be expressed as

$$d(\tau) = \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} d_{i+1} + \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} d_i. \tag{7.1}$$

From this, the following expression for $r(\tau)$ is derived

$$r(\tau) = -\frac{1}{\tau} \ln\left(\frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} d_{i+1} + \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} d_i \right), \tag{7.2}$$

when we discretize r for each day k, we obtain

$$r_k = -\frac{1}{\tau_k} \ln\left(\frac{\tau_k - \tau_i}{\tau_{i+1} - \tau_i} d_{i+1} + \frac{\tau_{i+1} - \tau_k}{\tau_{i+1} - \tau_i} d_i \right). \tag{7.3}$$

The forward rate function is defined as $f(\tau) = \frac{d}{d\tau} r(\tau)\tau$. After further calculations, we arrive at the following expression for the forward rate

$$f(\tau) = \frac{d_i - d_{i+1}}{(\tau - \tau_i)d_{i+1} + (\tau_{i+1} - \tau)d_i}, \tag{7.4}$$

discretizing f, we have

$$f_k = \frac{r_{k+1}\tau_{k+1} - r_k\tau_k}{\tau_{k+1} - \tau_k} = \frac{\ln\left(\frac{\tau_k - \tau_i}{\tau_{i+1} - \tau_i} d_{i+1} + \frac{\tau_{i+1} - \tau_k}{\tau_{i+1} - \tau_i} d_i \right) - \ln\left(\frac{\tau_{k+1} - \tau_i}{\tau_{i+1} - \tau_i} d_{i+1} + \frac{\tau_{i+1} - \tau_{k+1}}{\tau_{i+1} - \tau_i} d_i \right)}{\tau_{k+1} - \tau_k} \tag{7.5}$$

(Hagan and West n.d.).

7.3 Linear on Spot Rates

This method utilizes the rates at times i and $i+1$ to interpolate the values in $r(\tau)$. Specifically, this interpolation results in a weighted average of the rates at these two time points, expressed as

$$r(\tau) = \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} r_{i+1} + \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} r_i, \tag{7.6}$$

when we discretize r for each day k, we obtain the following expression

$$r_k = \frac{\tau_k - \tau_i}{\tau_{i+1} - \tau_i} r_{i+1} + \frac{\tau_{i+1} - \tau_k}{\tau_{i+1} - \tau_i} r_i. \tag{7.7}$$

The forward rate function is defined as

$$f(\tau) = \frac{d}{d\tau} r(\tau)\tau, \tag{7.8}$$

this leads to the following expression for the forward rate

$$f(\tau) = \frac{2\tau - \tau_i}{\tau_{i+1} - \tau_i} r_{i+1} + \frac{\tau_{i+1} - 2\tau}{\tau_{i+1} - \tau_i} r_i, \tag{7.9}$$

discretizing this gives us

$$f_k = \frac{r_{k+1}\tau_{k+1} - r_k\tau_k}{\tau_{k+1} - \tau_k} =$$
$$\tau_{k+1}\left(\frac{\tau_{k+1} - \tau_i}{\tau_{i+1} - \tau_i} r_{i+1} + \frac{\tau_{i+1} - \tau_{k+1}}{\tau_{i+1} - \tau_i} r_i\right) \quad - \tau_k\left(\frac{\tau_k - \tau_i}{\tau_{i+1} - \tau_i} r_{i+1} + \frac{\tau_{i+1} - \tau_k}{\tau_{i+1} - \tau_i} r_i\right)\frac{\tau_{k+1} - \tau_k}{\tau_{k+1} - \tau_k}.$$

The same logic applies as before, indicating that the rates are not continuous (Hagan and West n.d.).

7.4 Raw Interpolation

Raw interpolation employs a linear method based on the logarithm of discount factors. This approach produces forward curves that remain consistently constant across segments. A notable characteristic of this method is its stability in comparison to earlier techniques. For segmented linear interpolation to perform effectively, it is essential that each segment retains a uniform area between its nodes. In this context, the area refers to the integral or accumulated value between two points on the curve.

Additionally, this method serves as an excellent baseline for comparison, given its stability and its tendency to provide a good approximation. The logarithm of discount factors is defined as

$$\ln d(\tau) = \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} \ln d_i + \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} \ln d_{i+1}, \tag{7.10}$$

this can also be expressed as

$$r(\tau) = \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} \frac{\tau_{i+1}}{\tau} r_{i+1} + \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} \frac{\tau_i}{\tau} r_i. \tag{7.11}$$

The discretized r for each day k gives the following expression:

$$r_k(\tau) = \frac{\tau_k - \tau_i}{\tau_{i+1} - \tau_i} \frac{\tau_{i+1}}{\tau_k} r_{i+1} + \frac{\tau_{i+1} - \tau_k}{\tau_{i+1} - \tau_i} \frac{\tau_i}{\tau_k} r_i. \tag{7.12}$$

By multiplying by τ and taking the derivative with respect to τ, we derive the forward rate function

$$f(\tau) = \frac{\tau_{i+1}}{\tau_{i+1} - \tau_i} r_{i+1} - \frac{\tau_i}{\tau_{i+1} - \tau_i} r_i. \tag{7.13}$$

The forward rate f_k is defined as

$$f_k = \frac{r_{k+1}\tau_{k+1} - r_k\tau_k}{\tau_{k+1} - \tau_k}, \tag{7.14}$$

Substituting r_k into this expression it gives us the forward rate f_k

$$f_k = \frac{\left(\frac{\tau_{k+1} - \tau_i}{\tau_{i+1} - \tau_i} \frac{\tau_{i+1}}{\tau_{k+1}} r_{i+1} + \frac{\tau_{i+1} - \tau_{k+1}}{\tau_{i+1} - \tau_i} \frac{\tau_i}{\tau_{k+1}} r_i \right) \tau_{k+1} - \left(\frac{\tau_k - \tau_i}{\tau_{i+1} - \tau_i} \frac{\tau_{i+1}}{\tau_k} r_{i+1} + \frac{\tau_{i+1} - \tau_k}{\tau_{i+1} - \tau_i} \frac{\tau_i}{\tau_k} r_i \right) \tau_k}{\tau_{k+1} - \tau_k} \tag{7.15}$$

Hagan and West n.d.

7.5 Linear on the Logarithm of Rates

Given the expression

$$\ln r(\tau) = \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} \ln r_{i+1} + \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} \ln r_i, \tag{7.16}$$

rewriting the equation as

$$r(\tau) = r_{i+1}^{\frac{\tau - \tau_i}{\tau_{i+1} - \tau_i}} r_i^{\frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i}}, \tag{7.17}$$

then, the discretized version for each day k is

$$r_k = r_{i+1}^{\frac{\tau_k - \tau_i}{\tau_{i+1} - \tau_i}} r_i^{\frac{\tau_{i+1} - \tau_k}{\tau_{i+1} - \tau_i}}. \tag{7.18}$$

The forward rate f_k is defined as

$$f_k = \frac{r_{k+1}\tau_{k+1} - r_k\tau_k}{\tau_{k+1} - \tau_k}, \tag{7.19}$$

substituting r_k into the forward rate

$$f_k = \frac{r_{i+1}^{\frac{\tau_{k+1}-\tau_i}{\tau_{i+1}-\tau_i}} r_i^{\frac{\tau_{i+1}-\tau_{k+1}}{\tau_{i+1}-\tau_i}} \tau_{k+1} - r_{i+1}^{\frac{\tau_k-\tau_i}{\tau_{i+1}-\tau_i}} r_i^{\frac{\tau_{i+1}-\tau_k}{\tau_{i+1}-\tau_i}} \tau_k}{\tau_{k+1}-\tau_k}. \tag{7.20}$$

This method does not guarantee positive forward rates (Hagan and West n.d.).

7.6 Newton-Raphsons Method

For equation systems with k variables and k number of functions as we have in our case, the iteration transforms into:

$$x_{t+1} = x_t + \gamma J_F(x_t)^{-1} F(\mathbf{x}) \tag{7.21}$$

where J_F represents the Jacobian of the function f at x_t, γ represents the step size and F is a vector comprised of the following functions:

$$F(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_k(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2, \ldots, x_k) \\ f_2(x_1, x_2, \ldots, x_k) \\ \vdots \\ f_k(x_1, x_2, \ldots, x_k) \end{bmatrix}, \tag{7.22}$$

which will be used to iteratively find the solution (Akram and Ann 2015).

The step length in each step is decided through the following algorithm. If the error after a step of $\gamma = 1$ is taken the error is evaluated, should the error be larger than before the step the usage of interval halving of γ will be performed iteratively until the error after the step is smaller than before. A pseudo-code algorithm for this can be written as follows:

Set $\gamma = 1$
While $\|F(x_{t+1})\| > \|F(x_t)\|$:
  $\gamma = \frac{\gamma}{2}$
Return γ

7.6.1 Jacobian Derivatives for Interpolation Methods in Newton-Raphson

For y_{OIS} defined as:

$$y_{\text{OIS}} = \frac{P_{t,T_0} - P_{t,T_K}}{\sum_{i=1}^{K} \Delta T_i P_{t,T_i}} \tag{7.23}$$

we compute the Jacobian by differentiating with respect to the discount factors P_{t,T_i}, incorporating the inner derivative vector:

$$\frac{\partial d(\tau)}{\partial \mathbf{d}} = \left[\frac{\partial d(\tau)}{\partial d_i}, \frac{\partial d(\tau)}{\partial d_{i+1}}\right]$$

as defined in Section 7.7.

Case 1: $i = 0$ For $i = 0$, differentiating with respect to P_{t,T_0}:

$$\frac{\partial y_{\text{OIS}}}{\partial P_{t,T_0}} = \frac{1}{\sum_{i=1}^{K} \Delta T_i P_{t,T_i}} \cdot \frac{\partial d(\tau)}{\partial \mathbf{d}} \tag{7.24}$$

Case 2: $i = 1, 2, \ldots, K-1$ For $i = 1, \ldots, K-1$, differentiating with respect to P_{t,T_i}:

$$\frac{\partial y_{\text{OIS}}}{\partial P_{t,T_i}} = -\frac{\Delta T_i (P_{t,T_0} - P_{t,T_K})}{\left(\sum_{i=1}^{K} \Delta T_i P_{t,T_i}\right)^2} \cdot \frac{\partial d(\tau)}{\partial \mathbf{d}} \tag{7.25}$$

Case 3: $i = K$ Finally, for $i = K$, differentiating with respect to P_{t,T_K}:

$$\frac{\partial y_{\text{OIS}}}{\partial P_{t,T_K}} = -\left(\frac{1}{\sum_{i=1}^{K} \Delta T_i P_{t,T_i}} + \frac{\Delta T_K \cdot (P_{t,T_0} - P_{t,T_K})}{\left(\sum_{i=1}^{K} \Delta T_i P_{t,T_i}\right)^2}\right) \cdot \frac{\partial d(\tau)}{\partial \mathbf{d}} \tag{7.26}$$

7.7 Interpolation Methods and Inner Derivatives

This section expands the inner derivatives of $\frac{\partial d(\tau)}{\partial d_i}$ and $\frac{\partial d(\tau)}{\partial d_{i+1}}$ as referred to under Section 7.6.1 for each chosen interpolation method.

7.7.1 Linear on Discount Factors

For linear interpolation on discount factors:

$$d(\tau) = \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} d_{i+1} + \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} d_i \tag{7.27}$$

Differentiating with respect to d_i and d_{i+1}:

$$\frac{\partial d(\tau)}{\partial d_i} = \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i}, \quad \frac{\partial d(\tau)}{\partial d_{i+1}} = \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} \tag{7.28}$$

7.7.2 Linear on Spot Rates

The spot rates r_i are related to the discount factors via $r_i = -\frac{\ln(d_i)}{\tau_i}$. Substituting this into the interpolation formula, the transformed discount factor becomes:

$$d(\tau) = \exp\left(-\tau\left(\frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} \cdot \left(-\frac{\ln(d_{i+1})}{\tau_{i+1}}\right) + \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} \cdot \left(-\frac{\ln(d_i)}{\tau_i}\right)\right)\right) \tag{7.29}$$

Differentiating with respect to d_i and d_{i+1}:

$$\frac{\partial d(\tau)}{\partial d_i} = \tau \cdot d(\tau) \cdot \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} \cdot \frac{1}{\tau_i d_i}, \quad \frac{\partial d(\tau)}{\partial d_{i+1}} = \tau \cdot d(\tau) \cdot \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} \cdot \frac{1}{\tau_{i+1} d_{i+1}} \tag{7.30}$$

7.7.3 Raw Interpolation

For raw interpolation, the logarithmic representation of the discount factor is:

$$\ln d(\tau) = \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} \ln d_i + \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} \ln d_{i+1} \tag{7.31}$$

Thus:

$$d(\tau) = \exp\left(\frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} \ln d_i + \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} \ln d_{i+1}\right) \tag{7.32}$$

Differentiating with respect to d_i and d_{i+1}:

$$\frac{\partial d(\tau)}{\partial d_i} = d(\tau) \cdot \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} \cdot \frac{1}{d_i}, \quad \frac{\partial d(\tau)}{\partial d_{i+1}} = d(\tau) \cdot \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} \cdot \frac{1}{d_{i+1}} \tag{7.33}$$

7.7.4 Linear on the Logarithm of Rates

The logarithm of rates is interpolated as:

$$r(\tau) = r_{i+1}^{\frac{\tau - \tau_i}{\tau_{i+1} - \tau_i}} r_i^{\frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i}} \tag{7.34}$$

Substituting $r_i = -\frac{\ln(d_i)}{\tau_i}$,

$$r(\tau) = \left(-\frac{\ln(d_{i+1})}{\tau_{i+1}}\right)^{\frac{\tau - \tau_i}{\tau_{i+1} - \tau_i}} \cdot \left(-\frac{\ln(d_i)}{\tau_i}\right)^{\frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i}}. \tag{7.35}$$

Also, the discount factor is given by

$$d(\tau) = \exp\left(-\tau \cdot r(\tau)\right). \tag{7.36}$$

Differentiating with respect to d_i and d_{i+1}:

$$\begin{aligned} \frac{\partial d(\tau)}{\partial d_i} &= \tau \cdot d(\tau) \cdot \frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} \cdot \frac{1}{\tau_i d_i} \cdot \left(-\frac{\ln(d_i)}{\tau_i}\right)^{\frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i} - 1}, \\ \frac{\partial d(\tau)}{\partial d_{i+1}} &= \tau \cdot d(\tau) \cdot \frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} \cdot \frac{1}{\tau_{i+1} d_{i+1}} \cdot \left(-\frac{\ln(d_{i+1})}{\tau_{i+1}}\right)^{\frac{\tau - \tau_i}{\tau_{i+1} - \tau_i} - 1}. \end{aligned} \tag{7.37}$$

8 Model Evaluation, Testing and Validation

This section describes how testing and validation of the model will be conducted. It is essential to confirm that the model performs accurately across different data sets and is reliable. Furthermore, validating that the model achieves improved accuracy compared to other methods, such as interpolation techniques, is a vital objective, and the planned test for this is also highlighted below.

8.1 KKT Test

The KKT conditions are first-order derivative tests that provide necessary conditions for a solution to be optimal in a optimization problem. These conditions do not guarantee that a solution to the problem is optimal, but they have to be fulfilled for a optimal solution. They are sufficient if the objective function and non-linear constraints are convex and the linear constraints are linear. The KKT conditions consist of four parts, stationarity, primal feasibility, dual feasibility, and complementary slackness. Complementary slackness will not be tested since our constraints must fulfill strict equality (Wikipedia 2024).

By the problem $\min_x \; f(x)$ subject to $Ax = a$ and $Bx = b$, we can set the Lagrangian as $L = f(x) - y_a^T(Ax - a) \; - y_b^T(Bx - b)$ (Blomvall 2024a). With the objective function

$$
\begin{aligned}
f(\Delta x_t) = & \sum_{t=1}^{T} \left(c_t + a_t^T \Delta w_t + \frac{1}{2} \Delta w_t^T A_t \Delta w_t + b_t^T \Delta v_t + \frac{1}{2} \Delta v_t^T B_t \Delta v_t + e_t^T \Delta x_t + \frac{1}{2} \Delta x_t^T E_t \Delta x_t \right) \\
& + c_0 + e_0^T \Delta x_0 + \frac{1}{2} \Delta x_0^T E_0 \Delta x_0, \qquad (8.1)
\end{aligned}
$$

it results in the Lagrangian

$$
\begin{aligned}
L = & \sum_{t=1}^{T} \left(c_t + a_t^T \Delta w_t + \frac{1}{2} \Delta w_t^T A_t \Delta w_t + b_t^T \Delta v_t + \frac{1}{2} \Delta v_t^T B_t \Delta v_t + e_t^T \Delta x_t + \frac{1}{2} \Delta x_t^T E_t \Delta x_t \right) \\
& + c_0 + e_0^T \Delta x_0 + \frac{1}{2} \Delta x_0^T E_0 \Delta x_0 + \sum_{t=1}^{T} \left[\lambda_t^T \left(\Delta x_t - F_t \Delta x_{t-1} - \Delta w_t - \delta_t^w \right) + \mu_t^T \left(H_t \Delta x_t + \Delta v_t + \delta_t^v \right) \right] \\
= & f(\Delta x_t) + \sum_{t=1}^{T} \left[\lambda_t^T \left(\Delta x_t - F_t \Delta x_{t-1} - \Delta w_t - \delta_t^w \right) + \mu_t^T \left(H_t \Delta x_t + \Delta v_t + \delta_t^v \right) \right]. \qquad (8.2)
\end{aligned}
$$

Now, we want to solve for the Lagrangian multipliers λ_t^T and μ_t^T. This will address the primal feasibility conditions. We begin by using the derivative of the Lagrangian with respect to Δw_t and set it to zero.

$$
\begin{aligned}
\nabla_{\Delta w_t} L = a_t + A_t \Delta w_t - \lambda_t = 0 \Rightarrow \lambda_t = a_t + A_t \Delta w_t, \qquad (8.3) \\
\iff \Delta w_t = \Delta x_t - F_t \Delta x_{t-1} - \delta_t^w \qquad (8.4)
\end{aligned}
$$

which gives us a solution for λ_t^T.

To solve for μ_t^T, we instead use the derivative with respect to Δv_t and set to zero

$$\nabla_{\Delta v_t} L = b_t + B_t \Delta v_t + \mu_t = 0 \Rightarrow \mu_t = -(b_t + B_t \Delta v_t), \tag{8.5}$$

$$\iff \Delta v_t = -H_t \Delta x_t - \delta_t^v \tag{8.6}$$

which gives us a solution for μ_t^T.

Now, to address the dual feasibility, we use the derivative with respect to Δx_t and set to zero

$$\nabla_{\Delta x_t} L = e_t + E_t \Delta x_t + \lambda_t + H_t^T \mu_t = 0, \quad \text{for } t = 1, \ldots, T, \tag{8.7}$$

$$\nabla_{\Delta x_0} L = = e_0 + E_0 \Delta x_0, \quad \text{for } t = 0. \tag{8.8}$$

8.2 Linear Kalman Test

As the derivations for the objective functions are based on reasoning similar to a regular kalman filtration, we are going to implement a standard linear kalman filtration as a confirmation of the derivations. Consider

$$x_k = F x_{k-1} + w_k, \tag{8.9}$$

$$z_k = H x_k + v_k, \tag{8.10}$$

$$\hat{x} k|k-1 = F \hat{x} k-1|k-1, \tag{8.11}$$

$$P_{k|k-1} = F P_{k-1|k-1} F^\top + Q, \tag{8.12}$$

with the same definitions as stated in section 4.3. Then, by implementing a standard Kalman filter and compute the necessary parts so that we can do the prediction and update steps

$$y_k = z_k - H \hat{x}_{k|k-1}, \tag{8.13}$$

$$S_k = H P_{k|k-1} H^\top + R, \tag{8.14}$$

$$K_k = P_{k|k-1} H^\top S_k^{-1}, \tag{8.15}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k y_k, \tag{8.16}$$

$$P_{k|k} = (I - K_k H) P_{k|k-1}, \tag{8.17}$$

which we can compare with the model as we derived earlier. We compare our model to the linear Kalman filter by simply solving a arbitrary linear problem with the linear Kalman filter and our model. The reasoning behind this test is that the second order terms of the linear problem will be zero, and thus we should receive the same optimal solution for our model and for the linear Kalman model. This will allow us to test our implementation and make sure that it contains no errors in the first order terms.

8.3 Out-of-Sample Statistical Test

To determine if a model achieves better accuracy than an alternative model, statistical test based on out-of-sample observations will be used. Due to the high number of parameters used in the model presented above, testing with all data is less relevant, as parameter penalization may make the model appear less effective by emphasizing the impact of having many parameters. By using out-of-sample data to test the model's optimal performance, the influence of the large number of parameters is minimized (Blomvall 2024a). Blomvall (2021) outlines a statistical test to compare two models with this approach.

Firstly, the difference in log-values between model 1 and the alternative model, model 2, is calculated as

$$\bar{d}_i = \ln \bar{f}_{i,1}(\bar{z}_i, \hat{\bar{z}}_i) - \ln \bar{f}_{i,2}(\bar{z}_i, \hat{\bar{z}}_i), \tag{8.18}$$

where I is the set of all instruments, $\hat{\bar{z}}_i$ is the vector of the calculated quotes from model j and $\bar{z}_i$ is the actual retrieved quotes. The average log-values difference across all time steps is given by

$$d_i = \frac{1}{T}\sum_{t=1}^{T} \bar{d}_i, \tag{8.19}$$

where $\bar{d}_i = (d_{i,1}, ..., d_{i,T})$ The standard error s of this average is

$$s = \frac{\sigma}{\sqrt{T}}, \tag{8.20}$$

where $\sigma^2 = \frac{1}{T-1}\sum_{t=1}^{T}(d_{i,t} - d_i)^2$ is the sample variance of the differences d_i.
Under the assumption that $d_{i,t}$ values are independent and that T is large enough, the Central Limit Theorem suggests that the test statistic $\frac{d_{i,t}}{s}$ follows an approximate standard normal distribution, i.e.

$$\frac{d_{i,t}}{s} \sim N(0,1), \tag{8.21}$$

for all t. We can use this to determine if model 1 is statistically better than model 2 at a given confidence level (Blomvall 2024a). Let $F_N(\omega)$ denote the cumulative distribution function for a standard normal variable. Then, it holds that

$$\frac{d_i}{s} \geq F_N^{-1}(\alpha). \tag{8.22}$$

So, by choosing α arbitrarily close to 1 and confirming that equation (8.22) holds, one can conclude that model 1 is better than the alternative model.

8.3.1 Squared Error

One approach is to calculate out-of-sample estimation errors of market quotes from each model. Thus we can compute the probability that one model would have smaller estimation errors of the market quotes, which will indicate that the model has greater performance. This method is suitable since we can evaluate two models with the same out-of-sample data, the market quotes, and since we can calculate the quotes $\hat{z}_{i,t}$ of a instrument i for the time step t using the forward curve f_t. This quote will be compared with the actual retrieved quote z_i.

For each time step $t = 1, \dots, T$, the error function vector $\bar{f}_{i,j} = (f_{i,j,1}, ..., f_{i,j,T})$ is computed, for model j and for out-of-sample instrument $i \in I$ as the squared estimation error

$$\bar{f}_{i,j}(\bar{z}_i, \hat{\bar{z}}_i) = (\bar{z}_i - \hat{\bar{z}}_i)^2, \tag{8.23}$$

where $\bar{z}_i$ and $\hat{\bar{z}}_i$ is defined as in Section 8.3. The resulting $\bar{f}_{i,j}$ is then used in the Equation 8.18 for two different models.

8.3.2 Log-likelihood Test

Another approach we can have for the statistical test, to compare two different models, is by calculating log-likelihood values of each model and compare log-likelihood value for each time step for the two models on out-of-sample data. Doing so for the pricing function, z, described in Section 4.1, that is

$$z_t = g_t(x_t^s) + I_t^u x_t^u + v_t, \quad t = 1, \dots, T, \tag{8.24}$$

where $g_t(x_t^s)$ is defined in Equation 4.5. This results in a random variable, z, that does not have a specific distribution, or at least not without many assumptions regarding the other terms in the equation. Therefore, when calculating the log-likelihood function for z convolution will be used. The probability density function is now obtained as

$$f_{g_t + I_t^z x_t^u}(w) = \int_{-\infty}^{\infty} f_{g_t}(g_t) f_{I_t^z x_t^u}(w - g)\, dg, \tag{8.25}$$

$$f_{z_t}(z) = \int_{-\infty}^{\infty} f_{g_t + I_t^z x_t^u}(w) f_{v_t}(z - w)\, dw, \tag{8.26}$$

where this holds since $g(x_t^s)$, x_t^u and v_t are independent random variables. Nevertheless, this approach could be computationally demanding, particularly due to the dimensionality of for example v_t, which corresponds to the number of instruments and can be significantly large. However, when the log-likelihood function is computed, different models can again be evaluated using the method in Section 8.3.

8.3.3 Analytical Approximation of the Log-Likelihood Function

As mentioned before, the distribution of z does not have a closed form without specific assumptions regarding the terms in

$$z_t = g_t(x_t^s) + I_t^u x_t^u + v_t, \quad t = 1, \ldots, T. \tag{8.27}$$

But to illustrate how the likelihood function could be derived analytically the following assumptions are made:

- $x_t^s \sim \mathcal{N}(\mu_s, \Sigma_s)$, so that x_t^s follows a multivariate normal distribution,
- $x_t^u \sim \mathcal{N}(\mu_u, \Sigma_u)$, so that x_t^u also follows a multivariate normal distribution,
- $v_t \sim \mathcal{N}(0, \Sigma_v)$, the equivalence of Gaussian noise.

To handle the non-linear term $g_t(x_t^s)$, a first-order Taylor expansion is used to approximate it around its mean, μ_s, and the following is obtained

$$g_t(x_t^s) \approx g_t(\mu_s) + \nabla_{x_t^s} g_t(\mu_s)(x_t^s - \mu_s), \tag{8.28}$$

where $\nabla_{x_t^s} g_t(\mu_s)$ denotes the Jacobian of $g_t(x_t^s)$ evaluated at μ_s. This linearization transforms $g_t(x_t^s)$ into a linear function of x_t^s, keeping the Gaussian nature of x_t^s.

Given these approximations, z_t becomes a linear combination of independent normal random variables, meaning that

$$z_t \sim \mathcal{N}(\mu_z, \Sigma_z), \tag{8.29}$$

where

$$\begin{cases} \mu_z &= g_t(\mu_s) + I_t^u \mu_u, \\ \Sigma_z &= \nabla_{x_t^s} g_t(\mu_s) \Sigma_s (\nabla_{x_t^s} g_t(\mu_s))^T + I_t^u \Sigma_u (I_t^u)^T + \Sigma_v. \end{cases} \tag{8.30}$$

Therefore, the multivariate normal probability density function for z_t is given by

$$f_z(z_t) = \frac{1}{(2\pi)^{n_z/2} \det(\Sigma_z)^{1/2}} \exp\left(-\frac{1}{2}(z_t - \mu_z)^T \Sigma_z^{-1} (z_t - \mu_z)\right), \tag{8.31}$$

where n_z is the dimension of z. The likelihood function is

$$\mathcal{L}_z(z_t) = \prod_{t=1}^{T} f_z(z_t) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{n_z/2} \det(\Sigma_z)^{1/2}} \exp\left(-\frac{1}{2}(z_t - \mu_z)^T \Sigma_z^{-1} (z_t - \mu_z)\right). \tag{8.32}$$

Finally, the log-likelihood function for z_t can be written as

$$\ell_z(z_t) = -\frac{T}{2}(n_z \log(2\pi) + \log(\det(\Sigma_z))) - \frac{1}{2} \sum_{t=1}^{T} (z_t - \mu_z)^T \Sigma_z^{-1} (z_t - \mu_z). \tag{8.33}$$

This derivation provides a different approach to calculating the log-likelihood function analytically by assuming normal distributions of x_t^s, x_t^u and v_t, and by linearizing $g_t(x_t^s)$. These approximations significantly reduce computational complexity, when comparing to the method in Section 8.3.2, allowing for efficient evaluation of different models as described in Section 8.3.

8.3.4 Prediction Error

We can test different models on its ability to predict the next quote of each instrument. This prediction can be compared to a reference prediction which is that the quotes will be equal to the quote in the previous time step. The difference between the prediction and the actual value will be presented as a MSE.

The model is thus tested on the last data point t: (x_t, z_t) and it makes a prediction $\hat{z}_{t+1}$ for the same instance based on the trained model

$$\hat{z}_{t+1} = \Phi(x_t), \tag{8.34}$$

where Φ is the trained model. The prediction error

$$e_{t+1} = \hat{z}_{t+1} - z_{t+1}, \tag{8.35}$$

is the difference between the prediction and the actual quote. The model's performance is based on how accurately it predicts the data point that was excluded. The overall performance is summarized using an error metric, MSE, over all the instruments

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} e_i^2, \tag{8.36}$$

where the total error is the average of the squared errors for each instrument (Khadka 2023). Furthermore, Equation 8.35 could be used in the statistical test described previously with $\bar{f}_j(\bar{z}, \hat{\bar{z}}) = \bar{e}$.

8.4 Q-Q Plot

For this model, assumptions regarding distributions for noise will be made and to verify the assumptions are accurate, a Q-Q plot will be used to compare empirical data distribution with theoretical assumed distribution.

In this case, the noise v will be hypothesized to follow a Student's t distribution, while w may follow a Normal distribution. When constructing the Q-Q plot, the quantiles of the acutal data will be plotted against the quantiles of the theoretical distribution, simply by sampling data from the distribution. If the distribution is correct, the points in the plot should closely follow a straight line, see Figure 2.

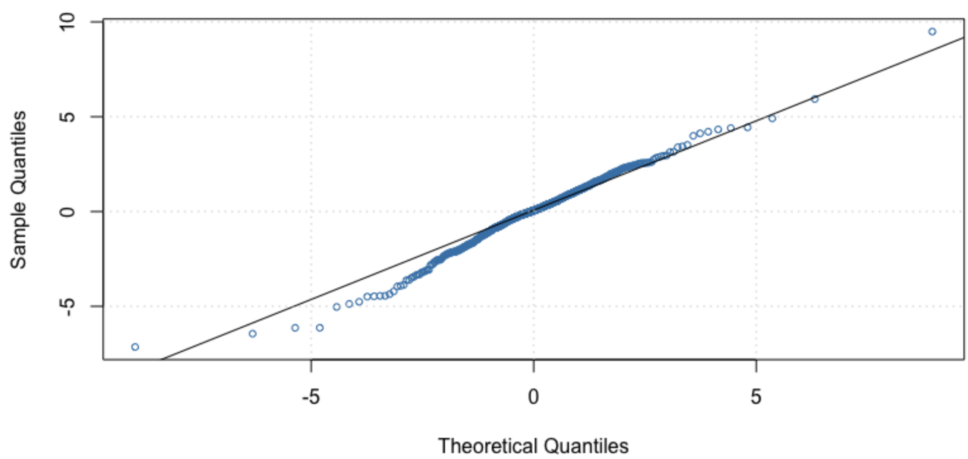

Figure 2: Q-Q plot comparing different GARCH models using the Student's t-distribution. This plot visually assesses the model fit to the assumed distribution, where deviations from the line indicate departures from the theoretical quantiles (Gyamerah 2019).

9 Results

In this section, the results from the project are presented. This involves key finding from the kalman filter model, interpolation and testing.

9.1 Simple interpolation methods

Below the results for the four different interpolation methods are shown.

9.1.1 Linear on Discount Factors

This method can exhibit jagged behavior with sharp changes at tenor nodes. As shown in Figure 4, there are more fluctuations at the beginning, primarily due to the presence of numerous OIS contracts with shorter maturities, such as 1-week, 2-week, 2-month, and so on. After the second year, however, there is only one contract for each subsequent year. The same pattern applies to the other three methods as well.

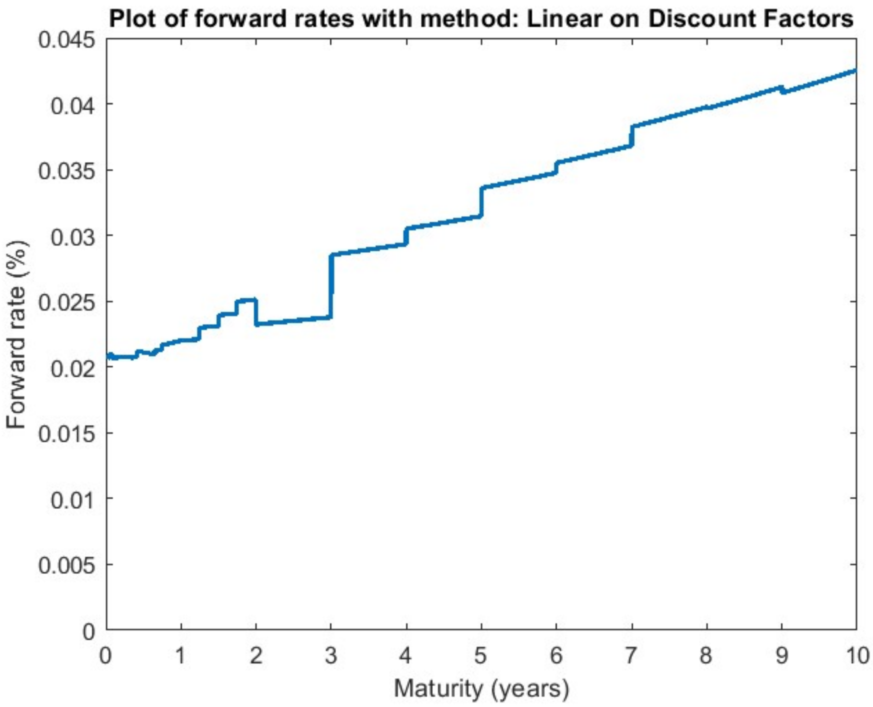

Figure 3: Linear on Discount Factors on 2005-08-15

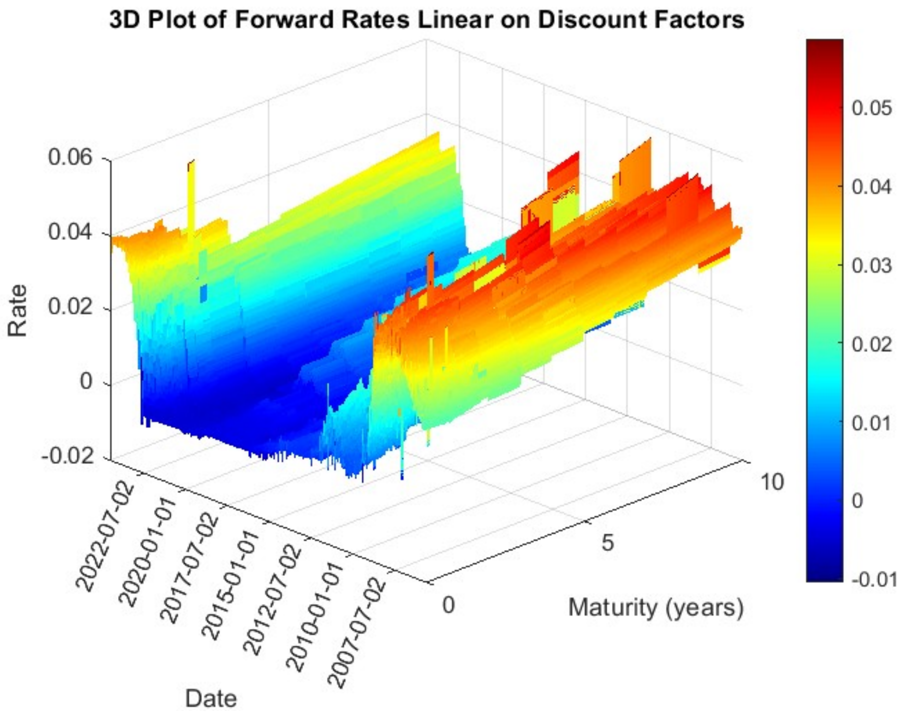

Figure 4: Linear on Discount Factors over the entire dataset

9.1.2 Linear on Spot Rates

This method minimizes abrupt transitions between discount factor nodes, resulting in a forward curve that is smoother and less susceptible to sharp fluctuations. However, for very short maturities, such as overnight or 1-week tenors, the linear interpolation of spot rates may introduce minor inaccuracies compared to methods like Raw Interpolation.

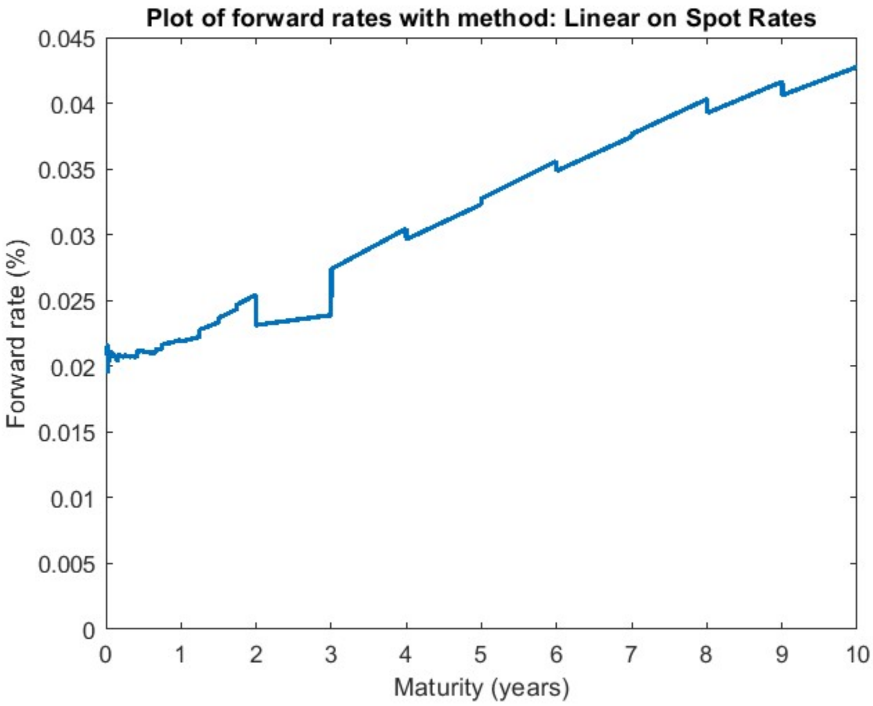

Figure 5: Linear on Spot Rates on 2005-08-15

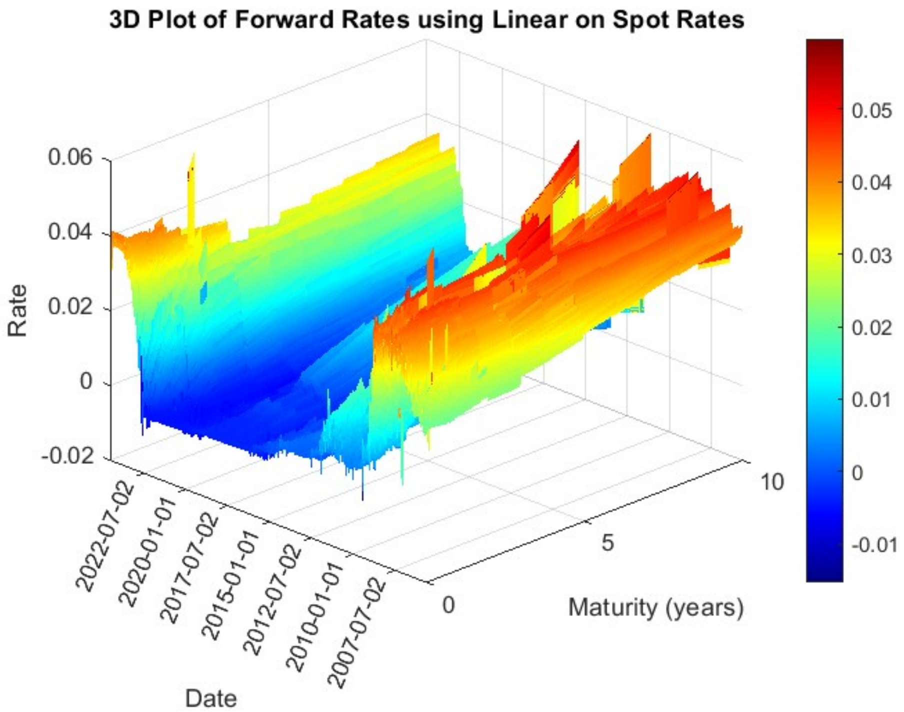

Figure 6: Linear on Spot rates over the entire dataset

9.1.3 Raw Interpolation

The curve generated using Raw Interpolation is piecewise constant between the known points.

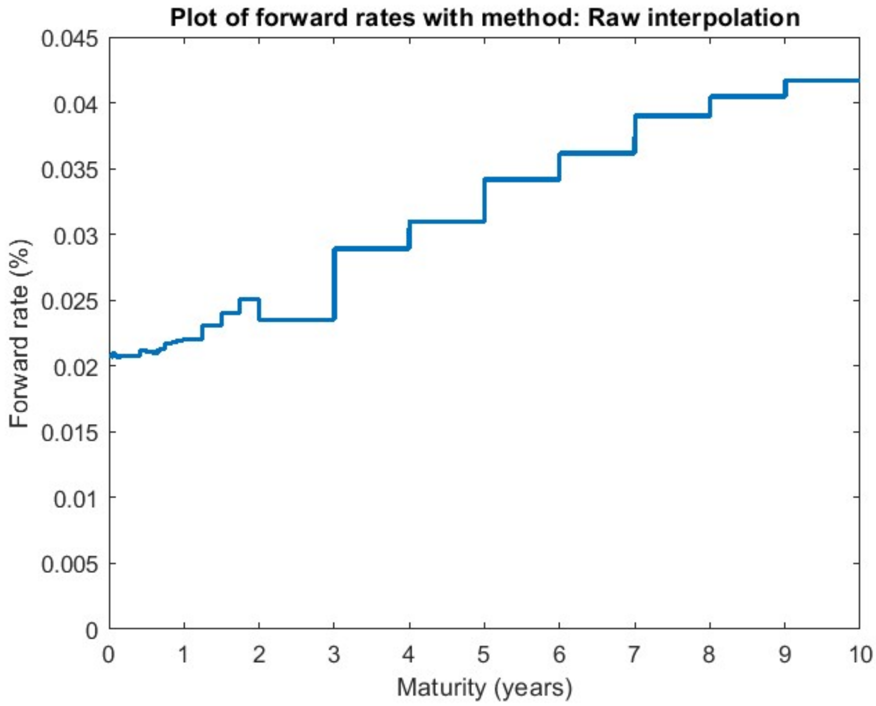

Figure 7: Raw Interpolation on 2005-08-15

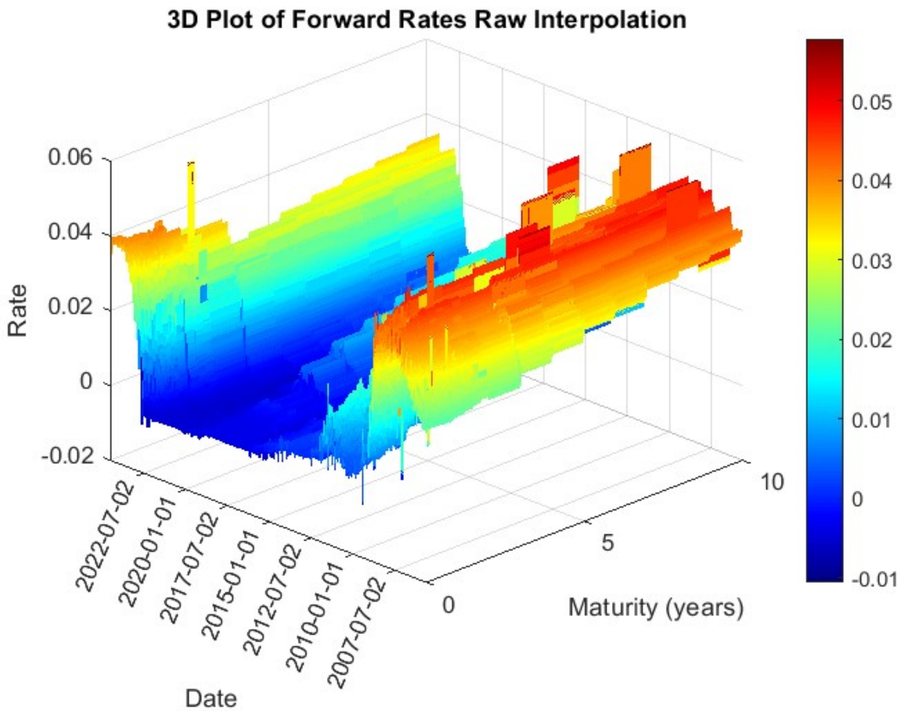

Figure 8: Raw Interpolation over the entire dataset

9.1.4 Linear on the Logarithm of Rates

The logarithmic transformation helps to reduce the oscillations encountered in simpler approaches.

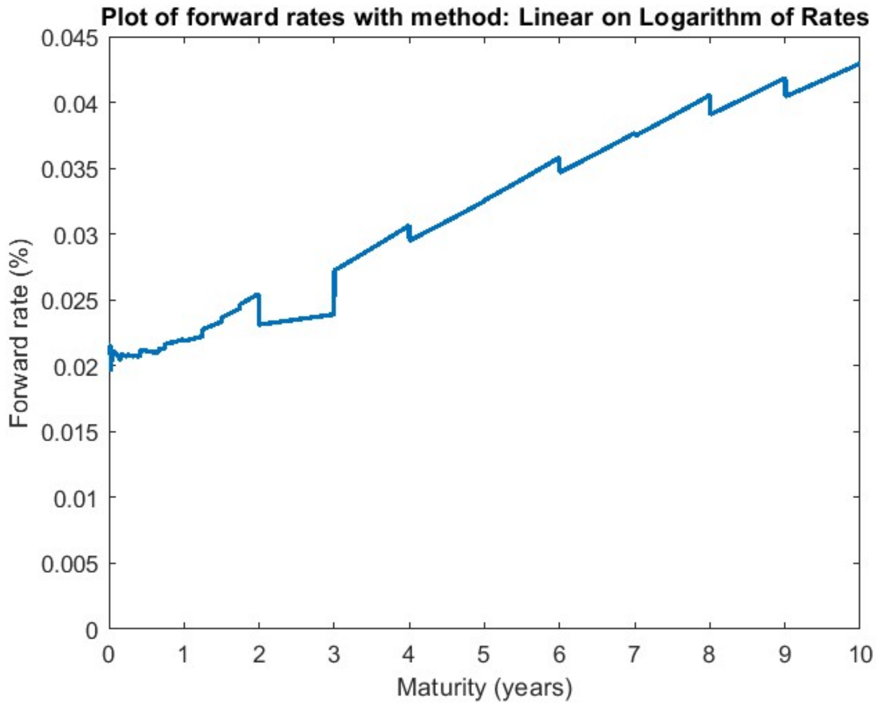

Figure 9: Linear on the Logarithm of Rates on 2005-08-15

Due to negative rates that occurred during the period 2005 to 2024, no three-dimensional graph can be plotted over the entire interval.

9.2 Optimization Model

The optimization model utilizes a variation of a Kalman filter, called Rauch Tung Striebel, which acts as a smoothener, shown in Figure 10 and Figure 11. This optimization model has been compared to realized prices of OIS contracts and has had an error $\epsilon \leq 10^{-5}$, especially showing accuracy for OIS contracts with higher maturities. The optimization model's covariance matrices, e.g. state change-, measurement- and process noise, have been specially chosen to fit the realized forward rates.

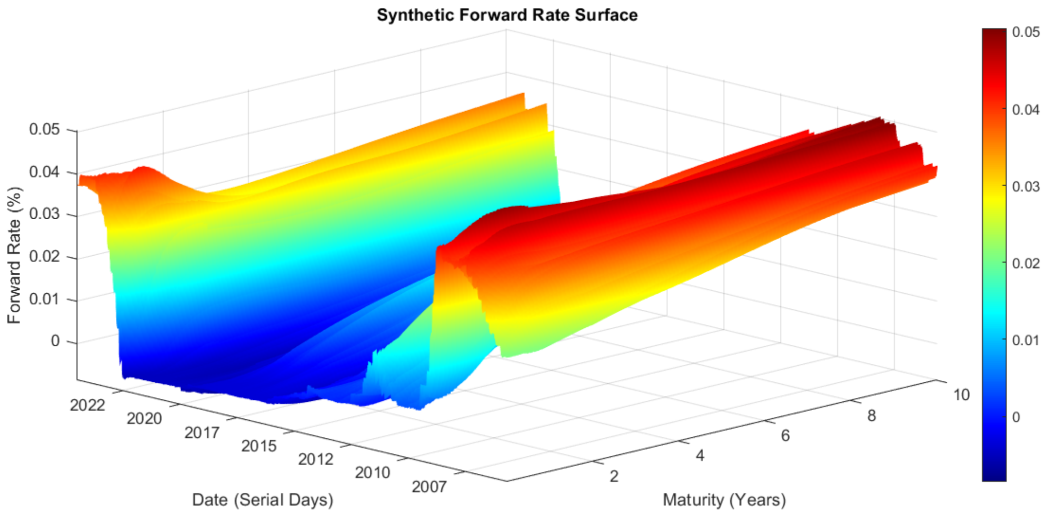

Figure 10: 3D plot over Forward Rates generated by the optimization model from in-sample data

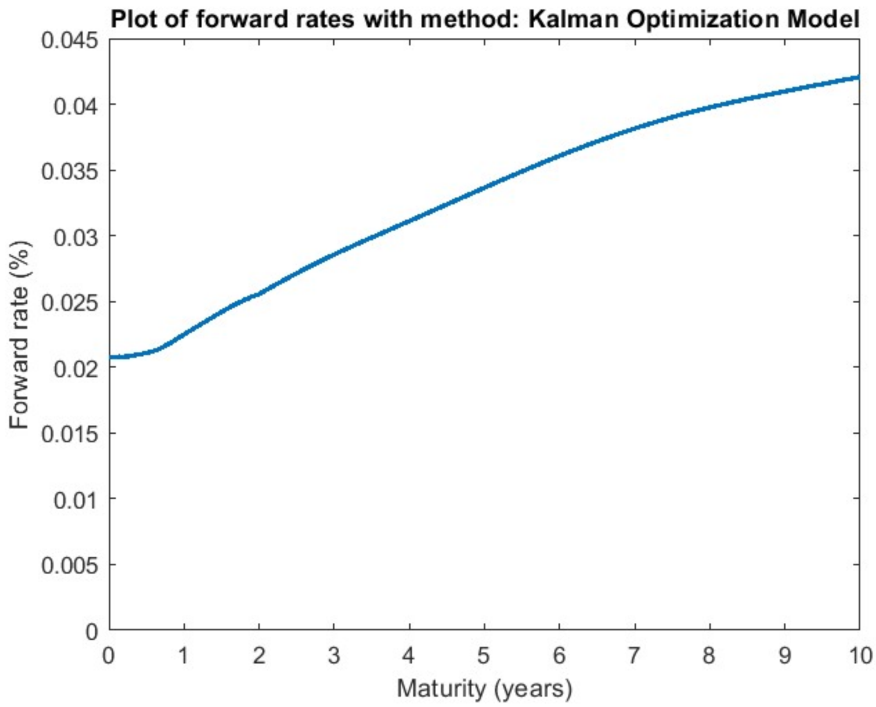

Figure 11: Kalman Optimization model forward rates on 2005-08-15

9.3 Validation

In this section, the results from the validation of the model will be presented.

9.3.1 LOOCV

In the optimization model, we performed a LOOCV to check the robustness of the model by removing an OIS contract that constructs the forward curve. Since we have 27 OIS contracts, a decision was made to remove only a total of three separate contracts. The selection of which contracts to remove have been carefully selected to represent the majority of the OIS contracts, meaning one short-term, one medium-term, and one long-term OIS contract. The following Table 1, shows the chosen contracts, the corresponding MSE between that instrument's price and its realized price over all time periods, and its MSE when the tenor was excluded from the data set.

Optimization model

OIS contract	MSE	MSE excl. one tenor
3 Week	2.0905e-08	3.6940e-08
1 Year	1.2249e-09	2.2416e-09
5 Year	7.6583e-09	1.0752e-08

Table 1: OIS contracts being removed, its mean squared errors for LOOCV, and regular mean squared errors, both over all of its time periods for the optimization model.

In Table 1, we observe that the excluding of all 3 week tenors increased the MSE from repricing the swap contract. We can observe the same pattern in the other tenors, which was expected since the exclusion of a tenor essentially removes the constraint which includes the prices of that tenor. However, the MSE is still rather small.

Interpolation

OIS Contract	Linear on Discount Factors	Linear on Spot Rates	Raw Interpolation
3 Week	5.1351e-07	6.5419e-07	5.1333e-07
1 Year	2.2178e-07	2.2022e-07	2.2387e-07
5 Year	5.1556e-06	5.9477e-06	5.0060e-06

Table 2: Mean squared errors (MSE) for LOOCV excluding one tenor at a time across the entire dataset for for the three interpolation methods: Linear on Discount Factors, Linear on Spot Rates, and Raw Interpolation.

After removing one contract, the resulting error is comparable across the different interpolation methods. Raw Interpolation provides the lowest error when excluding the 3-week and 5-year tenors with Linear on Discount Factors not far behind. Linear on Spot Rates is worse than both methods for the two tenors, however for the 1-year tenor it narrowly outperforms the others. Overall, the lowest errors occur for the 1-year tenor with a substantial increase in error for the 5-year tenor.

All compared interpolation methods exhibit very similar MSE in the non-LOOCV results. All resulting repricing errors here are on the magnitude of the tolerance level squared or lower, for the same tolerance level as in the LOOCV results of ten to the power of minus ten.

OIS Contract	MSE
3 Week	6.8810e-31
1 Year	1.3730e-27
5 Year	2.3528e-23

Table 3: MSE for the repricing error of Linear on Discount not excluding any contract.

The Linear on the Logarithm of Rates method was excluded from the validation process, as the model failed to generate sufficient usable results across the trade days, making a statistical comparison impractical.

Statistical Test

To evaluate the LOOCV performed for the different methods on the different tenors with statistical significance, statistical tests for the different combinations have been implemented, shown in Figure 12, 13, and 14.

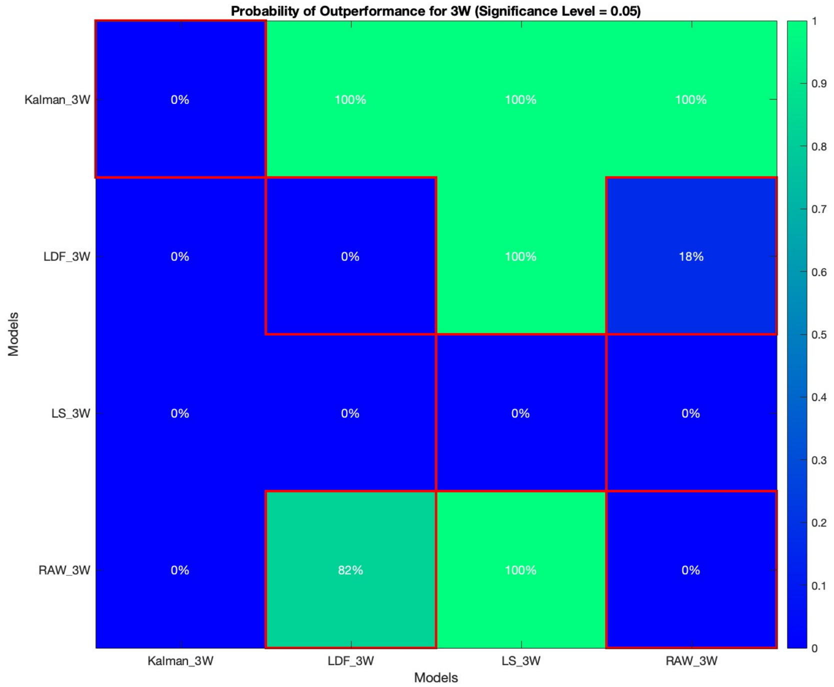

Figure 12: Statistical test for 3W OIS contract being repriced after LOOCV shows the probability that model (row) is better than model (column), and red-border shows if it is statistically insignificant.

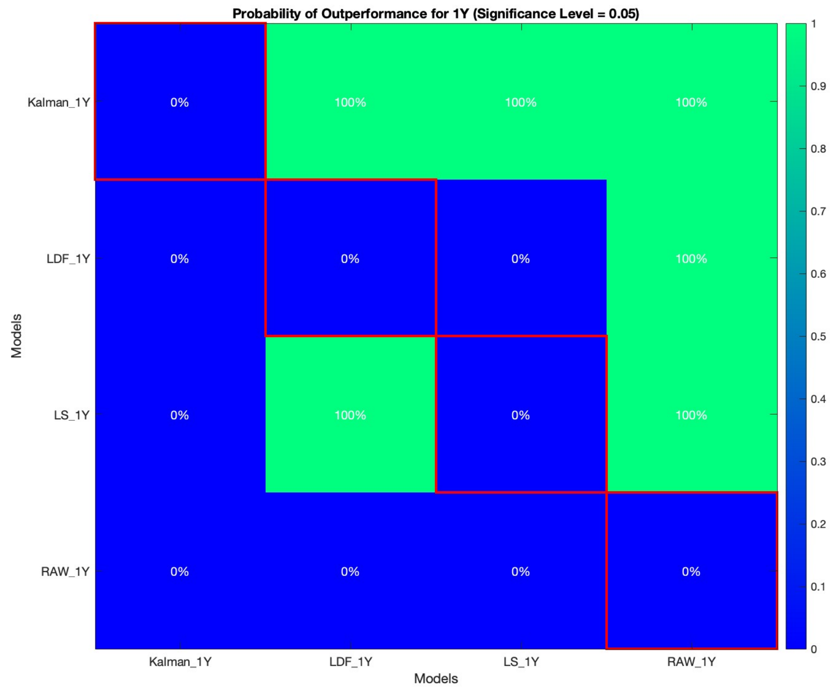

Figure 13: Statistical test for 1Y OIS contract being repriced after LOOCV shows the probability that model (row) is better than model (column), and red-border shows if it is statistically insignificant.

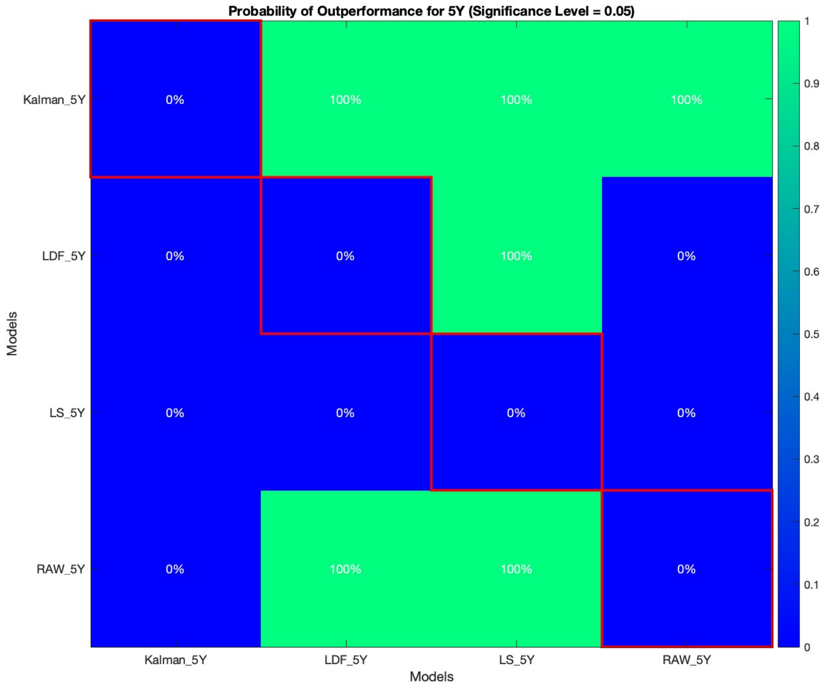

Figure 14: Statistical test for 5Y OIS contract being repriced after LOOCV shows the probability that model (row) is better than model (column), and red-border shows if it is statistically insignificant.

10 Conclusions

This section presents conclusions based on the results.

10.1 Optimization Model

From Figure 10, we can see that for longer maturities, the forward rates remain stable. It would be unreasonable for OIS contracts with similar maturities, such as a 7-year and a 9-year contract, to differ significantly in their forward rates. Since the difference in maturities is relatively small, and for the long-term interest rates is generally expected to follow a consistent trend based on current economic conditions and central bank policies. According to the theory of liquidity preference, investors typically require a premium for holding longer-term securities due to the increased risk. If the forward rates for these contracts were to differ greatly, it would imply an abnormal risk or liquidity premium that is not in accordance with the market's expectations for stable economic conditions. Liquidity preference theory often helps explain the general shape of the yield curve. In the case of these OIS contracts, the stability of the forward rates for similar maturities aligns with this theory, indicating that the market expects a consistent interest rate environment.

Between 2007 and 2010, forward rates were relatively high due to efforts to manage inflation before the financial crisis. After the crisis in 2008, rates dropped sharply as central banks took steps to support economic recovery. From 2020 onward, rates began to rise again as economies improved and inflation became a growing concern.

10.1.1 LOOCV

By a statistical test of the MSE from the LOOCV we can conclude that our optimization model outperforms the three interpolation methods, which can be seen in Figure 12, 13, and 14, over all the three different tenors compared. Since the different tenors is one of short duration, one of medium duration and one of long duration, it's very likely that the optimization model outperforms the interpolation methods for all the tenors. The statistical test did however indicate that raw interpolation of was the best interpolation method for longer tenors, this is however something that should be further evaluated to make sure that this result isn't unique for the 5Y OIS contract. Despite of this, as mentioned before the optimization method still outperforms all the interpolation methods for all tenors.

The values of the Tables 1 and 2 indicates that the out-performance of the optimization model compared to the interpolations methods is larger for the longer duration, as we can observe that the MSE of the interpolation methods is approximately 20 times bigger than for the optimization model. This strong out-performance is somewhat expected since the method models the long term swaps with the components from PCA, which captures the curvature of the forward curve, which means that contracts priced with the curve, even for tenors between the observed value is priced quite well. This essentially means that a 6.5 year swap can be priced with the forward curve quite well, even as the model has no observed values of these contracts.

10.2 Simple Interpolation Methods

The figures in Section 9.1 clearly show that the methods generate curves that closely resemble Figure 10. However, all the curves are more jagged and less smooth, which is a result of the curves needing to exactly solve the system. The interpolation methods are sensitive to the noise in the data, leading to curves of lower quality. Simple interpolation methods tend to produce lower-quality curves when contracts with similar maturities are present. Methods using spot rates are more prone to this issue, as they can exhibit greater oscillations. In contrast, Raw Interpolation as well as Linear on Discount Factors are less sensitive to this problem.

The results from the LOOCV and Statistical Test show the lowest repricing error for both the 3-week and 5-year tenor Raw Interpolation. However, there is no statistical significance for it beating Linear on Discount Factors for the 3-week tenor. From Figure 5 one can see that the rate curve for Linear on Spot Rates is very volatile at the start of the curve which could be the reason for its performance compared to Raw Interpolation and Linear on Discount Factors early on. The Linear on Spot Rates method also beats the other interpolation methods for the medium-length 1-year tenor with statistical significance, but its poor performance on both short and long-term tenors shows it unsuitable to use as a reference method for the Kalman Optimization model.

10.3 Further research

Topics of interest for future investigation include implementing other interpolation methods such as Cubic Spline. Although it was anticipated that these methods would not yield better results.

Given that the chosen covariance matrices are selected by perceived optimum in the optimization model, the covariance matrices have an arbitrary nature. Hence, improvements can be made by iteratively estimating the covariance matrices to have a more accurate joint distribution. Furthermore, there was an ambition to have student-t distributions instead of normal distribution, but because of time constraints this was not adequately assessed.

Another promising direction for further research would be applying interpolation techniques to contracts in different currencies, such as USD. This would provide valuable insights into how the implemented methods perform across various contract types and markets.

Due to time constraints, we were unable to implement additional validation techniques such as the QQ plot and the Karush-Kuhn-Tucker (KKT) test. Implementing these methods would have provided further rigor to the analysis by offering multiple layers of validation. Including these additional validation methods could have strengthened the reliability of the results. They remain a valuable area for future exploration to enhance the overall credibility and accuracy of the findings.

References

Akram, Saba and Quarrat Ul Ann (2015). “Newton Raphson Method”. In: *International Journal of Scientific & Engineering Research* 6.7, pp. 1748–1752.

Bartholomew-Biggs, M. et al. (2000). “Automatic Differentiation of Algorithms”. In: *Journal of Computational and Applied Mathematics* 124.1, pp. 171–190.

Baydin, A. G. et al. (2018). “Automatic Differentiation in Machine Learning: A Survey”. In: *arXiv preprint*. URL: https://arxiv.org/abs/1502.05767.

Blomvall, JÃ¶rgen (2024a). “Lecture Notes on Financial Optimization”.

– (2024b). “Maximum Likelihood - Kalman”.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). “Maximum Likelihood from Incomplete Data via the EM Algorithm”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 39.1, pp. 1–38. URL: http://www.jstor.org/stable/2984875.

Einicke, G. A. et al. (n.d.). “Riccati Equation and EM Algorithm Convergence for Inertial Navigation Alignment”. In: ().

Färm, Hannes and Fredrik Hammarbäck (2024). *Determining Intraday Market Prices for Interest Rate Derivatives*. Tech. rep. LinkÃ¶ping University.

Gyamerah, Samuel (2019). *Estimating the Volatility of Bitcoin Using GARCH Models*. DOI: 10.48550/arXiv.1909.04903.

Hagan, Patrick S. and Graeme West (n.d.). *Interpolation Methods for Curve Construction*. Unpublished Work.

Hull, J. C. (2017). *Options, Futures, and Other Derivatives*. 10th. Boston: Pearson Education.

Khadka, Nirajan (2023). *How Leave-One-Out Cross Validation (LOOCV) Improves Model Performance*. Accessed 7 Nov. 2024. URL: https://dataaspirant.com/leave-one-out-cross-validation-loocv/.

Rauch, H. E., F. Tung, and C. T. Striebel (1965). “Maximum Likelihood Estimates of Linear Dynamic Systems”. In: *AIAA Journal* 3.8, pp. 1445–1450. DOI: 10.2514/3.3166.

Rydholm, Clara and Gustav Villwock (2022). *Improving Term Structure Measurements by Incorporating Steps in a Multiple Yield Curve Framework*. Tech. rep. LinkÃ¶ping University.

Wikipedia, contributors (2024). *Karush-Kuhn-Tucker Conditions*. Last accessed 14 Nov. 2024. URL: https://en.wikipedia.org/w/index.php?title=Karush%E2%80%93Kuhn%E2%80%93Tucker_conditions&oldid=1228996335.