UC SANTA BARBARA

Dynamic Bond Ladder Investment Decision Making

Lei, Avery, Tony, Manny, Chi

Supervised by: Dr.Gareth Peters and Zimo Zhu

Department of Statistics and Applied Probability

University of California, Santa Barbara

March 24, 2024

Abstract

Bond ladder portfolios are a popular investment strategy used to manage interest rate risk, generate steady cash flows, and maintain liquidity. A key challenge in managing such portfolios is determining the optimal times to reinvest maturing bonds or make adjustments based on market conditions. In this project, we develop a systematic framework that employs an optimal stopping time method to identify the best moments for taking action. Our approach integrates yield curve bootstrapping, which accounts for part of our data preparation and allows us to work with data containing missing values, as well as the Dynamic Nelson Siegel (DNS) model, which is used to forecast the yield curve in the future to evaluate the mean and covariance of the future value of a bond portfolio.

1 Introduction

Bond ladder portfolios are a widely used strategy in fixed-income investments to manage interest rate risk, generate steady cash flows, and maintain liquidity. By staggering bond maturities, investors gain periodic reinvestment opportunities, balancing higher yields from long-term bonds with short-term liquidity. However, traditional bond ladder strategies follow mechanical reinvestment approaches that fail to account for changing interest conditions. This study introduces an adaptive decision-making framework that employs an optimal stopping time method to improve reinvestment timing and enhance portfolio performance.

Yield curve modeling is crucial in this approach. Traditional methods, such as smoothed bootstrapping, provide segmented yield curve estimates but often suffer from liquidity constraints. The Static Nelson-Siegel model offers a structured alternative by modeling yield curve levels, slopes, and curvatures[Diebold and Li, 2006]. Its extension, the Dynamic Nelson-Siegel (DNS) model, incorporates time-varying state variables estimated using the Kalman filter[Kalman, 1960], enabling real-time forecasting and portfolio optimization.

A key challenge in bond ladder management is determining optimal reinvestment and liquidation timing in a stochastic interest rate environment. This study formulates the decision process as a sequential optimization problem, leveraging Monte Carlo simulations to forecast yield curve scenarios. The optimal stopping time framework allows investors to maximize long-term portfolio value by dynamically adjusting their reinvestment strategy.

To construct the bond ladder portfolio, we manage our risk by choosing the best possible action to take at the optimal time. This requires accurate yield curve forecasting, highlighting the significance of statistical modeling techniques such as the Kalman filter.

Empirical analysis using historical Treasury bond data from the St. Louis Federal Reserve tests the effectiveness of our framework. We apply cubic spline bootstrapping to construct a smooth yield curve, ensuring accurate parameter estimation in the DNS model. Our results compare traditional and adaptive strategies, demonstrating that an adaptive approach, guided by optimal stopping rules and yield curve forecasts, enhances returns while mitigating downside risks.

This study contributes to fixed-income portfolio management by integrating yield curve modeling, Monte Carlo simulations, and optimal stopping theory to refine bond ladder strategies. The findings underscore the advantages of adaptive reinvestment, improving resilience against interest rate fluctuations and optimizing long-term performance.

This study is structured as follows. Section 2 introduces the concept of optimal stopping time and its application to our portfolio management. Section 3 defines the dynamic portfolio value process and outlines how different actions—such as buying, selling, or stripping bonds—affect the portfolio's value over time. In Section 4, we present the Nelson-Siegel model for yield curve modeling, covering both the static version and dynamic case using the Kalman filter. Section 5 describes the Treasury yield data and the bootstrapping methods used to generate smooth term structures. Section 6 presents case studies using simulated and empirical data to demonstrate model performance. We conclude in Section 7 with a summary of findings and directions for future research.

=

2 Optimal Stopping Time

An adaptive bond ladder portfolio involves managing a sequence of bond investments with strategic buy, sell, or strip decisions. The challenge lies in determining the optimal stopping times to execute these actions to maximize the portfolio's value. Mathematically, this problem can be formulated as an optimal stopping time problem, where we define a process $W(t)$ that represents the difference in portfolio value between taking an optimal action and taking no action at time t.

This section develops the stopping time framework adapted from classical optimal stopping theory and applies it to bond ladder portfolio optimization.

Notation

All relevant notation for this section if defined throughout the section.

2.1 Reward Function Formulation

Consider a bond portfolio evolving over time, where at each time step t, the investor observes the reward process $\{W(t)\}_{t=1}^{T}$. This process is defined as:

$$W(t) = \max_{A_t \in \mathcal{A}} \Pi_T(A_t) - \Pi_T(A_t^{\text{null}}),$$

where:

- $\Pi_T(A_t)$ is the projected portfolio value at time T if an action taken at time t.
- $\Pi_T(A_t^{\text{null}})$ is the projected portfolio value at time T if no action is taken at t.
- $\mathcal{A}$ represents the set of available actions.

Thus, $W(t)$ represents the opportunity cost of inaction at time t.

2.2 Optimal Stopping Rule

Definition 1. *A **single** stopping time $\delta \in \mathbb{N}$ is a $\mathcal{F}_t$-measurable random variable such that $\delta \leq T$ almost surely, where T is the end date of a given investment period. The decision to stop at time δ means executing an action (such as selling or stripping a bond) to maximize expected portfolio value.*

*A k-**multiple** stopping time $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_k]^T \in \mathbb{N}^k$ is analogous to a single stopping time, but for several actions occurring one after the other, i.e. $\delta_1 < \cdots < \delta_k \leq T$ almost surely.*

Definition 2. *In the The gain function associated with a k-multiple stopping time τ is given by:*

$$g(\boldsymbol{\delta}) = \sum_{i=1}^{k} W(\delta_i).$$

The objective is to find an optimal stopping rule that maximizes the expected gain:

$$v = \sup_{\boldsymbol{\delta} \in \mathcal{S}} \mathbb{E}[g(\boldsymbol{\delta})],$$

where $\mathcal{S}$ is the set of all stopping times.

Theorem 1 *Let $W(1), W(2), \ldots, W(T)$ be a sequence of independent random variables with known distribution functions $F_1, F_2, \ldots, F_T$, and the gain function $g(\boldsymbol{\delta}) = \sum_{j=1}^{k} W(\delta_j)$. Let $v^{L,l}$ be the value of a game where the agent is allowed to stop l times ($l \leq k$) and there are L ($L \leq T$) steps remaining. If there exist $\mathbb{E}[W(1)], \mathbb{E}[W(2)], \ldots, \mathbb{E}[W(T)]$ then the value of the game is given by*

$$\begin{aligned}
v^{1,1} &= \mathbb{E}[W(T)], \\
v^{L,1} &= \mathbb{E}\Big[\max\{W(T-L+1), v^{L-1,1}\}\Big], \quad 1 < L \leq T, \\
v^{L,l+1} &= \mathbb{E}\Big[\max\{v^{L-1,l} + W(T-L+1), v^{L-1,l+1}\}\Big], \quad l+1 < L \leq T, \\
v^{l,l} &= \mathbb{E}\Big[v^{l-1,l-1} + W(T-l+1)\Big].
\end{aligned}$$

If we put

$$\begin{aligned}
&\delta_1^* = \min\{m_1 : 1 \leq m_1 \leq T-k+1, W(m_1) \geq v^{T-m_1,k} - v^{T-m_1,k-1}\}; \\
&\delta_i^* = \min\{m_i : \delta_{i-1}^* < m_i \leq T-k+i, W(m_i) \geq v^{T-m_i,k-i+1} - v^{T-m_i,k-i}\}, \quad i = 2, \ldots, k-1; \\
&\delta_k^* = \min\{m_k : \delta_{k-1}^* < m_k \leq T, W(m_k) \geq v^{T-m_k,1}\}; \qquad (1)
\end{aligned}$$

then $\boldsymbol{\delta}^* = (\delta_1^*, \ldots, \delta_k^*)$ is the optimal multiple stopping rule.

3 Bond Ladder Portfolio Present Value Process

In this project, we define a **stochastic portfolio** whose value evolves over time based on the actions taken within our bond ladder management framework. The total portfolio value at any given time consists of three key components:

1. **The present value of the bond holdings** – representing the discounted value of all bonds currently held in the portfolio.

2. **The present value of the cash position** – including any proceeds from bond sales or uninvested funds, which accrue interest over time.

3. **The present value of outstanding debt** – capturing any borrowed funds used to finance bond purchases when the cash position is insufficient.

Each of these components is influenced by past decisions regarding bond purchases, sales, and cash allocations. For instance, if a bond is sold, the proceeds are added to the cash position, which then accrues interest. Given this dynamic structure, we establish the following assumptions to define the decision-making process within our framework:

3.1 Investment and Transaction Assumptions

1. **Periodic Bond Purchases**: Bonds are acquired at the beginning of each month in a deterministic manner. Each month constitutes an **investment period** during which bond purchases occur. If sufficient cash is available, it is used to finance the purchase; otherwise, the portfolio incurs debt to maintain the desired investment level.

2. **Borrowing and Lending Terms**: The interest rate for both borrowing and lending is **constant and independent of the yield curve**. This ensures that the cost of borrowing and the return on uninvested cash remain stable throughout the investment horizon.

3. **Discrete Action Timing**: Portfolio adjustments beyond the regular bond purchases—such as selling a bond or adjusting cash allocations—can only be made **once per investment period**. This constraint reflects a structured decision-making process, preventing continuous or high-frequency trading.

4. **No Transaction Costs or Market Frictions**: Bonds can be sold at their **present value at any time after purchase**, without incurring transaction fees or market frictions. This assumption simplifies the valuation process and ensures that asset liquidity does not impose constraints on decision-making.

These assumptions establish a well-defined framework for managing the bond ladder portfolio, enabling a structured approach to decision-making under uncertainty. By incorporating these constraints, our model ensures **realistic yet computationally manageable** portfolio dynamics, allowing us to evaluate the effectiveness of different investment strategies through our optimal stopping methodology.

3.2 Mathematical Construction of the Portfolio Process:

We propose a mathematical framework that models the evolution of a bond ladder portfolio over time. Our portfolio consists of three components:

- P_t: The present value of all bond holdings at time t.
- S_t: The surplus, representing available cash reserves at time t.
- D_t: The deficit, representing borrowed funds at time t.

At the initial time $t = 0$, the total portfolio value is given by:

$$\Pi_0 = P_0 + S_0 + D_0$$

where the surplus and deficit are determined by the initial cash position C_0 relative to the bond purchases:

$$D_0 = \mathbb{1}_{C_0<P_0} \cdot (P_0 - C_0), \quad S_0 = \mathbb{1}_{C_0\geq P_0} \cdot (C_0 - P_0)$$

The bond holdings at time 0 are expressed as:

$$P_0 = \sum_{i=1}^{m} N_i B_{w_{i,0},\tau_i}$$

where N_i represents the number of bonds of type i, and B_{t,T_i} denotes the present value at time t of all future cash flows of a bond with maturity T_i.

3.2.1 Bond Valuation with Future Cash Flows

Each bond's value depends on its future coupon payments and principal repayment. The present value of a bond is given by:

$$B_{t,T_i} = \sum_{j=1}^{v_i} \left[f_i(\exp(-\tau_{i,j} \cdot y_t(\tau_{i,j}))) \cdot \mathbb{1}_{T_{i,j}>0} \right] \cdot \mathcal{G}_1(A_{1:t})$$

$$+ \exp(-\tau_{i,j} \cdot y_t(\tau_{i,j})) \mathbb{1}_{\tau_{i,j}>0} \cdot \mathcal{G}_2(A_{1:t})$$

where $y_t(\tau)$ is the yield at time t for maturity τ, and $\mathcal{G}_1$ and $\mathcal{G}_2$ are indicator functions that ensure we only account for future cash flows that have not been sold:

$$\mathcal{G}_1(A_{1:t}) = \prod_{k=1}^{t} |\mathbb{1}_{\{\text{sell}=A_k \text{ or strip}=A_k\}} - 1|$$

$$\mathcal{G}_2(A_{1:t}) = \prod_{k=1}^{t} |\mathbb{1}_{\{\text{sell}=A_k\}} - 1|$$

3.2.2 Portfolio Evolution Over Time

At time t, the total portfolio value is:

$$\Pi_t = P_t + S_t + D_t$$

where the bond holdings, surplus, and deficit evolve as:

$$P_t = \sum_{i=1}^{m} B_{t,T_i}$$

$$D_t = (D_{t-1}e^r + h_{1,t}(A_t))_+$$

$$S_t = (S_{t-1}e^r + h_{2,t}(A_t))_+$$

The functions $h_{1,t}(A_t^{(q)})$ and $h_{2,t}(A_t^{(q)})$ capture the effect of the qth action taken at time t on the deficit and surplus, respectively.

3.2.3 Effect of Actions on the Portfolio

At each time step, different actions can be taken, each affecting the portfolio in specific ways:

1. Doing Nothing If no action is taken, the surplus and deficit simply grow at the risk-free rate:

$$h_{1,t}(A_t) = h_{2,t}(A_t) = 0$$

2. Selling Bonds of Type i When selling a bond, the proceeds increase the surplus, while the deficit adjusts accordingly:

$$h_{1,t}(A_t) = -N_i B_{t,T_i}$$

$$h_{2,t}(A_t) = -\min(0, D_{t-1}e^r + h_{1,t}(A_t))$$

3. Buying Bonds of Type i When purchasing a bond, cash reserves are used first, with any shortfall covered by borrowing:

$$h_{1,t}(A_t) = -\min(0, S_{t-1}e^r + h_{2,t}(A_t))$$

$$h_{2,t}(A_t) = -N_i B_{t,T_i}$$

4. Stripping Bonds of Type i Stripping a bond means selling all future coupon payments for their present value while retaining the principal:

$$h_{1,t}(A_t) = -\left[N_i(B_{t,T_i} - e^{-\tau_i y_t(\tau_i)} \cdot \mathbb{1}_{\tau_i>0})\right]$$

$$h_{2,t}(A_t) = -\min(0, D_{t-1} + h_{1,t}(A_t))$$

This framework defines a dynamic bond ladder portfolio that evolves through time based on investment actions. The recursive structure enables the modeling of reinvestment strategies, debt accumulation, and cash flow management while incorporating optimal decision-making rules.

4 Model Overview and Estimation

The Nelson-Siegel model is a widely used approach for modeling the yield curve, providing a parsimonious representation of interest rate dynamics. In our previous research, we employed the static Nelson-Siegel model, which assumes that the level, slope, and curvature factors remain constant over time. Under this assumption, we estimated the parameters using Ordinary Least Squares (OLS) and Generalized Least Squares (GLS), depending on the structure of the covariance matrix. The optimal decay parameter, λ, was determined via a grid search by maximizing the profile likelihood.

However, the static Nelson-Siegel model does not account for time variation in yield curve dynamics, limiting its applicability in forecasting and risk management. In this study, we extend our previous approach by implementing the Dynamic Nelson-Siegel (DNS) model, which allows the factor loadings to evolve over time. This extension captures the temporal structure of interest rates more effectively, improving both in-sample fit and out-of-sample forecasting accuracy. The DNS model is formulated as a state-space representation, where the latent factors follow a vector autoregressive process. We estimate the model parameters using the Kalman filter and maximum likelihood estimation, enabling a more flexible and dynamic characterization of yield curves.

By incorporating time variation in the Nelson-Siegel framework, our approach enhances the ability to model and forecast bond yields, making it particularly useful for applications in fixed income portfolio management and interest rate risk assessment. This study builds on our previous work while addressing its limitations, demonstrating the advantages of a dynamic factor structure in yield curve modeling.

4.1 Dynamic Nielson Siegel Model

Due to time-varying economic conditions, static parameters may not adequately capture the dynamics of the yield curve. Additionally, the results of Case III in Section 6 suggest that the dynamic Nelson-Siegel model may provide a better fit for the yield curve. Instead, we employ dynamic parameters within the Nelson-Siegel model to address these changes:

$$\mathbf{Y}_t(\boldsymbol{\tau}) = \boldsymbol{\Phi}(\lambda; \boldsymbol{\tau})\mathbf{X}_t + \epsilon_t, \tag{2}$$

where $\epsilon_t \overset{i.i.d}{\sim} \mathcal{MVN}(\mathbf{0}, \mathbf{Q})$, $\mathbf{v}_t \overset{i.i.d}{\sim} \mathcal{MVN}(\mathbf{0}, \mathbf{R})$, $\epsilon_t \perp \mathbf{v}_t$.

Extend it we will have

$$Y_t(\tau_i) = L_t + S_t \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} + C_t \left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} \right) + \varepsilon_t(\tau_i), \tag{3}$$

where $Y_t(\tau_i)$ is the bootstrapped yield at time t with maturity τ_i, L_t, S_t, C_t are the level, slope, and curvature at time t, and λ is the decay parameter.

where $\boldsymbol{\beta}_t = [L_t, S_t, C_t]^T$, $\mathbf{Y}_t = [Y_t(\tau_1), \ldots, Y_t(\tau_N)]^T$, $\boldsymbol{\Phi}(\lambda, \boldsymbol{\tau})$ is the $N \times 3$ design matrix with the i-th row being

$$\left[1, \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i}, \left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} \right) \right],$$

and $\boldsymbol{\varepsilon}_t = [\varepsilon_t(\tau_1), \ldots, \varepsilon_t(\tau_N)]^T \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma)$.

4.2 Nelson Siegel Model Under the Kalman Filter

Kalman Filter is a very efficient algorithm for estimating the actual state of a dynamic system from noisy measurements. It is particularly well adapted to systems that can be described by linear state transition and measurement models corrupted with Gaussian noise. Kalman Filter is a recursive algorithm that acts on new data as soon as it becomes available and is computationally expedient and thus widely applied in control systems, signal processing, computer vision, and navigation. Here we assumed the parameters matrix A, B, C, D, Q, R to be deterministic and constant overtime. Kalman Filter function can be written in:

$$\mathbf{X}_t = \mathbf{A}\,\mathbf{X}_{t-1} + \mathbf{B}\,\mathbf{W}_t \quad \mathbf{W}_t \sim \mathcal{MVN}(\mathbf{0}, \mathbf{Q}), \tag{4}$$

$$\mathbf{Y}_t = \mathbf{C}(\lambda; \boldsymbol{\tau})\,\mathbf{X}_t + \mathbf{D}\,\mathbf{Z}_t, \quad \mathbf{Z}_t \sim \mathcal{MVN}(\mathbf{0}, \mathbf{R}) \tag{5}$$

- $\mathbf{X}_t \in \mathbb{R}^{3\times 1}$: $\mathbf{X}_t = [L_t, S_t, C_t]^T$ is the state vector at time t, representing the level, slope, and curvature factors.
- $\mathbf{A} \in \mathbb{R}^{3\times 3}$: The state transition matrix.
- $\mathbf{B} \in \mathbb{R}^{3\times p}$: State noise transformation matrix, dimension correction for $\mathbf{W}_t$. (It is identity if the dimension of $\mathbf{W}_t$ is the same as $\mathbf{X}_t$).
- $\mathbf{W}_t \sim \mathcal{MVN}(\mathbf{0}, \mathbf{Q})$: The process noise, where $\mathbf{Q} \in \mathbb{R}^{3\times 3}$
- $\mathbf{Y}_t \in \mathbb{R}^n$: $\mathbf{Y}_t = [Y_t(\tau_1), \ldots, Y_t(\tau_n)]^T$ is the observed yield vector at different maturities τ.

- $\mathbf{C} \in \mathbb{R}^{n \times m}$: $\mathbf{C}(\lambda, \boldsymbol{\tau})$ is the $n \times 3$ design matrix of Nelson Siegel basis.
- $\mathbf{D} \in \mathbb{R}^{n \times r}$: Observation noise transformation matrix, dimension correction for $\mathbf{Z}_t$). (It is identity if the dimension of $\mathbf{Z}_t$ is the same as $\mathbf{Y}_t$)
- $\mathbf{Z}_t \sim \mathcal{MVN}(\mathbf{0}, \mathbf{R})$: The measurement noise, where $\mathbf{R} \in \mathbb{R}^{r \times r}$

State Prediction

We derived the distribution by calculating the marginal distribution:

$$f(\mathbf{X}_t|\mathbf{Y}_{1:t-1}) = \int \pi(\mathbf{X}_t|\mathbf{X}_{t-1})\pi(\mathbf{X}_{t-1}|\mathbf{Y}_{1:t-1})\, d\mathbf{X}_{t-1}$$

Hence the **state prediction** follows:

$$\mathbf{X}_t \mid \mathbf{Y}_{1:t-1} \sim \mathcal{MVN}(\hat{\mathbf{X}}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}) \tag{6}$$

$$\hat{\mathbf{X}}_{t|t-1} = \mathbf{A}\hat{\mathbf{X}}_{t-1|t-1}$$
$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{A}\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T$$

Observation Prediction

We derived the distribution by calculating the marginal distribution:

$$f(\mathbf{Y}_t|\mathbf{Y}_{1:t-1}) = \int \pi(\mathbf{Y}_t|\mathbf{X}_t)\pi(\mathbf{X}_t|\mathbf{Y}_{1:t-1})\, d\mathbf{X}_t$$

Hence the **observation prediction** follows:

$$\mathbf{Y}_t \mid \mathbf{Y}_{1:t-1} \sim \mathcal{MVN}(\hat{\mathbf{Y}}_{t|t-1}, \mathbf{F}_t) \tag{7}$$

$$\hat{\mathbf{Y}}_{t|t-1} = \mathbf{C}\hat{\mathbf{X}}_{t|t-1} = \mathbf{C}\mathbf{A}\hat{\mathbf{X}}_{t-1|t-1}$$
$$\mathbf{F}_t = \mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T + \mathbf{D}\mathbf{R}\mathbf{D}^T = \mathbf{C}(\mathbf{A}\boldsymbol{\Sigma}_{t-1}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T)\mathbf{C}^T + \mathbf{D}\mathbf{R}\mathbf{D}^T$$

State Update

We derived the distribution by calculating the conditional distribution over multivariate normal distribution:

$$f(\mathbf{X}_t \mid \mathbf{Y}_{1:t}) = \frac{f(\mathbf{Y}_t \mid \mathbf{X}_t) f(\mathbf{X}_t \mid \mathbf{Y}_{1:t-1})}{f(\mathbf{Y}_t \mid \mathbf{Y}_{1:t-1})}$$

Hence the **state update** follows:

$$\mathbf{X}_t \mid \mathbf{Y}_t \sim \mathcal{MVN}(\hat{\mathbf{X}}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \tag{8}$$

$$\hat{\mathbf{X}}_{t|t} = \hat{\mathbf{X}}_{t|t-1} + \mathbf{K}_t(\mathbf{Y}_t - \mathbf{C}\hat{\mathbf{X}}_{t|t-1})$$
$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t\mathbf{C}^T\boldsymbol{\Sigma}_{t|t-1}$$
$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1} = \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T(\mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T + \mathbf{D}\mathbf{R}\mathbf{D}^T)^{-1}$$

4.2.1 Parameter Estimation By MLE

Since the only available information for the dynamic linear system is the observation $\mathbf{Y}_{1:t}$, our goal is to maximize its likelihood using gradient descent.

Gradient Descent

We utilize gradient descent for parameter estimation until the likelihood is maximized, indicating that all parameters have converged to an optimal point. We update the parameters using the Newton's method step:

$$\Theta^{(k+1)} = \Theta^{(k)} - \alpha \frac{\partial \ell^{(k)}}{\partial \Theta}.$$

where the parameter vector Θ is constructed using the vech operator:

$$\Theta = \text{vech}(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{Q}, \mathbf{R}) = \left[A_{11}, A_{12}, ..., A_{mm}, B_{11}, ..., R_{rr}\right]^T.$$

Here, vech$(\cdot)$ denotes the vectorization operation that extracts all unique elements of the symmetric matrices $\mathbf{Q}$ and $\mathbf{R}$, along with all entries of the rectangular matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and $\mathbf{D}$, forming a single column vector. Thus, Θ_i represents the i-th element of Θ. Such an operator is necessary for numeric manipulation in matrix form.
Thus, we can update each entries of all parameters at every iteration k, where one iteration k is a loop from time 0 to time T for kalman filter

Learning Rate α

The learning rate α is a hyperparameter that controls the step size of updates during gradient descent. To ensure the convergence of stochastic approximation methods, the step size sequence α_k must satisfy the Robbins-Monro conditions:

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \quad \sum_{k=1}^{\infty} \alpha_k^2 < \infty$$

Partial log-Likelihood

By Equation (7), the marginal likelihood of multivariate normal distribution is given by :

$$\begin{aligned} L(\Theta \mid \mathbf{Y}_{1:T}) &= \int f(\mathbf{Y}_{1:T}, \mathbf{X}_{1:T} \mid \Theta)\, d\mathbf{X}_{1:T} \\ &= \prod_{T=1}^{T} L(\mathbf{Y}_t \mid \mathbf{Y}_{1:t-1}, \Theta), \end{aligned}$$

Then, we can compute log-marginal-likelihood of the observations:

$$\begin{aligned} \ell(\Theta \mid \mathbf{Y}_{1:T}) &= \sum_{t=1}^{N} \log L\big(\mathbf{Y}_t \mid \mathbf{Y}_{1:t-1}, \Theta\big) \\ &= \text{constant} - \frac{1}{2} \sum_{t=1}^{N} \ln |\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^{N} \mathbf{e}_t^T \mathbf{F}_t^{-1} \mathbf{e}_t. \end{aligned}$$

Each term $\ell(\Theta \mid \mathbf{Y}_{1:T})$ can be written in terms of

$$\mathbf{e}_t \;=\; \mathbf{Y}_t - \mathbf{C}\,\hat{\mathbf{X}}_{t|t-1}, \quad \mathbf{F}_t \;=\; \mathbf{F}_t = \mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T + \mathbf{D}\mathbf{R}\mathbf{D}^T$$

To derive the gradients of l with respect to the model parameters $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{Q}, \mathbf{R}$, we have a general case for partial derivatives over Θ_i as follow, and we can calculate the partial derivatives of each entries in the appendix in Section8 (including $\frac{\partial \ell}{\partial A_{ij}}$, $\frac{\partial \ell}{\partial Q_{ij}}$, $\frac{\partial \ell}{\partial R_{ij}}$).

Since we aim to estimate the parameters, we compute the partial derivatives at time t based on the results from time $t-1$, initializing the partial derivatives at time 0 with initial guesses. For each time t, we derive the partial derivatives for **state prediction**, **state prediction covariance**, **observation prediction**, and **prediction covariance** as follows:

$$\begin{aligned}
\frac{\partial \hat{\mathbf{X}}_{t|t-1}}{\partial \Theta_i} &= \frac{\partial}{\partial \Theta_i}\mathbf{A}\hat{\mathbf{X}}_{t-1|t-1} \\
&= \frac{\partial \mathbf{A}}{\partial \Theta_i}\hat{\mathbf{X}}_{t-1|t-1} + \mathbf{A}\frac{\partial \hat{\mathbf{X}}_{t-1|t-1}}{\partial \Theta_i}
\end{aligned} \tag{9}$$

$$\begin{aligned}
\frac{\partial \boldsymbol{\Sigma}_{t|t-1}}{\partial \Theta_i} &= \frac{\partial}{\partial \Theta_i}(\mathbf{A}\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T) \\
&= \frac{\partial \mathbf{A}}{\partial \Theta_i}\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}^T + \mathbf{A}\frac{\partial \boldsymbol{\Sigma}_{t-1|t-1}}{\partial \Theta_i}\mathbf{A}^T + \mathbf{A}\boldsymbol{\Sigma}_{t-1|t-1}\frac{\partial \mathbf{A}^T}{\partial \Theta_i} \\
&\quad + \frac{\partial \mathbf{B}}{\partial \Theta_i}\mathbf{Q}\mathbf{B}^T + \mathbf{B}\frac{\partial \mathbf{Q}}{\partial \Theta_i}\mathbf{B}^T + \mathbf{B}\mathbf{Q}\frac{\partial \mathbf{B}^T}{\partial \Theta_i}
\end{aligned} \tag{10}$$

$$\begin{aligned}
\frac{\partial \hat{\mathbf{X}}_{t|t}}{\partial \Theta_i} &= \frac{\partial}{\partial \Theta_i}\big(\hat{\mathbf{X}}_{t|t-1} + \mathbf{K}_t(\mathbf{Y}_t - \mathbf{C}\hat{\mathbf{X}}_{t|t-1})\big) \\
&= \frac{\partial}{\partial \Theta_i}\big(\hat{\mathbf{X}}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{e}_t\big) \\
&= \frac{\partial \hat{\mathbf{X}}_{t|t-1}}{\partial \Theta_i} + \frac{\partial \boldsymbol{\Sigma}_{t|t-1}}{\partial \Theta_i}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{e}_t + \boldsymbol{\Sigma}_{t|t-1}\frac{\partial \mathbf{C}^T}{\partial \Theta_i}\mathbf{F}_t^{-1}\mathbf{e}_t \\
&\quad - \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \Theta_i}\mathbf{F}_t^{-1}\mathbf{e}_t + \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{e}_t}{\partial \Theta_i}
\end{aligned} \tag{11}$$

$$\begin{aligned}
\frac{\partial \boldsymbol{\Sigma}_{t|t}}{\partial \Theta_i} &= \frac{\partial}{\partial \Theta_i}(\boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t\mathbf{C}\boldsymbol{\Sigma}_{t|t-1}) \\
&= \frac{\partial \boldsymbol{\Sigma}_{t|t-1}}{\partial \Theta_i} - \frac{\partial \boldsymbol{\Sigma}_{t|t-1}}{\partial \Theta_i}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{C}\boldsymbol{\Sigma}_{t|t-1} \\
&\quad - \boldsymbol{\Sigma}_{t|t-1}\frac{\partial \mathbf{C}^T}{\partial \Theta_i}\mathbf{F}_t^{-1}\mathbf{C}\boldsymbol{\Sigma}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \Theta_i}\mathbf{F}_t^{-1}\mathbf{C}\boldsymbol{\Sigma}_{t|t-1} \\
&\quad + \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{C}}{\partial \Theta_i}\boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{C}\frac{\partial \boldsymbol{\Sigma}_{t|t-1}}{\partial \Theta_i}
\end{aligned} \tag{12}$$

Referring (9)(10)(11)(12), we can calculate the partial derivatives of innovation $\mathbf{e}_t$ and innovation covariance $\mathbf{F}_t$

$$\begin{aligned}\frac{\partial \mathbf{F}_t}{\partial \Theta_i} &= \frac{\partial}{\partial \Theta_i}\left(\mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T + \mathbf{D}\mathbf{R}\mathbf{D}^T\right) \\ &= \frac{\partial \mathbf{C}}{\partial \Theta_i}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}^T + \mathbf{C}\frac{\partial \boldsymbol{\Sigma}_{t|t-1}}{\partial \Theta_i}\mathbf{C}^T + \mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\frac{\partial \mathbf{C}^T}{\partial \Theta_i} \\ &\quad + \frac{\partial \mathbf{D}}{\partial \Theta_i}\mathbf{R}\mathbf{D}^T + \mathbf{D}\frac{\partial \mathbf{R}}{\partial \Theta_i}\mathbf{D}^T + \mathbf{D}\mathbf{R}\frac{\partial \mathbf{D}^T}{\partial \Theta_i}\end{aligned} \tag{13}$$

$$\begin{aligned}\frac{\partial \mathbf{e}_t}{\partial \Theta_i} &= \frac{\partial}{\partial \Theta_i}(\mathbf{Y}_t - \mathbf{C}\mathbf{A}\hat{\mathbf{X}}_{t-1|t-1}) \\ &= -\frac{\partial \mathbf{C}}{\partial \Theta_i}\mathbf{A}\hat{\mathbf{X}}_{t-1|t-1} - \mathbf{C}\frac{\partial \mathbf{A}}{\partial \Theta_i}\hat{\mathbf{X}}_{t-1|t-1} - \mathbf{C}\mathbf{A}\frac{\partial \hat{\mathbf{X}}_{t-1|t-1}}{\partial \Theta_i}\end{aligned} \tag{14}$$

Hence, with Equation(13) and (14), we have partial log-likelihood at time t

$$\frac{\partial \ell_t}{\partial \Theta_i} = -\frac{1}{2}\mathrm{Tr}\left(\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \Theta_i}\right) - \frac{1}{2}\left[\frac{\partial \mathbf{e}_t^T}{\partial \Theta_i}\mathbf{F}_t^{-1}\mathbf{e}_t - \mathbf{e}_t^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \Theta_i}\mathbf{F}_t^{-1}\mathbf{e}_t + \mathbf{e}_t^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{e}_t}{\partial \Theta_i}\right] \tag{15}$$

Finally, the gradient of the likelihood function over the entire Kalman filter from time 0 to time t can be obtained by stacking all the partial log-likelihood over time.

$$\begin{aligned}\frac{\partial \ell}{\partial \Theta_i} &= -\frac{1}{2}\sum_{t=1}^{T}\mathrm{Tr}\left(\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \Theta_i}\right) - \frac{1}{2}\sum_{t=1}^{N}\left[\frac{\partial \mathbf{e}_t^T}{\partial \Theta_i}\mathbf{F}_t\mathbf{e}_t + \mathbf{e}_t^T\frac{\partial \mathbf{F}_t^{-1}}{\partial \Theta_i}\mathbf{e}_t + \mathbf{e}_t^T\mathbf{F}_t\frac{\partial \mathbf{e}_t}{\partial \Theta_i}\right] \\ &= -\frac{1}{2}\sum_{t=1}^{T}\mathrm{Tr}\left(\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \Theta_i}\right) - \frac{1}{2}\sum_{t=1}^{N}\left[\frac{\partial \mathbf{e}_t^T}{\partial \Theta_i}\mathbf{F}_t\mathbf{e}_t - \mathbf{e}_t^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \Theta_i}\mathbf{F}_t^{-1}\mathbf{e}_t + \mathbf{e}_t^T\mathbf{F}_t\frac{\partial \mathbf{e}_t}{\partial \Theta_i}\right]\end{aligned} \tag{16}$$

The partial log-likelihood $\frac{\partial \ell}{\partial \Theta_i}$ over time 0 to time t can be used for gradient descent at the k-th iteration of parameter estimation.

Convergence Criterion

Here, we have several ways to check the convergence of parameter estimation.

$$\|\Theta_k - \Theta_{k-1}\| \leq \delta.$$

This method is widely used, but we must carefully choose a small value of δ to ensure that the convergence criterion is not smaller than the learning rate. Otherwise, the convergence check will always be satisfied if the learning rate is too small. So we have to use different values $\delta^{(k)}$ for each iteration.

5 Data Overview and Preparation

To apply the Nelson Siegel Model to a portfolio of Treasury bonds for optimal weighting, we take St. Louis Federal Reserve spot rate curve estimates and apply a bootstrapping procedure to obtain a set of uniformly spaced spot rate curve fits. We then take these as inputs to the Nelson Siegel model for the purpose of providing a parsimonious and interpretable parameterization of yield curve dynamics. The journey of how these inputs are obtained are explained in full in this section.

5.1 An Overview of Treasury Bonds and Yields

United States treasuries are bonds for which the U.S. government is the counterparty. In return for providing a future set of cash flows to the buyer, the U.S. government receives a lump sum of capital. This capital goes towards funding government operations in the short term, and the low risk level expected of these bonds helps ensure liquidity for many financial institutions in markets around the globe. In fact, the U.S. government can theoretically print any amount of money it needs to meet these debt obligations, which is why the cash flows determined by these bonds are often considered risk-free.

As such, the vast majority of bonds issued by the United States specify a fixed set of cash flows on predetermined dates. We denote the attributes of such bonds in the following manner:

Symbol	**Description**
a	Number of payments per year
b	Number of years until bond matures
c	Total number of payments: $c = ab$
α	Nominal coupon rate
r	Coupon rate by period: $r = \frac{\alpha}{a}$
T	Date of Maturity of Bond
t	Current calendar date
τ	Time remaining until bond matures: $\tau = T - t$
I	Nominal annual interest rate
j	Interest rate by period: $j = \frac{I}{a}$
i	Annual effective interest rate
P	Market Price of Bond
F	Face Value of Bond
R	Redemption Value
g	Modified coupon rate: $g = \frac{Fr}{C}$
$y_t(\tau)$	Yield to maturity; e.g., spot rate, IRR, par rate

Table 1: Notation used in this study.

Among the measurements listed above, the annual yield of the bonds indicated by i is the most useful for comparing the performances of other bonds. This represents the expected internal rate of return of a bond over a year, and this metric is among those chosen to determine whether a bond should go in an investor portfolio. For a bond with c regularly spaced cash flows purchased at price P with coupon rate r, face value F, redemption value C, and interest rate per period j, the bond price is as follows:

$$P = rFa^{j} + \frac{C}{(1+j)^{c}}, \tag{17}$$

$$a^{j} = \sum_{k=1}^{c} (1+j)^{-k} \tag{18}$$

Solving for j given all other parameters requires a computational root solver, but once found the annual yield i may be derived from j. However, whenever a bond is sold in markets after issuance, especially after the payment of coupons, an adjusted formula is necessary to determine the bond yield of the remaining cash flows given the most recent price. Most large data providers accomplish this by using the clean bond pricing formula.

Let $B_k = rFa^{j}_{\overline{k}|} + Cv^{k}$ and $\tau = \lfloor \tau \rfloor + f$. Then denote the dirty bond price by $D_\tau = (1+fj)B_{\lfloor \tau \rfloor}$. This price does not account for the most immediate cash flow, which we subtract from the dirty bond price to arrive at the clean bond price as follows.

$$C_\tau = D_\tau - Cg(\frac{(1+j)^{f} - 1}{j}) \tag{19}$$

C_τ denotes the clean bond price for a given time to maturity. Solving for j in this equation gives the clean yield estimate corresponding to the payment period for any bond that pays fixed coupons. Most databases will report the nominal yield adjustment of this result, so in most cases the yield should then be converted to the annual effective yield rate for comparing and modeling similar bonds.

When deriving the corresponding time to maturity for a bond of a given yield on a certain date, the day-counting convention must be accounted for [Åkerlund, 2024]. Many different bond issuers including the US Treasury employ different counting conventions to define the unit of time for the accrual of interest on bond issuance. Hence it is necessary when obtaining raw bond data to refer to the corresponding counting convention for each bond to accurately compare and model yields of different tenors. Obtaining the time to maturity for every yield of a subset of treasury bonds over time allows for the construction of a Treasury yield curve.

5.2 St. Louis FRED Treasury Spot Rate Curves

The data set of interest for this study concerns spot rate curves, which are yield curves that represent returns on bonds that pay only a single cashflow. For a data set of bonds that pay out multiple cash flows as coupons over time, deriving the spot rate curve requires either matching the cash flows of bonds of varying tenors to compute the spot rates directly or performing a change of measure to arrive at a zero coupon bond equivalent yield. In this case, the St. Louis Federal Reserve performs the former method by employing recent par bond yields.

Par bonds are bonds whose market prices closely match their redemption value, and typically trade at par closer to their issuance because the Federal Reserve auctions them such that their coupon rate matches their yield rate. These auctions occur in what are known as primary markets, where large institutions and investors have the opportunity to directly purchase treasuries from the federal government. Upon purchase, some of these participants may choose to sell their Treasury bonds prior to maturity in what are

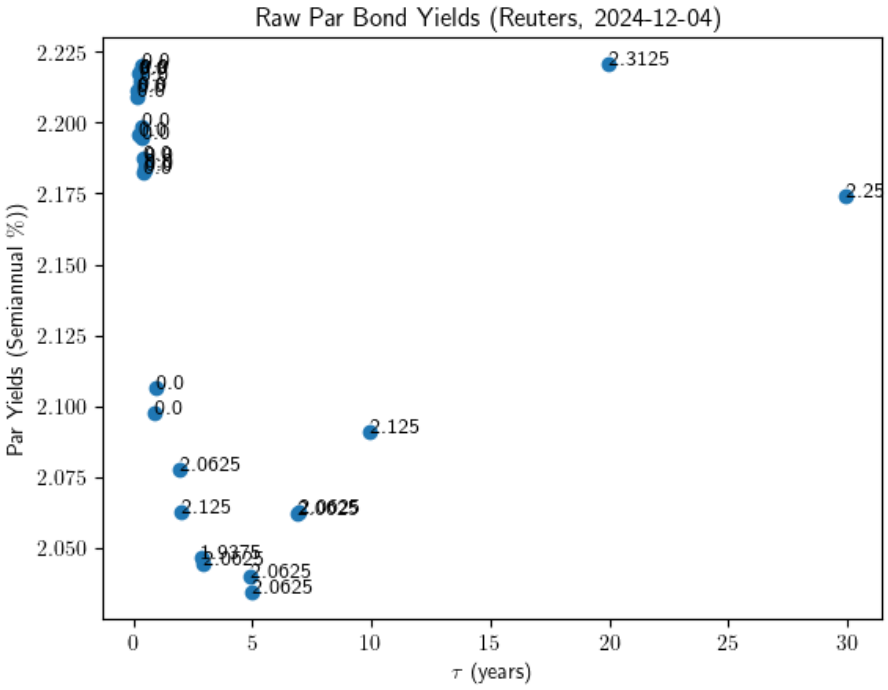

Figure 1: Bonds issued within last 2 months of December 4, 2012. Bonds without coupons are denoted with a coupon value of 0. Observe that those without coupons fall roughly close to their yields, which the Fed may consider as par bonds, and that the par rates of zero coupon bonds are equivalent to spot rates.

known as secondary markets. It is in these secondary markets where the St. Louis Federal Reserve collects par yield curve data to estimate spot rates [of the Treasury, 2021].

Since the cash flows of recently issued par bonds roughly align, spot rate estimates may be computed from them first trivially via iterative methods and then sampled from any tenor through an interpolation method. The interpolation method currently utilized by the Fed is called the Monotone Convex Method. This technique ensures that the forward rate curve, which defines the rate of return from a future point in time to an even further point in time, remains positive [Hagan and West, 2006]. From this interpolation procedure, the St. Louis Federal Reserve posts spot rate curve estimates at commonly issued tenors for par bonds (e.g. 3 months, 1 year, 3 years, etc.) after the bond market concludes trading everyday.

5.3 Features of Fed Data Set

Figure 7 visually shows all the information contained within a single date of the data set, whereas figure 3 shows the full data set across every date. Spanning from 1962 to current year, we observe that the yield curve may be perceived as a surface in 3-D space whose yields span over tenors and dates.

However, the bond market will not necessarily trade on every conceivable date, especially on weekends or holidays recognized by the US treasury market. So missingness will be anticipated on such dates in this data set. In addition, there are also some tenors for which the United States government simply did not issue bonds. This culmination of this missingness may be visualized as follows:

Much of the missingness may be observed from the tenors associated with treasury

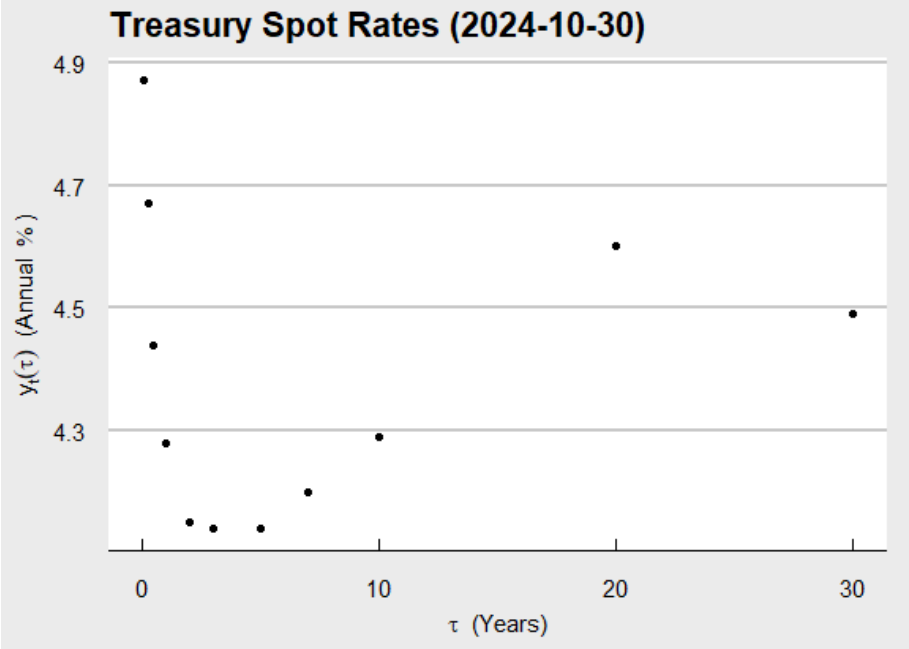

Figure 2: Treasury yields posted for October 30, 2024.

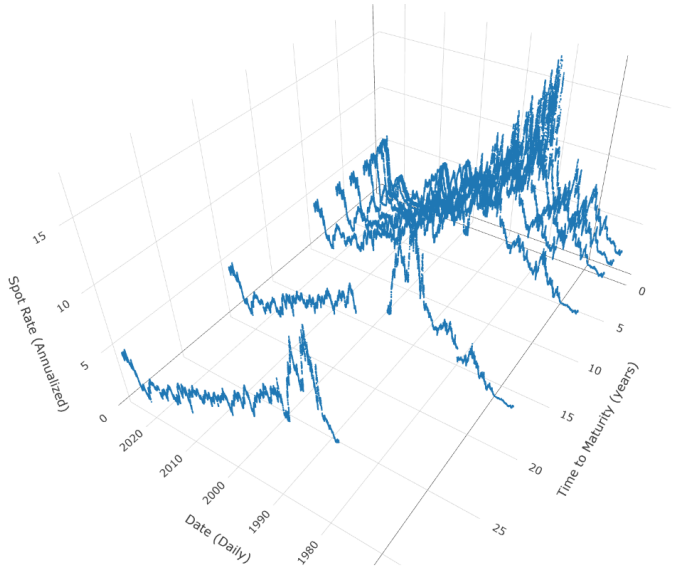

Figure 3: Treasury full data set posted in Full.

bills (1 month, 3 month, and 6 month tenors). Since the data assessed in this table retain its order, it follows that either the U.S. government did not issue some T-bills for a certain period of time or that the government simply did not fit the curve at those tenors for the spot rates reported. Further inspection of missingness at the 20 and 30 year tenors are also ongoing, but to deal with this missingness in our data set, we apply a bootstrap procedure to tenors over every date of the time series.

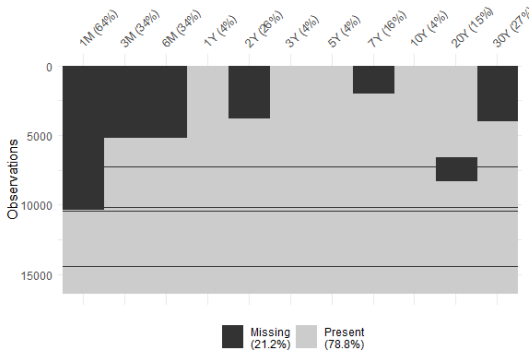

Figure 4: Missingness in yields table (see vis_miss package in R naniar package)

5.4 Processing Treasury Yield Data

Taking the time series of these Fed fits across the commonly issued tenors, the goal is to bootstrap them so that a sufficient number of uniformly spaced tenors may be supplied to the Nelson Siegel Model for further modeling. Bootstrapping is broadly defined in a fixed income setting as an interpolation of yields across their respective tenors. For this paper, we chose the method of cubic smoothing splines to bootstrap the treasury spot rate curve from the St. Louis Federal Reserve fits of par treasury spot rates. The reasoning behind the selection for cubic the selection of cubic smoothing splines over splines of other degree may be explained by their optimality on the L_2 norm [Wahba, 1990], and so we proceed with a model that takes advantage of this result.

Suppose we divide a bounded interval of possible tenors into sub-intervals whose points of separation we define as knots. Denote this set of knots $\{\tau_1, \tau_2, ..., \tau_M\}$ such that $0 \leq \tau_1 < \tau_2 < ... < \tau_M$. Then the spot rate $y_t(\tau)$ for $\tau \in [0, \tau_M]$ may be expressed as follows.

$$y_t(\tau) = \sum_{i=1}^{M+2} \alpha_{t,i} B_{t,i,3}(\tau), \tag{20}$$

$$B_{t,i,n}(\tau) = \frac{\tau - \tau_i}{\tau_{i+k} - \tau_i} B_{t,i,n-1}(\tau) + \frac{\tau_{i+k+1} - \tau}{\tau_{i+k+1} - \tau_{i+1}} B_{t,i+1,n-1}(\tau). \tag{21}$$

In the equation above, for each $i \in \{1, 2, ..., M+2\}$ α_i denotes a set of coefficients corresponding to basis functions $B_{t,i,3}(\tau)$. The Cox-de-Boor recursion on the basis functions listed above allows for exact estimation of the cubic b-spline basis functions for a supplied set of knots at a given tenor [de Boor, 1972]. Expressed in matrix form, the equivalent representation is

Moreover, from this natural cubic smoothing spline representation we may also derive an expression for penalized cubic splines via the ridge regression representation [Hoerl and Kennard, 1970].

$$\mathbf{y}_t = \mathbf{B}_t \boldsymbol{\alpha}_{t,ridge}, \tag{22}$$

$$\boldsymbol{\alpha}_{t,ridge} = (\mathbf{B}_t^\top \mathbf{B}_t + \delta_t \mathbf{I})^{-1} \mathbf{B}_t^\top \mathbf{y}_t \tag{23}$$

δ denotes the tuning parameter which helps to control over fitting. Among the possible selection of penalized cubic splines and natural cubic splines, we decided to examine the performance of natural cubic spline fits for yield curve interpolation and examine the performance of penalized cubic splines in future studies. Furthermore, we will also examine the performance of clamped splines on the yield curve in the future to explore the impact of assuming the yield on at expiration of a bond to be zero.

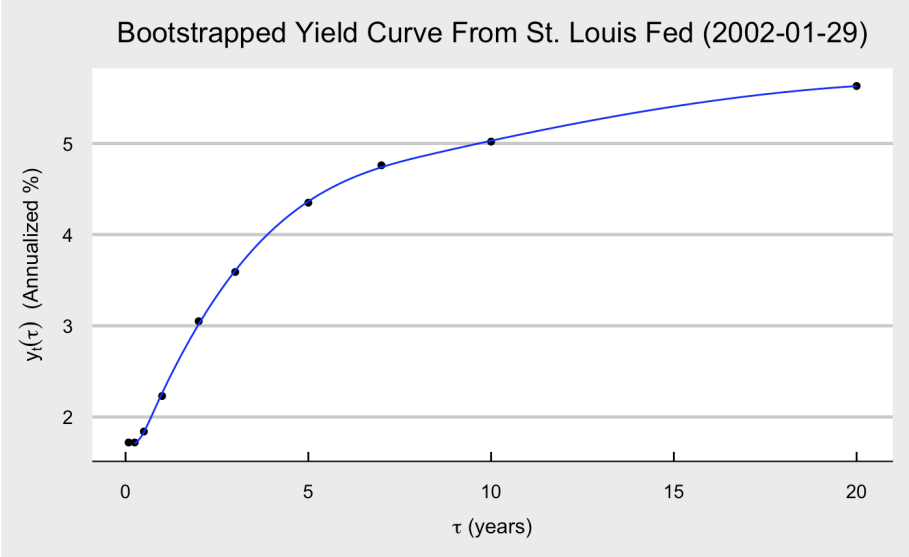

Figure 5: Natural Cubic Smoothing Spline Bootstrap fit.

For a select set of knots, we observe that our bootstrap procedure is prone to overfitting as may be suspected from figure 5. To this end, taking advantage of penalized splines is certainly one approach to resolving this problem. However, methods for optimal knot selection and placement should be considered prior to avoid reducing model parsimony. To find the optimal knots for fitting the yield curve, we currently select common 5 quantiles which split the range of tenors by 25% and employ those for our knots. Future efforts will go towards exploring more consistent methods optimal knot selection and placement. Sampling from the natural cubic smoothing spline fits at yearly intervals, we next explore the estimation techniques of the Nelson Siegel model under different scenarios.

6 Results

In an interest rate environment such that the underlying true level, slope, and curvature are presumed to be constant throughout time, we compare an adaptive portfolio process with a stationary portfolio process. The stationary process entails no actions beyond simply buying bonds at regularly spaced intervals and letting them mature with time, whereas the adaptive process takes advantage of our stopping rule and makes decisions which maximize the forecast performance of the bond ladder portfolio.

6.1 Model Constraints

6.1.1 Common Portfolio Assumptions

Throughout each case study, we begin with 100,000 dollars and allow for the purchase of 5000 units worth of treasuries at the end of every month. Each unit corresponds to a product of the face value of the bond and the number of bonds purchased, and scales the present value of the respective bond type. At the start, 70,000 dollars are pooled across three bonds that share the same coupon rate, frequency, and tenor at issuance. The only variation across these bond types are the issuance dates and maturity dates, and all new bond purchases corresponding to the 5000 unit delegation are of one-year tenors as of purchase. For remaining cash or debt after purchase, they are assumed to accrue at the same fixed annual interest rate of three percent.

6.1.2 Constraints on the Adaptive Portfolio

Following the same assumptions listed prior, the Adaptive Portfolio gains the ability to take one action every day between new bond purchases. Across each asset, the decision to sell, strip, or continue the position is weighed across all assets and determined by the value function of the asset on the date of evaluation. Moreover, an action entails the adjustment of a single asset. Hence no two different types of bonds may be acted upon in the same day less continuing the position by default.

6.2 Performance under Stationary Interest Rate Regime

Under conditions where the yield curve possesses relatively the same shape across time, the stationary and adaptive portfolios are implemented with unsurprising results.

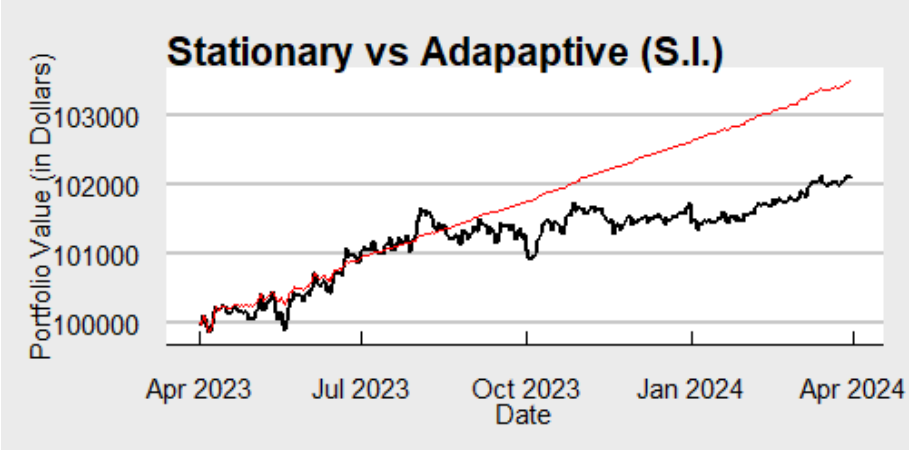

Figure 6: Yield Curve Under Stationary Interest Rate Regime

The yields simulated for this interest rate regime tend to fall within a range of one to three percent. Since the lending rate is pegged at 3 percent, the bond portfolio gradually exits bond positions until it has only one bond remaining at most within each period. This rate of return is preferable overall to the smaller range offered by the simulated bond yields. So in the figure above, the adaptive portfolio (in red) surpasses our stationary portfolio (in black) gradually over time.

6.3 Performance under Adaptive Interest Rate Regime

Enabling the shape of the yield curve to invert and jump every month still grants an edge to the adaptive portfolio. We observe below that despite frequently flipping between a range of four to five percent in annual treasury yields to a range of zero to one percent, the stationary portfolio eventually falls short.

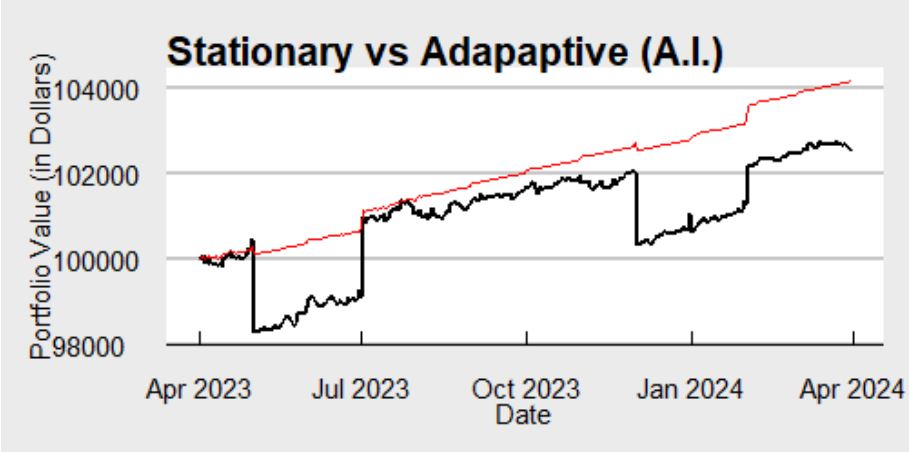

Figure 7: Yield Curve Under Adaptive Interest Rate Regime

Although the frequent, sudden changes in the yield curve may skew the forecasts, the stationarity maintained within each period prior to the yield curve changing does not hinder the performance of the adaptive portfolio. Part of this may also be explained by the zero to one percent interest rate regime skewing forecasts to being less than three percent on the balance, and may thereby lead to more bonds being sold or stripped.

6.4 Remarks on Testing of Real Interest Rate Data

Before testing real interest rate data, we wish to ensure that the Dynamic Nelson Siegel Model can capture more complex interest rate regimes. As it currently stands, the performance of the yield forecasts may depend heavily on the time period the model is trained on. In modeling more recent years, the dynamic Nelson Siegel model does not tend to perform well given the Fed rate hikes, rate cuts, and varying shape of the yield curve taken throughout 2020 to 2024. Hence, improving upon the Nelson Siegel model framework will be of great interest for upcoming research.

7 Future Studies

7.1 Covariance Regression

We can also model the ϵ_t in the Nelson-Siegel model by state space model to accommodate the situation of dynamic covariance matrix:

$$\epsilon_t = \gamma_t B X_t + \zeta_t, \tag{24}$$

where $\gamma_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $\zeta_t \overset{i.i.d}{\sim} \mathcal{MVN}(0,\Psi)$,$X_t$ are the observed exogenous factors, and $E(\gamma_t \times \zeta_t) = 0$.

Then, we have:

$$Var(\mathbf{Y}_t(\boldsymbol{\tau})|X_t, \boldsymbol{\beta}_t; B, \Psi) = BX_tX_t^TB^T + \Psi. \tag{25}$$

And after detrending: $E_t = \mathbf{Y}_t(\boldsymbol{\tau}) - \boldsymbol{\Phi}(\lambda; \boldsymbol{\tau})\hat{\boldsymbol{\beta}}_t$, we could use the Expectation-Maximization algorithm(EM algorithm) to estimate B and Ψ, also known as Covariance-Regression ([Hoff and Niu, 2012])

7.2 Extend Kalman Filter with GARCH

We can also use a GARCH model to incorporate volatility, resulting in a more conservative estimate of the confidence interval.

$$\mathbf{X}_t = \mathbf{A}\,\mathbf{X}_{t-1} + \mathbf{B}\,\mathbf{W}_t, \quad \mathbf{W}_t \sim \mathcal{MVN}(\mathbf{0}, \mathbf{Q}_t), \tag{26}$$

$$\mathbf{Y}_t = \mathbf{C}\,\mathbf{X}_t + \mathbf{D}\,\mathbf{Z}_t, \quad \mathbf{Z}_t \sim \mathcal{MVN}(\mathbf{0}, \mathbf{R}), \tag{27}$$

$$\mathbf{Q}_t = \boldsymbol{\omega}_0 + \sum_{j=1}^{q} \boldsymbol{\alpha}_j \mathbf{W}_{t-j}\mathbf{W}_{t-j}^\top + \sum_{i=1}^{p} \boldsymbol{\beta}_i \mathbf{Q}_{t-i} \tag{28}$$

References

[de Boor, 1972] de Boor, C. (1972). On calculating with b spliens. *Journal of Approximation Theory*.

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

[Hagan and West, 2006] Hagan, P. and West, G. (2006). Methods for constructing a yield curve. Technical report, Wilmott Magazine. Available at: https://downloads.dxfeed.com/specifications/dxLibOptions/HaganWest.pdf.

[Hoerl and Kennard, 1970] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*.

[Hoff and Niu, 2012] Hoff, P. D. and Niu, X. (2012). A covariance regression model. *Statistica Sinica*, pages 729–753.

[Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.

[of the Treasury, 2021] of the Treasury, U. D. (2021). Yield curve methodology change informmation sheet. Technical report, U.S. Department of the Treasury. Available at: https://home.treasury.gov/policy-issues/financing-the-government/yield-curve-methodology-change-information-sheet.

[Salmond, 1989] Salmond, D. J. (1989). Tracking and state estimation in uncertain environments. *IEE Proceedings F (Radar and Signal Processing)*, 136(2):73–82.

[Tao, nd] Tao, T. (n.d.). What's new – terence tao's blog. Accessed: 2024-03-21.

[Wahba, 1990] Wahba, G. (1990). Spline models for observational data. *SIAM*.

[Åkerlund, 2024] Åkerlund, E. (2024). Day count conventions. *Treasury Systems*. See https://support.treasurysystems.com/support/solutions/articles/103000058036-day-count-conventions for more information on day counting conventions.

8 Appendix

Notation

Symbol	Description	Section
Π_t	Total portfolio value at time t	Portfolio Process
P_t	Position value of bonds at time t	Portfolio Process
S_t	Surplus at time t	Portfolio Process
D_t	Deficit at time t	Portfolio Process
N_i	Number of bonds of type i with maturity T_i	Portfolio Process
B_{t,T_i}	Present value at t of bond with face value 1 and maturity T_i	Portfolio Process
r_i	Coupon rate of bond type i	Portfolio Process
n_t	Number of remaining unpaid coupons at time t	Portfolio Process
m	Number of different bond types in the portfolio	Portfolio Process
$y_t(\tau)$	Yield curve function at time t for maturity τ	Portfolio Process
A_t	Action taken at time t	Portfolio Process
$\mathcal{G}_1(A_{1:t})$	Indicator function for unsold future cash flows	Portfolio Process
$\mathcal{G}_2(A_{1:t})$	Indicator function for unsold face value	Portfolio Process
$h_{1,t}(A_t)$	Effect of action A_t on deficit	Portfolio Process
$h_{2,t}(A_t)$	Effect of action A_t on surplus	Portfolio Process
$Y_t(r)$	Bootstrapped yield at time t for maturity r	DNS Model
λ	Decay parameter controlling term structure curvature	DNS Model
$\beta_{\mathbf{t}} = [L_t, S_t, C_t]^T$	State vector containing level, slope, and curvature factors	DNS Model
$\mathbf{\Phi(A, r)X_t}$	Nelson-Siegel design matrix transformation	DNS Model
ϵ_t	Process noise term	DNS Model
$\mathbf{X_t}$	State vector at time t	Kalman Filter
$\mathbf{A}$	State transition matrix	Kalman Filter
$\mathbf{W_t}$	Process noise vector	Kalman Filter
$\mathbf{Q}$	Covariance matrix of process noise	Kalman Filter
$\mathbf{C}$	Observation matrix (Nelson-Siegel design basis)	Kalman Filter
$\mathbf{D}$	Observation noise transformation matrix	Kalman Filter
$\mathbf{Z_t}$	Measurement noise	Kalman Filter
$\mathbf{R}$	Covariance matrix of observation noise	Kalman Filter
$\mathbf{\Sigma_{t|t-1}}$	State prediction covariance matrix	Kalman Filter
$\mathbf{F_t}$	Covariance matrix of observation prediction	Kalman Filter
$\mathbf{K_t}$	Kalman gain matrix	Kalman Filter
$\mathbf{X_{t|t}}$	Updated state estimate after measurement	Kalman Filter
$W(t)$	Reward function at time t	Stopping Time
$\Pi_T(A_t)$	Projected portfolio value at time T given action A_t	Stopping Time
$\Pi_T(A_t^{\text{null}})$	Projected portfolio value at time T if no action is taken	Stopping Time
$\mathcal{A}$	Set of available actions	Stopping Time
δ	Stopping time (random variable)	Stopping Time
$g(\delta)$	Gain function for a stopping time δ	Stopping Time
S	Set of all stopping times	Stopping Time
v	Optimal expected gain	Stopping Time
F_t	Distribution function of reward variable $W(t)$	Stopping Time
$v^{L,1}$	Value function for stopping problem with L steps left	Stopping Time
$v^{L,k}$	Value function when k stopping actions are allowed	Stopping Time
δ_i	Time step at which the i-th stopping action is taken	Stopping Time

9 Derivation of Gaussian Convolution Integral

Consider the State and Observation equations[Salmond, 1989] [Tao, nd]:

$$\begin{cases} X_t = AX_{t-1} + W_t \\ Y_t = CX_t + Z_t \end{cases}$$

Where:

$$\begin{cases} W_t \sim \mathcal{N}(0, Q) \\ Z_t \sim \mathcal{N}(0, R) \end{cases}$$

We have:

$$\begin{cases} \pi(y_t|x_t) \sim \mathcal{N}(Cx_t, R) \\ \pi(x_t|x_{t-1}) \sim \mathcal{N}(Ax_{t-1}, P_{t|t-1}) \end{cases}$$

Where:

$$\begin{cases} P_{t|t-1} = AP_{t-1|t-1}A^T + Q \\ \hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1} \end{cases}$$

We can obtain the distribution of $\pi(y_t|y_{t-1})$ with the integral:

$$\pi(y_t|y_{t-1}) = \int \pi(y_t|x_t)\pi(x_t|y_{t-1})dx_t$$

$$\pi(y_t|y_{t-1}) = \frac{\int \exp\Big\{ -\frac{1}{2}\Big((y_t - Cx_t)^T R^{-1}(y_t - Cx_t) + (x_t - \hat{x}_{t|t-1})^T P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1})\Big)\Big\}dx_t.}{(2\pi)^{(m+n)/2}|R|^{1/2}|P_{t|t-1}|^{1/2}}$$

We first deal with the term in the exponential (ignoring the $-\frac{1}{2}$) for now

$$(y_t - Cx_t)^T R^{-1}(y_t - Cx_t) + (x_t - \hat{x}_{t|t-1})^T P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1})$$

Where each term is rewritten as:

$$(y_t - Cx_t)^T R^{-1}(y_t - Cx_t) = y_t^T R^{-1} y_t - y_t^T R^{-1} C x_t - x_t^\top C^\top R^{-1} y_t + x_t^T C^T R^{-1} C x_t.$$

And:

$$(x_t - \hat{x}_{t|t-1})^T P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1}) = x_t^T P_{t|t-1}^{-1} x_t - \hat{x}_{t|t-1}^T P_{t|t-1}^{-1} x_t - x_t^\top P_{t|t-1}^{-1}\hat{x}_{t|t-1} + \hat{x}_{t|t-1}^T P_{t|t-1}^{-1}\hat{x}_{t|t-1}.$$

By combining the terms and completing the square, we get:

$$\begin{aligned}
& x_t^T(C^T R^{-1} C + P_{t|t-1}^{-1})x_t - x_t^T(C^T R^{-1} y_t + P_{t|t-1}^{-1}\hat{x}_{t|t-1}) - (y_t^\top R^{-1} C + \hat{x}_{t|t-1}^\top P_{t|t-1}^{-1})x_t \\
& + y_t^T R^{-1} y_t + \hat{x}_{t|t-1}^T P_{t|t-1}^{-1}\hat{x}_{t|t-1} \\
&= x_t^\top H^{-1} x_t - x_t^\top(C^\top R^{-1}(y_t - C\hat{x}_{t|t-1}) + H^{-1}\hat{x}_{t|t-1}) - ((y_t - C\hat{x}_{t|t-1})^\top R^{-1} C + \hat{x}_{t|t-1}^\top H^{-1})x_t \\
& + y_t^\top R^{-1} y_t + \hat{x}_{t|t-1}^\top P_{t|t-1}^{-1}\hat{x}_{t|t-1} \\
&= x_t^\top H^{-1} x_t - x_t^\top H^{-1} h - h^\top H^{-1} x_t + y_t^\top R^{-1} y_t + \hat{x}_{t|t-1}^\top P_{t|t-1}^{-1}\hat{x}_{t|t-1} \\
&= x_t^\top H^{-1} x_t - x_t^\top H^{-1} h - h^\top H^{-1} x_t + h^\top H^{-1} h - h^\top H^{-1} h + y_t^\top R^{-1} y_t + \hat{x}_{t|t-1}^\top P_{t|t-1}^{-1}\hat{x}_{t|t-1} \\
&= (x_t - h)^\top H^{-1}(x_t - h) - h^\top H^{-1} h + y_t^\top R^{-1} y_t + \hat{x}_{t|t-1}^\top P_{t|t-1}^{-1}\hat{x}_{t|t-1}
\end{aligned}$$

Where

$$H^{-1} = C^T R^{-1} C + P_{t|t-1}^{-1} \quad \text{and} \quad h = \hat{x}_{t|t-1} + HC^\top R^{-1}(y_t - C\hat{x}_{t|t-1})$$

Now, we deal with the term:

$$-h^\top H^{-1} h + y_t^\top R^{-1} y_t + \hat{x}_{t|t-1}^\top P_{t|t-1}^{-1} \hat{x}_{t|t-1}$$

Rewrite using the definition of h and H Using basic matrix operations, while keeping in mind that $P_{t|t-1}$, R, and therefore H are all symmetric and invertible:

$$\begin{aligned}
&- h^\top H^{-1} h + y_t^\top R^{-1} y_t + \hat{x}_{t|t-1}^\top P_{t|t-1}^{-1} \hat{x}_{t|t-1} \\
&= -(\hat{x}_{t|t-1} + HC^\top R^{-1}(y_t - C\hat{x}_{t|t-1}^\top))^\top H^{-1}(\hat{x}_{t|t-1} + HC^\top R^{-1}(y_t - C\hat{x}_{t|t-1}^\top)) \\
&\quad + y_t^\top R^{-1} y_t + \hat{x}_{t|t-1}^\top P_{t|t-1}^{-1} \hat{x}_{t|t-1} \\
&= -\hat{x}_{t|t-1}^\top H^{-1} \hat{x}_{t|t-1} - (y_t - C\hat{x}_{t|t-1})^\top R^{-1} CHC^\top R^{-1}(y_t - C\hat{x}_{t|t-1}) - \hat{x}_{t|t-1}^\top C^\top R^{-1} y_t + \hat{x}_{t|t-1}^\top C^\top R^{-1} C\hat{x}_{t|t-1} \\
&\quad - y_t^\top R^{-1} C\hat{x}_{t|t-1} + \hat{x}^\top C^\top R^{-1} C\hat{x}_{t|t-1} + y_t^\top R^{-1} y_t + \hat{x}_{t|t-1}^\top P_{t|t-1}^{-1} \hat{x}_{t|t-1} \\
&= -(y_t - C\hat{x}_{t|t-1})^\top R^{-1} CHC^\top R^{-1}(y_t - C\hat{x}_{t|t-1}) + (y_t - C\hat{x}_{t|t-1}) R^{-1}(y_t - C\hat{x}_{t|t-1}) \\
&\quad - \hat{x}_{t|t-1^\top} H^{-1} \hat{x}_{t|t-1} + \hat{x}_{t|t-1}^\top (C^\top R^{-1} C + P_{t|t-1}^{-1}) \hat{x}_{t|t-1} \\
&= -(y_t - C\hat{x}_{t|t-1})^\top R^{-1} CHC^\top R^{-1}(y_t - C\hat{x}_{t|t-1}) + (y_t - C\hat{x}_{t|t-1}) R^{-1}(y_t - C\hat{x}_{t|t-1}) \\
&= (y_t - C\hat{x}_{t|t-1})^\top [-R^{-1} C(C^\top R^{-1} C + P_{t|t-1}^{-1})^{-1} C^\top R^{-1} + R^{-1}](y_t - C\hat{x}_{t|t-1}) \\
&= (y_t - C\hat{x}_{t|t-1})^\top (R + CP_{t|t-1} C^\top)^{-1} (y_t - C\hat{x}_{t|t-1})
\end{aligned}$$

Where the last equality uses Woodbury's Identity for inverse matrices. Thus, the following equality is proved:

$$(y_t - Cx_t)^T R^{-1}(y_t - Cx_t) + (x_t - \hat{x}_{t|t-1})^T P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1})$$

$$=$$

$$(x_t - h)^\top H^{-1}(x_t - h) + (y_t - C\hat{x}_{t|t-1})^\top (R + CP_{t|t-1} C^\top)^{-1}(y_t - C\hat{x}_{t|t-1})$$

All there is left is to show that:

$$|P_{t|t-1}||R| = |H||R + CP_{t|t-1} C^\top|$$

Using the definition of H^{-1}, properties of determinants of inverse matrices and the multiplicativity of determinants when the dimensions are the same, proving the following equality will suffice:

$$|P_{t|t-1}||H|^{-1} = |R + CP_{t|t-1} C^\top||R|^{-1}$$

$$|P_{t|t-1} H^{-1}| = |RR^{-1} + CP_{t|t-1} C^\top R^{-1}|$$

$$|P_{t|t-1}(P_{t|t-1}^{-1} + C^T R^{-1} C)| = |RR^{-1} + CP_{t|t-1} C^\top R^{-1}|$$

$$|I_m + P_{t|t-1} C^\top R^{-1} C| = |I_n + CP_{t|t-1} C^\top R^{-1}|$$

Let $U = P_{t|t-1} C^\top R^{-1}$
Then, we have

$$|I_m + P_{t|t-1} C^\top R^{-1} C| = |I_m + UC|$$

and

$$|I_n + CP_{t|t-1}C^\top R^{-1}| = |I_n + CU|$$

By the Weinstein–Aronszajn Identity, we know

$$|I_m + P_{t|t-1}C^\top R^{-1}C| = |I_n + CP_{t|t-1}C^\top R^{-1}|$$

Which implies

$$|P_{t|t-1}||R| = |H||R + CP_{t|t-1}C^\top|$$

Thus:

$$\pi(y_t|y_{t-1}) = \frac{\exp\Big\{-\frac{1}{2}(y_t - C\hat{x}_{t|t-1})^\top (R + CP_{t|t-1}C^\top)^{-1}(y_t - C\hat{x}_{t|t-1})\Big\}}{|R + CP_{t|t-1}C^\top|^{\frac{1}{2}} \cdot (2\pi)^{\frac{m}{2}}} \cdot \int_{\mathbb{R}^n} \mathcal{G}(x_t; h, H)dx_t$$

Where $\int_{\mathbb{R}^n} \mathcal{G}(x_t; h, H) = \int_{\mathbb{R}^n} \frac{\exp\{-\frac{1}{2}(x_t-h)^\top H^{-1}(x_t-h)\}}{|H|^{\frac{1}{2}}\cdot(2\pi)^{\frac{n}{2}}}dx_t = 1$ Then, the conditional distribution of y_t given the observations y_{t-1} is:

$$\pi(y_t|y_{t-1}) \sim \mathcal{N}(C\hat{x}_{t|t-1}, R + CP_{t|t-1}C^\top)$$

9.1 Partial Derivatives w.r.t. Parameters

Here, we present the partial derivatives with respect to each parameter, entry by entry.

Partial Log-likelihood w.r.t. A

$$\frac{\partial \ell}{\partial A_{ij}} = -\frac{1}{2}\sum_{t=1}^{N} \operatorname{Tr}\left(\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial A_{ij}}\right) - \frac{1}{2}\sum_{t=1}^{N}\left[\frac{\partial \mathbf{e}_t^T}{\partial A_{ij}}\mathbf{F}_t\mathbf{e}_t - \mathbf{e}_t^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial A_{ij}}\mathbf{F}_t^{-1}\mathbf{e}_t + \mathbf{e}_t^T\mathbf{F}_t\frac{\partial \mathbf{e}_t}{\partial A_{ij}}\right]$$

$$\frac{\partial \mathbf{e}_t}{\partial A_{ij}} = -\mathbf{C}J_{\mathbf{A}}^{(ij)}\hat{\mathbf{X}}_{t-1|t-1} - \mathbf{C}\mathbf{A}\frac{\partial \hat{\mathbf{X}}_{t|t-1}}{\partial A_{ij}}$$

$$\frac{\partial \mathbf{F}_t}{\partial A_{ij}} = \mathbf{C}\frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial A_{ij}}\mathbf{C}^T$$

$$\frac{\partial \hat{\mathbf{X}}_{t|t-1}}{\partial A_{ij}} = J_{\mathbf{A}}^{(ij)}\hat{\mathbf{X}}_{t-1|t-1} + \mathbf{A}\frac{\partial \hat{\mathbf{X}}_{t-1|t-1}}{\partial A_{ij}}$$

$$\frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial A_{ij}} = J_{\mathbf{A}}^{(ij)}\mathbf{\Sigma}_{t-1|t-1}\mathbf{A}^T + \mathbf{A}\frac{\partial \mathbf{\Sigma}_{t-1|t-1}}{\partial A_{ij}}\mathbf{A}^T + \mathbf{A}\mathbf{\Sigma}_{t-1|t-1}J_{\mathbf{A}}^{(ji)}$$

$$\frac{\partial \hat{\mathbf{X}}_{t|t}}{\partial A_{ij}} = \frac{\partial \hat{\mathbf{X}}_{t|t-1}}{\partial A_{ij}} + \frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial A_{ij}}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{e}_t - \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial A_{ij}}\mathbf{F}_t^{-1}\mathbf{e}_t + \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{e}_t}{\partial A_{ij}}$$

$$\frac{\partial \mathbf{\Sigma}_{t|t}}{\partial A_{ij}} = \frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial A_{ij}} - \frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial A_{ij}}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{C}\mathbf{\Sigma}_{t|t-1} + \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial A_{ij}}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{C}\frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial A_{ij}}$$

Partial w.r.t. Q

$$\frac{\partial \ell}{\partial Q_{ij}} = -\frac{1}{2}\sum_{t=1}^{N} \mathrm{Tr}\left(\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial Q_{ij}}\right) - \frac{1}{2}\sum_{t=1}^{N}\left[\frac{\partial \mathbf{e}_t^T}{\partial Q_{ij}}\mathbf{F}_t\mathbf{e}_t - \mathbf{e}_t^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial Q_{ij}}\mathbf{F}_t^{-1}\mathbf{e}_t + \mathbf{e}_t^T\mathbf{F}_t\frac{\partial \mathbf{e}_t}{\partial Q_{ij}}\right]$$

$$\frac{\partial \mathbf{e}_t}{\partial Q_{ij}} = -\mathbf{C}\mathbf{A}\frac{\partial \hat{\mathbf{X}}_{t|t-1}}{\partial Q_{ij}}$$

$$\frac{\partial \mathbf{F}_t}{\partial Q_{ij}} = \mathbf{C}\frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial Q_{ij}}\mathbf{C}^T$$

$$\frac{\partial \hat{\mathbf{X}}_{t|t-1}}{\partial Q_{ij}} = \mathbf{A}\frac{\partial \hat{\mathbf{X}}_{t-1|t-1}}{\partial Q_{ij}}$$

$$\frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial Q_{ij}} = \mathbf{A}\frac{\partial \mathbf{\Sigma}_{t-1|t-1}}{\partial Q_{ij}}\mathbf{A}^T + \mathbf{B}J_{\mathbf{Q}}^{(ij)}\mathbf{B}^T$$

$$\frac{\partial \hat{\mathbf{X}}_{t|t}}{\partial Q_{ij}} = \frac{\partial \hat{\mathbf{X}}_{t|t-1}}{\partial Q_{ij}} + \mathbf{\Sigma}_{t|t-1}\frac{\partial \mathbf{C}^T}{\partial Q_{ij}}\mathbf{F}_t^{-1}\mathbf{e}_t - \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial Q_{ij}}\mathbf{F}_t^{-1}\mathbf{e}_t + \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{e}_t}{\partial Q_{ij}}$$

$$\frac{\partial \mathbf{\Sigma}_{t|t}}{\partial Q_{ij}} = \frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial Q_{ij}} - \frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial Q_{ij}}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{C}\mathbf{\Sigma}_{t|t-1} + \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial Q_{ij}}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{C}\frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial Q_{ij}}$$

Partial w.r.t. R

$$\frac{\partial \ell}{\partial R_{ij}} = -\frac{1}{2}\sum_{t=1}^{N} \mathrm{Tr}\left(\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial R_{ij}}\right) - \frac{1}{2}\sum_{t=1}^{N}\left[\frac{\partial \mathbf{e}_t^T}{\partial R_{ij}}\mathbf{F}_t\mathbf{e}_t - \mathbf{e}_t^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial R_{ij}}\mathbf{F}_t^{-1}\mathbf{e}_t + \mathbf{e}_t^T\mathbf{F}_t\frac{\partial \mathbf{e}_t}{\partial R_{ij}}\right]$$

$$\frac{\partial \mathbf{e}_t}{\partial R_{ij}} = -\mathbf{C}\mathbf{A}\frac{\partial \hat{\mathbf{X}}_{t-1|t-1}}{\partial R_{ij}}$$

$$\frac{\partial \mathbf{F}_t}{\partial R_{ij}} = \mathbf{C}\frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial R_{ij}}\mathbf{C}^T + \mathbf{D}J_{\mathbf{R}}^{(ij)}\mathbf{D}^T$$

$$\frac{\partial \hat{\mathbf{X}}_{t|t-1}}{\partial R_{ij}} = \mathbf{A}\frac{\partial \hat{\mathbf{X}}_{t-1|t-1}}{\partial R_{ij}}$$

$$\frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial R_{ij}} = \mathbf{A}\frac{\partial \mathbf{\Sigma}_{t-1|t-1}}{\partial R_{ij}}\mathbf{A}^T$$

$$\frac{\partial \hat{\mathbf{X}}_{t|t}}{\partial R_{ij}} = \frac{\partial \hat{\mathbf{X}}_{t|t-1}}{\partial R_{ij}} + \mathbf{\Sigma}_{t|t-1}\frac{\partial \mathbf{C}^T}{\partial R_{ij}}\mathbf{F}_t^{-1}\mathbf{e}_t - \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial R_{ij}}\mathbf{F}_t^{-1}\mathbf{e}_t + \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{e}_t}{\partial R_{ij}}$$

$$\frac{\partial \mathbf{\Sigma}_{t|t}}{\partial R_{ij}} = \frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial R_{ij}} - \frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial R_{ij}}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{C}\mathbf{\Sigma}_{t|t-1} + \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial R_{ij}}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t-1}\mathbf{C}^T\mathbf{F}_t^{-1}\mathbf{C}\frac{\partial \mathbf{\Sigma}_{t|t-1}}{\partial R_{ij}}$$