A Machine Learning Model for Predicting the SOFR Term-structure

Yiming Guo

A Thesis
in the Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Degree of
Master of Science in Mathematics and Statistics at
Concordia University
Montréal, Québec, Canada

August 2025

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Mr. Yiming Guo

Entitled: A Machine-Learning Model for Predicting SOFR Term-Structure

and submitted in partial fulfillment of the requirements for the degree of

Master of Science

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Frédéric Godin

_____ Examiner
Dr. Patrice Gaillardetz

_____ Examiner

_____ Thesis Supervisor(s)
Dr. Cody Hyndman

_____ Thesis Supervisor(s)

Approved by _____
Dr. Lea Popovic — Chair of Department or Graduate Program Director

_____
Dr. Pascale Sicotte — Dean of Faculty of Arts & Science

ABSTRACT

A Machine Learning Model for Predicting the SOFR Term-structure

Yiming Guo, M.Sc.
Concordia University, 2025

The Secured Overnight Financing Rate (SOFR) has emerged as the leading benchmark for U.S. dollar-denominated interest rate derivatives, replacing LIBOR due to its transparency and robustness. This thesis develops a comprehensive framework for modeling and forecasting the SOFR term structure using machine learning methods, with a particular focus on CME SOFR futures. We first apply the official CME methodology to construct a daily, piecewise-constant SOFR forward rate curve, incorporating policy-driven discontinuities at Federal Open Market Committee (FOMC) dates. The curve is then fitted to the dynamic Nelson–Siegel (DNS) model, extracting time series of level, slope, and curvature factors. To capture and predict the evolution of these factors, we implement recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) architectures, and integrate a Kalman filter for state-space estimation. The performance of the proposed model is evaluated through out-of-sample forecasts under various loss functions, including mean squared error (MSE) and mean absolute error (MAE), and benchmarks against curve persistence. Our results show that the machine learning approach provides robust short-term forecasts for the SOFR term structure. However, accurately modeling and forecasting abrupt changes around monetary policy announcements remains a challenge, highlighting an important direction for future research. This research offers a practical and robust modeling strategy for interest rate risk management, with direct applications in the pricing and risk assessment of SOFR-linked financial products.

Keywords: SOFR, term structure modeling, CME SOFR Futures, dynamic Nelson-Siegel, Kalman filter, recurrent neural network, long short-term memory, FOMC, yield curve forecasting, machine learning, interest rate derivatives.

Acknowledgments

First, I would like to express my deepest gratitude to my supervisor, Dr. Cody Hyndman, for his clear guidance, encouragement, and invaluable advice throughout the coursework and research. I also thank my parents for their support and encouragement of my decision to pursue a second undergraduate degree in mathematics and then a master's degree. I am grateful to my grandparents for their deep love. Finally, I am thankful to my best friends Crystal and Jeffrey for their constant support and companionship.

Contents

List of Figures

List of Tables

List of Variables

Symbol	Description
P	the final settlement price of SOFR Futures contract
R_{settle}	the settlement rate of SOFR Futures
r_t	the SOFR for day t
r'_t	the published SOFR for day t
$N^{(1)}$	the total number of calendar days in the delivery month
$D^{(1)}$	the set of calendar days for the delivery month
$D^{(1-)}$	the set of calender days for the delivery month before the trading date (excluded)
$D^{(1+)}$	the set of calender days for the delivery month after the trading date (included)
P_{quoted}	the quoted settlement price of SOFR Futures contract
R_{quoted}	the quoted settlement rate of SOFR Futures contract
P_{final}	the final settlement price of SOFR Futures contract
R_{final}	the final settlement rate of SOFR Futures contract
d_t	the number of calender days where r_i is used
$N^{(3)}$	the total number of calendar days in the reference quarter
$D^{(3)}$	the set of business days during the reference quarter
$D^{(3-)}$	the set of business days during the reference quarter before the purchase day (excluded)
$D^{(3+)}$	the set of business days during the reference quarter after the purchase day (included)
CME Group	Chicago Mercantile Exchange Group
P_{SR}	the final price of the SOFR Future
P_{obs}	the selected price of the SOFR Future in an observed interval
V_{obs}	the transaction volume during an observed interval
V_{total}	the transaction volume during all observed intervals
t_0	The date starting to estimate the term SOFR
FOMC	Federal Open Market Committee
$FOMC_k$	the date of the k-th FOMC meeting occurring on or after t_0
R_0	the initial SOFR on t_0
$\mathbf{R}$	vector of R_i, which is the input of the objective function

Continued on next page...

Symbol	Description
ΔR_i	$R_i - R_{i-1}$, the jump size in the SOFR on the day after the i-th FOMC policy date
R_i	the SOFR on the day between the i-th and the $(i+1)$-th FOMC policy date
$SOFR(t;\mathbf{R})$	the SOFR on date t
$p_m^{(1)}$, $p_q^{(3)}$	observed blended price of $SR1$ and $SR3$ futures contracts
$\hat{p}_m^{(1)}(\mathbf{R})$, $\hat{p}_m^{(3)}(\mathbf{R})$	observed blended price of $SR1$ and $SR3$ futures contracts
$w_m^{(1)}$, $w_q^{(3)}$	the weighting parameters for pricing errors of $SR1$ and $SR3$ futures contracts
λ_C	the weighting parameter for the penalty function in CME method
$T_m^{(1)}$	the set of calendar days for the m-th month
$N_m^{(1)}$	the total number of calendar days in the m-th month
$T_0^{(1+)}$	the calendar days after t_0 (included) in the current month
$T_0^{(1-)}$	the calendar days before t_0 (excluded) in the current month
$T_q^{(3)}$	set of Business Days for the q-th reference quarter
$N_q^{(3)}$	total number of calendar days in the q-th reference quarter
$T_q^{(3+)}$	the business days passed in the current reference quarter, after the date starting to estimate the term SOFR (included)
$T_q^{(3-)}$	the business days before t_0 (excluded) in the reference quarter
T	terminal date where the term rate is estimated
$\tilde{T}(T)$	the set of Business Days from the term start date to date T days in the future
b_τ	the load parameters in DNS model
L	the level factor
S	the slope factor
C	the curvature factor
λ	the delay parameter
a	the activation function
$\mathfrak{W}$,$\mathfrak{b}$	the weights and bias in the dense layer
S'	the number of samples in the input layer
L'	the sequential length in the input layer
F'	the size of the feature in the input layer
(c,h)	c is the output and h is the hidden state
f_t, i_t, o_t	the forget gate, the input gate and the output gate
N	the number of input features in the mode
H	the number of hidden units in the model
$\mathcal{W},\mathcal{W}'$	the weight matrices
$\mathbf{b}$	the bias vector
$\kappa(c_t^I),\theta(c_t^I),\sigma(c_t^I)$	three dense layers
τ	the tenor
ϵ_t	the noise between the observation and the model

Continued on next page...

Symbol	Description
U_t	the covariance matrix of noise
Q_k	the conditional variance of X_{k+1}, $\mathrm{Var}[X_{k+1} \mid \mathcal{F}_k]$
E_k	the eigenvector matrix of κ_k
ζ	the eigenvalues of κ_k
V_k	the diagonal matrix with the eigenvalues ζ of κ_k on the diagonal
$L_2(\cdot)$	the mean squared error loss function
$L_1(\cdot)$	the mean absolute error loss function
$\hat{\vartheta}$	the optimal parameter vector that minimizes the loss function
α_k	the learning rate
m_t	biased first moment estimate
v_t	biased second raw moment estimate
g_t	the gradient of the loss function
g_t^2	the elementwise square $g_t \odot g_t$

Introduction

The Secured Overnight Financing Rate (SOFR) is a benchmark interest rate that reflects the cost of borrowing cash overnight while using U.S. Treasury securities as collateral. It is based on data collected from the repurchase agreement market. A repurchase agreement, or repo, is a short-term borrowing agreement. It allows one party to sell securities and agree to buy them back later, often the next day, at a higher price. The SOFR is published daily by the Federal Reserve Bank of New York and has become the preferred replacement for the London Interbank Offered Rate (LIBOR) in U.S. financial markets.

The SOFR was introduced to address the problem of the reliability and stability of LIBOR, which was previously the global benchmark for short-term interest rates. LIBOR was based on the estimations of the leading banks in London, rather than the actual transactions, which makes it vulnerable to manipulation. In contrast, the SOFR is an overnight and backward-looking rate. It is based on the actual transactions in U.S. Treasury repo market with a volume of more than 2 trillion dollars per day, which makes it more robust and transparent (Huggins and Schaller 2022).

The SOFR is computed based on transactions in U.S. Treasury repo market, which consists of three primary segments: the Tri-Party Repo (TPR) Market, with transactions collected from the Bank of New York Mellon; the General Collateral Financing (GCF) Repo Market, with transactions obtained from the Office of Financial Research (OFR); and the Bilateral Repo Market, which includes transactions cleared via the Delivery-Versus-Payment (DVP) service offered by the Fixed Income Clearing Corporation (FICC). The following are details of the data collection from the Federal Reserve Bank of New York (2025). In the TPR market, the Bank of New York Mellon collects specific-counterparty tri-party general collateral repo transactions secured by Treasury securities, excluding transactions in which the Federal Reserve is a counterparty. The specific counterparty means that at the time of trade, the counterparties know the identity of each other. In the DVP repo market, rather than choosing from a broader set of acceptable collateral as seen in the tri-party repo market, counterparties engage in the transactions where they specify the exact security to be delivered and received. This distinction allows the DVP market to serve as a mechanism for temporarily acquiring specific securities. If cash providers are willing to accept a lower return to obtain a particular security, the repo for specific-issue collateral may be traded at a rate below the general repos. These underlying securities are considered as "special". To minimize the impact of transactions on the measurement of the general cost of financing, 20 percent of the lowest-rate transaction volume from the DVP segment is excluded daily. Each day, the New York Fed carefully reviews the transaction data to identify any entries that may need to be excluded. This includes transactions that are not conducted at arm's

length, as well as those that seem anomalous or potentially erroneous. When necessary, the New York Fed may apply expert judgment to determine whether certain transactions should be excluded from the final rate calculation.

Once the data is collected from the three sources, the SOFR is determined using a volume-weighted median, which represents the rate at which transactions reach the 50th percentile of total transaction volume. This calculation involves sorting transactions in ascending order by rate, computing the cumulative transaction volume, and identifying the rate corresponding to the midpoint (50th percentile) of the total dollar volume. The daily SOFR will be published on the next business day on the New York Fed website around 8:00 a.m. ET, and it is rounded to the nearest basis point. On weekends or U.S. public holidays, the SOFR is not published, and the SOFR for the last business day is applied. In this paper, a business day is defined as any day on which U.S. government securities transactions are settled, as recognized by CME Group and the Federal Reserve Bank of New York. This excludes weekends and designated U.S. public holidays, with holidays defined according to the Securities Industry and Financial Markets Association (SIFMA) U.S. Holiday Schedule, during which Treasury markets are closed.

Federal Open Market Committee (FOMC)

Although the SOFR is derived from the market transactions in U.S. repo market, it is still highly sensitive to the broader monetary policy decisions like most short-term interest rates. The most influential decisions come from the Federal Open Market Committee (FOMC). The FOMC is one of the branches of the Federal Reserve System, see (Board of Governors of the Federal Reserve System 2025). It consists of twelve members: the seven members of the Board of Governors; the president of the Federal Reserve Bank of New York; and the last four will be chosen from the remaining eleven Reserve Bank presidents, who serve on a rotating basis for a one-year term. FOMC holds eight scheduled meetings per year regularly, but additional meetings will be added if necessary. During these sessions, the committee evaluates the current state of economic and financial conditions, decides the appropriate position of the monetary policy, and analyzes the potential risk to its long-term objectives of price stability and sustainable economic growth.

The most critical decision of the meeting is the settlement of a target range of the Federal Funds Rate (FFR) and adjusts the rates: rates that anchor overnight money markets: the Interest on Reserve Balances (IORB), the Overnight Reverse Repurchase Agreement (ON RRP) rate, and the Standing Repo Facility (SRF) rate. IORB is the rate the Federal Reserve pays to banks based on reserve balances. Since banks can always earn IORB at the Fed, banks are unlikely to lend at lower rates, which helps keep overnight rates from falling below that value. The ON RRP is the rate that Federal Reserve pays to non-bank counterparties such as money market funds. It gives money funds a safe overnight place to invest, so they will not lend much below that rate, which puts a floor under secured funding rates. The SRF is the rate at which the Federal Reserve lends cash overnight against Treasury or agency collateral. By offering cash at a posted rate, the SRF prevents secured overnight rates from rising too sharply. SOFR is computed from Treasury repo transactions, and hence anchored by these rate.

Related Literature on SOFR modeling

Term-structure modeling based on LIBOR rate benefited from decades of research, resulting in sophisticated models for the term structure of interest rates and techniques for the pricing of derivatives. There are several influential frameworks: the Heath–Jarrow–Morton (HJM) model, introduced by (Heath, Jarrow, and Morton 1992),which models the entire forward rate curve in an arbitrage-free manner, and the short-rate models, such as the Vašíček model (Vašíček 1977), the Cox–Ingersoll–Ross (CIR) model (Cox, Ingersoll Jr, and Ross 1985), and the Hull-White model (Hull and White 1990). The LIBOR Market Model (LMM), also referred to as the BGM model, was introduced by Brace, Gatarek, and Musiela (1997) and became the industry standard for modeling the dynamics of discrete forward LIBOR rates, thereby providing a framework that is consistent with the pricing of caplets and swaptions observed in the market. A comprehensive treatment of these methodologies is provided by Brigo and Mercurio (2007), who surveyed both theoretical and practical aspects of interest rate modeling (also see Bolder 2015). The flexibility and tractability of these classical models have made them the foundation for much of the work in both academic and applied finance, particularly in the context of LIBOR-linked derivatives and risk management.

The discontinuation of LIBOR and the transition to SOFR have motivated the development of new modeling approaches tailored to the features of overnight and nearly risk-free benchmarks. The SOFR is calculated from actual U.S. Treasury repo transactions, and it is inherently more transparent and robust than the LIBOR. However, it differs from LIBOR in important respects: it is more sensitive to day-to-day market liquidity and monetary policy action and is affected by structural features such as regulatory quarter-ends and scheduled FOMC meetings. These distinctions necessitate new modeling considerations, especially for the accurate valuation and risk management of SOFR-linked instruments. Recent literature has advanced several approaches to model the term structure and dynamics of SOFR. Gellert and Schlögl (2021) extended classical short-rate models to SOFR by a term structure model for SOFR in which the short rate is modeled as piecewise constant between scheduled FOMC meetings, with deterministic jumps at known policy dates. This approach captures the observed discontinuities in overnight rates resulting from monetary policy decisions and aligns with the institutional reality of how the Federal Reserve implements policy. Building on this, Schlögl, Skov, and Skovmand (2024) constructed a dynamic term-structure model in which scheduled policy jumps are explicitly embedded in the short-rate process. They demonstrate that failing to model these discontinuities leads to significant model misspecification and increases in prediction errors for SOFR futures and related derivatives.

Fontana, Grbac, and Schmidt (2024) took a forward curve perspective, extending the HJM framework to include stochastic discontinuities at known calendar dates, thereby accounting for the jump risk inherent in overnight reference rates such as SOFR. Their work demonstrates that standard term structure models can be successfully adapted to the realities of modern, transaction-based benchmarks when extended to include both continuous and jump risk components. Nelson–Siegel type models remain popular for their parsimonious representation of the yield curve and ease of estimation. Skov and Skovmand (2021) proposed an arbitrage-free three-factor Nelson–Siegel model to fit the SOFR futures term structure, using a shadow-rate extension to accommodate the zero lower bound. They demonstrated that the SOFR futures alone are sufficiently informative to identify yield curve dynamics

without requiring seasonal corrections or exogenous factors. The dynamic Nelson–Siegel (DNS) model, originally introduced by Diebold and Li (2006) and further developed by Christensen, Diebold, and Rudebusch (2011), remains a robust and versatile framework for yield curve modeling and forecasting, particularly when tailored to accommodate the unique characteristics of overnight benchmark rates.

The adoption of SOFR has also accelerated the use of machine learning and statistical learning methods in yield curve modeling. Kratsios and Hyndman (2019) introduced an arbitrage-free regularization approach designed to learn arbitrage-free factor models within a generalized HJM setting. The framework leverages deep learning to approximate classical factor models while ensuring that the resulting bond price processes are arbitrage-free. Gao (2021) proposed an arbitrage-free recurrent neural network framework for forecasting yield curves, integrating stochastic filtering techniques with deep learning to enhance out-of-sample performance. Bianchi, Büchner, and Tamoni (2020) demonstrated that non-linear machine learning methods, particularly neural networks, significantly improve the out-of-sample predictability of excess bond returns compared to classical linear models. Their results show that neural networks are able to capture complex, non-linear relationships between yields, macroeconomic variables, and bond returns, translating into substantial economic gains in asset allocation exercises.

The SOFR-linked derivatives market remains in a period of transition, as both market participants and researchers adapt to the benchmark's distinctive characteristics and operational realities. This evolving environment necessitates a fundamental re-evaluation of term-structure modeling approaches, especially given the challenges posed by discontinuities of SOFR, sensitivity to monetary policy, and relatively limited historical data. However, to our knowledge, there are very few published papers that focus specifically on SOFR modeling. As a result, there remains a substantial need for further methodological innovation and empirical work to develop robust models capable of supporting the pricing and risk management of SOFR-linked financial products.

Overview

The remainder of this thesis is structured as follows. Chapter 1 introduces the structure, conventions, and pricing mechanics of SOFR futures, with a detailed discussion of the one-month (SR1) and three-month (SR3) contracts traded on the CME Group exchange, thereby laying the foundation for understanding SOFR futures instruments and their role in term structure construction. Chapter 2 describes the official CME methodology for constructing forward-looking term SOFR reference rates from observed futures prices, including input data selection, the application of Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization, and a step-by-step replication of the CME approach, culminating in a piecewise-constant SOFR forward curve that accounts for FOMC meeting dates. Chapter 3 presents the core predictive modeling framework developed in this thesis: building on the dynamic Nelson–Siegel (DNS) representation, the evolution of yield curve factors is modeled using recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) architectures, and a Kalman filter for state-space estimation, with comprehensive details on model design, training, and analysis. Chapter 4 contains the empirical and numerical implementation, detailing data

preparation, model training, hyperparameter selection, and out-of-sample forecast results under various loss functions, along with comprehensive comparisons to benchmark models and analysis of model performance and robustness. Finally, Chapter 5 concludes the thesis by summarizing the main findings and contributions, and suggesting avenues for future research in the modeling and forecasting of SOFR term structures.

Chapter 1

SOFR Futures

SOFR futures contracts are standardized financial instruments designed to facilitate risk management and speculation on short-term interest rate fluctuations where the underlying reference rate is the SOFR (Huggins and Schaller 2022). The three-month SOFR futures and one-month SOFR futures contracts, traded on the CME Group exchange, are the most liquid SOFR-based derivatives, characterized by high trading volumes and tight bid-ask spreads.

1.1 One-Month SOFR Futures (SR1)

An 1-month SOFR futures (SR1) contract is a standardized derivative traded on the CME that provides a transparent benchmark to manage short-term interest rate risk. It allows market participants to hedge or gain exposure to the average the SOFR over a single calendar month. These contracts are cash-settled, with the final settlement price P determined by the simple arithmetic average of the daily SOFR values over the calendar (not business) days of the delivery month. The formula for the final settlement price of a 1-month SOFR futures contract is defined as 100 less the settlement rate R_{settle}, $100 - R_{settle}$. To formalize this, we give the following definition

Definition 1.1.1. *The final settlement rate for a 1-month SOFR futures contract is defined as*

$$R_{settle} = 100 \times \left(\frac{1}{N^{(1)}} \sum_{t \in D^{(1)}} r_t \right) \tag{1.1}$$

where r_t is the SOFR on day t and reported in percentage points, $N^{(1)}$ is the total number of calendar days in the delivery month, and $D^{(1)}$ is the set of those days. And the final settlement price of a 1-month SOFR futures contract is defined as

$$P = 100 - R_{settle} \tag{1.2}$$

The settlement rate R_{settle} denotes the annualized interest rate obtained by taking the simple arithmetic average of all daily SOFR rates during the delivery month. The rate will be rounded to the nearest 1/100th basis point. On weekends and U.S. public holidays, the

SOFR is not published, the most recent published SOFR is carried forward and used for non-business days in the calculation. For SR1 contracts, the delivery month corresponds to the full calendar month. For example, the SR1 contract for January begins on January 1 and ends on January 31. January 1 is always a public holiday in the U.S., so the SOFR is not published. The most recent SOFR will be applied on January 1. For instance, January 1, 2023, is a holiday and December 31, 2022, is a Saturday. Hence, the rate of December 30, 2022, will be applied for January 1, 2023. This convention ensures that every day in the month is represented in the settlement calculation, even if the SOFR is not published on that date. The timeline representation illustrated in Figure 1.1 shows that $R_{settle} = 100 \times \frac{1}{31} \sum_{i=1}^{31} r_i$.

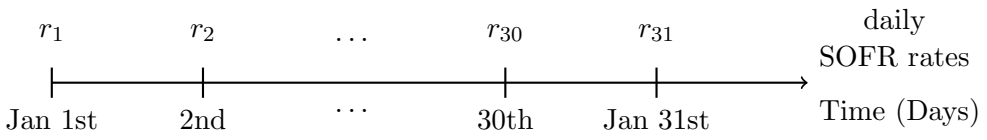


Figure 1.1: Timeline representation of the SOFR calculation for January.

In the CME SOFR futures settlement process, the published settlement rate (R_{settle}) is scaled by 100, so a market-implied interest rate of 5% appears as $R_{\text{settle}} = 5$. Throughout this thesis, when referring to the implied rate, an abuse of notation will be used as $100 - P = R\%$, where P is the quoted futures price and $R\%$ is the market-implied rate. For example, a futures price of \$95.00 implies an average SOFR rate of $100 - 95.00 = 5.00\%$.

This pricing structure also creates a distinct pricing dynamic, particularly when trading the contract in the middle of the delivery month. At any given time within the month, the settlement price is influenced by two components: the known SOFR rates starting from the month start to the day before purchase day, and the future unknown SOFR rates for the remaining days of the month. To formalize this, we introduce the following definition:

Definition 1.1.2. *The settlement rate during the delivery month for a 1-month SOFR futures contract is defined as*

$$R_{settle} = 100 \times \frac{1}{N^{(1)}} \left(\sum_{t \in D^{(1-)}} r'_t + \sum_{t \in D^{(1+)}} r_t \right) \tag{1.3}$$

where r'_t is the published SOFR on day t, $D^{(1-)}$ is the set of calendar days for the delivery month before the trading date (excluded), $D^{(1+)}$ is the set of calendar days for the delivery month after the trading date (included), and $N^{(1)}$ is the total number of calendar days in that month.

Consider an example displayed in Figure 1.2, in which a trader would like to purchase a January SR1 contract on January 15. So far, the SOFR rates have been published for January 1 to 14, which means the term $\sum_{t \in D^{(1-)}} r'_t$ is determined. The remaining SOFR rates for January 15 to 31, denoted by $\sum_{t \in D^{(1+)}} r_t$ in equation (1.3), are not yet known at the time of purchase. However, the current SR1 futures price reflects the market's expectation for the

average SOFR over the entire month, implicitly providing an estimate for the sum of the unknown daily rates.

Suppose the 1-month SOFR future price is quoted at \$95.70 on January 15 and the sum $\sum_{t=1}^{14} r'_t = 0.6$. There are 31 days in January, so $N^{(1)} = 31$ and the corresponding R_{settle} is $100 - 95.70 = 4.30$. Then the market expectation of the sum of the unknown rates $\sum_{t=15}^{31} r_t = R_{settle} \times \frac{N^{(1)}}{100} \times \sum_{t=1}^{14} r'_t = 4.3 \times \frac{31}{100} \times 0.6 = 0.7998$. If there is no FOMC meeting during the rest of the month, the market typically assumes the SOFR stays constant at the futures-implied rate. Hence, $r = 0.7998/17 \approx 0.047047$, which implies the daily SOFR is 4.71% from January 15 to 31.

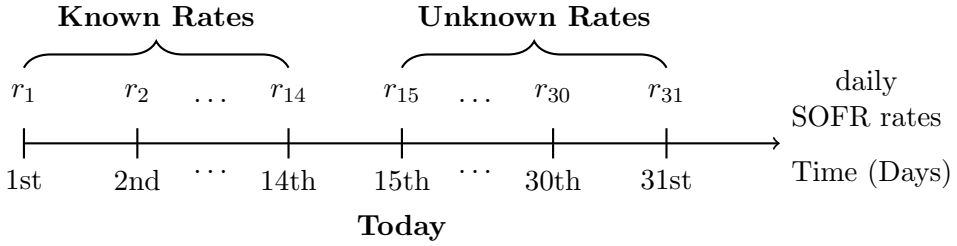


Figure 1.2: Timeline representation of the SOFR calculation for a sample month (January) when trade in the mid of the delivery month

The basis point value (BPV) of an SOFR futures contract is defined as the dollar amount gained or lost for each one basis point (0.01%) movement in the actual rate. For the 1-month SOFR futures (SR1), the BPV is \$41.67 per contract, defined by the CME Group. It means that every one basis point decrease in the settlement rate results in a profit of \$41.67 for the long position. The value is the same as the 30-day federal fund futures. While BPV is formally defined based on rate changes, traders in the futures market typically track profit and loss via changes in the quoted price. Because the price moves inversely to the rate and the quantities are scaled by a factor of 100, every increase of \$0.01 in the futures price corresponds to a decrease of 1 basis point of settlement rate, which leads to a profit of \$41.67 per contract for the long party. Thus, profit or loss for a long position always moves in the same direction as the price, a convention that is both intuitive and standard in futures trading.

The quoted price of an SOFR futures contract is the market price at which the contract is traded, defined as $P_{quoted} = 100 - R_{quoted}$, where R_{quoted} is the corresponding implied quoted rate. At the end of the delivery month, the final settlement price P_{final} is determined as $P_{final} = 100 - R_{final}$, where R_{final} is the realized 1-month SOFR rate for the contract period. For example, consider a trader who enters a long position in a 3-month SOFR futures contract. Suppose the quoted price is \$95.00, implying the quoted rate is $100-95.00 = 5.00\%$. At the end of the delivery period, suppose the final settlement price becomes \$95.06. The realized rate is therefore $100 - 95.06 = 4.94\%$. In this case, the final settlement rate is decreased from 5% to 4.94%, by 6 basis points, compared to the initial quoted rate, and the future price is increased from \$95.00 to \$95.06 by \$0.06. Hence, the long party realizes a profit of $6 \times \$41.67 = \$250.02$.

1.2 Three-Month SOFR Futures (SR3)

The 3-month SOFR futures contract (SR3) is a standardized, cash-settled futures contract that provides market participants with exposure to the average SOFR rates over a standardized 3-month period. Like the 1-month SOFR futures (SR1), the SR3 contract is traded on CME and is cash settled based on the observed SOFR rates during its reference period. The main distinction between SR1 and SR3 lies in the method used to calculate the settlement rate. While the SR1 contract uses the simple arithmetic average of the SOFR rates over a single calendar month, the SR3 contract uses the compounded average of the daily SOFR rates over a reference quarter of three months. The compounding reflects conventions commonly applied to money market instruments with longer maturities and results in a settlement rate that captures the cumulative effect of the daily SOFR movements over the quarter. The contract reference quarter of an SR3 future starts on the third Wednesday (included) of the starting month and ends on the third Wednesday (excluded) of the next quarterly month. We can purchase the 3-month SOFR future starting in any month, but the most liquid SR3 contracts start in March, June, September, and December, aligning with the standard International Monetary Market (IMM) quarterly cycles.

The contract code for SOFR futures, such as "SR3M0", follows a convention in which the prefix ("SR3" or "SR1") specifies the contract type, for 3-month and 1-month SOFR futures respectively, the single-letter code denotes the start month of the reference period, and the final digit indicates the year. This coding applies equally to both SR1 and SR3 contracts. The mapping between code letters and months is shown in Table 1.1. The final digit reflects the last digit of the contract year: "1" for 2021, "2" for 2022, "3" for 2023, and so on. This coding scheme is standard for CME futures and enables market participants to easily identify the contract's delivery (start) month and year. For example, the contract "SR3M0" begins on June 12, 2020 (M = June, 0 = 2020), and ends on September 16, 2020.

Table 1.1: CME Futures Contract Month Codes

Code Letter	**Month**	**Code Letter**	**Month**
F	January	N	July
G	February	Q	August
H	March	U	September
J	April	V	October
K	May	X	November
M	June	Z	December

Definition 1.2.1. *The formula of the final settlement rate of an SR3 contract is defined as*

$$R_{settle} = 100 \times \left(\frac{360}{N^{(3)}} \left(\prod_{t \in D^{(3)}} \left(1 + \frac{r_t \times d_t}{360} \right) - 1 \right) \right) \tag{1.4}$$

where $N^{(3)}$ is the total number of calendar days in the reference quarter, $D^{(3)}$ is the set of business days during the reference quarter, r_t is the SOFR for the day t, d_t is the number of

calendar days where r_i is used. And the final settlement price of a 3-month SOFR futures contract is defined as

$$P = 100 - R_{settle} \tag{1.5}$$

The settlement rate R_{settle} denotes the annualized interest rate obtained by compounding all the SOFR rates during the reference quarter. The rate will be rounded to the nearest 1/100th basis point. Simple interest will be used for all non-business days in the reference quarter using the most recent SOFR. This simple interest represents the compounded rate for the period. After that, the interest for all business days is compounded, and the resulting annualized rate is expressed according to the money market day count convention (Actual/360). Consider an example of compounding the SOFR rates for one week during the reference quarter. From Monday to Friday, the daily SOFR rates are r_1 to r_5, while for the weekends, Saturday and Sunday, the most recent SOFR r_5 will be applied. The final compounded SOFR rate will be $(\Pi_{i=1}^{4}(1+r_i/360))\times(1+3r_5/360)$ as illustrated in Figure 1.3.

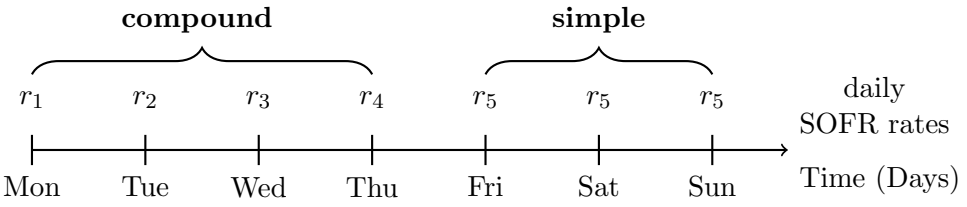


Figure 1.3: Timeline representation of the SOFR calculation for reference quarter

Similar to SR1 contracts, when trading an SR3 contract in the middle of the delivery month, the settlement rate is determined by combining the published SOFR rates up to the date of trade (excluded) and the projected rates for the remaining days.

Definition 1.2.2. *The formula of R_{settle} is formulated as*

$$R_{settle} = 100 \times \frac{360}{N^{(3)}} \left(\left(\prod_{t \in D^{(3-)}} \left(1 + \frac{r'_t \times d_t}{360} \right) \right) \right. \tag{1.6}$$

$$\left. \times \left(\prod_{t \in D^{(3+)}} \left(1 + \frac{r_t \times d_t}{360} \right) \right) - 1 \right) \tag{1.7}$$

where $D^{(3-)}$ is the set of business days during the reference quarter before the purchase day (excluded), $D^{(3+)}$ is the set of business days during the reference quarter after the purchase day (included), and $N^{(3)}$ is the total number of calendar days in the reference quarter.

The basis point value of a 3-month SOFR future is \$25 per contract, which is the same as the 3-month Eurodollar future. This means compared to R_{quoted}, for every basis point the R_{settle} drops, the long party will make a profit of \$25. Since the architecture of the SR3 contract is similar to that of the SR1 contract, the relationship between price and rate

changes remains the same: every increase of \$0.01 in the SR3 futures price corresponds to a one basis point decrease in the implied settlement rate, and yields a profit of \$25*per* contract for the long party. Suppose a trader buys one SR3 contract when the quoted price is \$95.00, implying a futures rate of $100 - 95.00 = 5.00\%$. At the end of the reference period, suppose the contract settles at \$95.03, which corresponds to a final rate of 4.97%. The final settlement rate is decreased by 3 basis points, and the final settlement is increased by \$0.03. Hence the long party realized a profit of $3 \times \$25 = \$75$.

When trading SOFR futures, the mark-to-market (MTM) procedure ensures that all gains and losses are settled on a daily basis. At the end of each trading day, the clearinghouse compares the previous day's settlement price to the current day's settlement price. The profit or loss for each position is determined by the change in the settlement rate, multiplied by the contract value per basis point, \$41.67 for SR1 and \$25 for SR3. This amount, known as variation margin, is credited to or debited from the margin account of each participant. Specifically, when the settlement price increases compared to the previous day, the long position earns a profit equal to the contract value per basis point for each basis point of increase in R_{settle}. Conversely, when the settlement price decreases, the long position incurs a loss of the same amount per basis point. Daily settlement via variation margin ensures that exposures remain fully collateralized and minimizes counterparty risk throughout the life of the contract.

To illustrate, consider an investor who enters a long position in a 3-month SOFR future contract with an initial margin \$2000. Suppose the settlement price at the end of day 1 is \$95.88, which is the initial quoted price. At the end of day 2, the settlement price rises to \$95.91. The increase in price is \$0.03 which corresponds to the decrease of 3 basis points in the rate. Hence, the investor's profit for that day is $3 \times \$25 = \$75$ as variation margin, credited to the margin account. The settlement price declines to \$95.84 at the end of day 3. The drop of \$0.07 leads to an increase of 7 basis points of the rate, resulting in a loss of $7 \times \$25 = \$175$, which is debited from the margin account. On day 4, the price is \$95.85, the investor earns \$25 and similar calculations apply for the subsequent days.

Table 1.2 provides a visualization of this mark-to-market process and summarizes the daily settlement prices, basis point moves in price, daily profit or loss, margin account balances, and the explanations for each day. This mark-to-market procedure is repeated each trading day, with gains credited and losses debited to the margin account. In this way, all profit and loss is settled in cash throughout the life of the contract, ensuring that exposures are continuously collateralized and counterparty risk is minimized until the contract is closed or expires.

In summary, the preceding sections provide an overview of the SOFR, its connection to U.S. monetary policy, and the structure of SOFR futures contracts. This shows the significance and challenges involved in modeling the SOFR term structure. Chapter 2 outlines the methodology and procedures for building the SOFR term structure from the SOFR futures data.

Table 1.2: Daily Mark-to-Market Example for a Long Position in 3-month SOFR futures

Day	Settlement Price ($)	Change	Daily P/L ($)	Margin Account ($)
1	95.88	–	–	2,000.00
2	95.91	+3	+75.00	2,075.00
3	95.84	−7	-175.00	1,900.00
4	95.85	+1	+25.00	1,925.00
5	95.83	−2	-50.00	1,875.00
6	...	...	...	...

Chapter 2

CME Term SOFR Reference Rates Methodology

In this chapter, we introduce the CME Term SOFR Reference Rates Methodology. This methodology provides a systematic approach for building a forward-looking and piecewise-constant SOFR term structure with the data from SOFR futures contracts (CME Group 2025). The method involves selecting input data, applying Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization techniques, building the SOFR curve, and generating the term rates for various maturities.

2.1 Input Data

The CME methodology combines 1-month and 3-month SOFR futures contracts. The rates will estimate tenors of 1 month, 3 months, 6 months, and 12 months. To fully cover the tenors, the algorithm uses 13 consecutive 1-month SOFR futures (SR1) and 5 consecutive quarterly (Mar, Jun, Sept, Dec) 3-month SOFR futures (SR3). The sampling market hours of the price of the SR1 and SR3 contracts are 7:00 a.m. CT until 2:00 p.m. CT. They are divided into 14 intervals of 30 minutes each. A set of Volume Weighted Average Prices (VWAP) is calculated based on transaction prices recorded within each observation interval. Also, they record a snapshot of executable bid/ask prices at a random moment during the observation interval. In each observation interval, a set of candidate solution prices is constructed for the relevant SOFR futures contracts based on available market data. For each contract, a seed value is first determined on a VWAP or the midpoint of the bid/ask price of the snapshot if there are no transactions occurring. A candidate solution price is the price within 5 increments of 0.01 basis points of the seed value. If there are transactions during the interval, the candidate solution will be the VWAP; if there is not, then the candidate solution will be the midpoint of the ask and bid price on that snapshot. For the first 7 SR1 contracts and the 5 SR3 contracts, a linear optimization algorithm selects the final price from the candidate set. This algorithm minimizes deviations from the VWAPs while ensuring the selected prices are consistent with the maximum number of outright, calendar spread, and butterfly market quotations, subject to bid/ask and spread constraints. For the remaining SR1 contracts (eighth to thirteenth), a boundary rule is applied: the selected price is the

VWAP if it is within the bid/ask bounds, or otherwise, the closest boundary value. After obtaining the selected price for each interval, the final price will be calculated by a weighted average of the selected price for each contract, where the weight is the transaction volume in each observed interval over the total transaction volume.

$$P_{SR} = \sum P_{obs} \times \frac{V_{obs}}{V_{total}} \tag{2.1}$$

where P_{SR} is the final price of the SOFR futures, P_{obs} is the selected price of the SOFR futures in an observed interval, V_{obs} is the volume of transactions during the observed interval, and V_{total} is the volume of transactions during all observed intervals. The methodology by which the selected and final prices are determined is not reproduced in this thesis, the analysis applies the official final prices as directly provided by CME Group.

As a typical example, from the official CME Term SOFR Constituent File, Figures 2.1, 2.2, and 2.3 show the actual Excel data files provided by CME for a single sample day (February 29, 2024). These figures illustrate the organization of interval input data, contract-specific interval prices, and the resulting official end-of-day projection file. Figure 2.1 displays the head (first few rows) of the CME Term SOFR Interval Input Data file for this day. The complete file records all observed intervals and contracts, containing selected prices, transaction volumes, market quotes, and timestamps for every 30-minute interval across the entire trading day. Figure 2.2 presents the interval data for the SR3H4 contract across all 14 observation intervals on the same date, illustrating the process for a single contract. Figure 2.3 shows the official CME Final Projection Inputs file. For each SOFR futures contract on the sample day, the file reports the final VWAP as well as the date of the next scheduled FOMC meeting.

It should be noted that the detailed internal methodology by which CME determines the selected price for each interval is not fully disclosed. Therefore, while interval-level prices are provided, it is not always possible to precisely reproduce their calculation. Furthermore, even if the selected prices for each interval are known, a direct calculation of the final VWAP using these values may still not exactly match the official CME published price. For example, in the sample Excel file, the VWAP calculated from the selected prices and volumes is 94.68609, while the official final VWAP published by CME is 94.6859. Any minor discrepancies can be attributed to rounding conventions, data eligibility, or other undisclosed adjustments made by CME. For this reason, all analyzes in this thesis rely directly on the Final VWAP in Figure 2.3 as published by CME.

Date	Interval	Contract	Selected Price	Initial Price	Volume	Bid	Ask	Bid/Ask Snapshot Time
29/02/2024	Interval 1	SR1G4	94.6912	94.6912	0	94.69	94.6925	7:29:49
29/02/2024	Interval 1	SR1H4	94.6899	94.69	684	94.6875	94.69	7:29:49
29/02/2024	Interval 1	SR1J4	94.6849	94.6848	2777	94.68	94.69	7:29:49
29/02/2024	Interval 1	SR1K4	94.7224	94.7224	430	94.72	94.725	7:29:49
29/02/2024	Interval 1	SR1M4	94.7907	94.7907	508	94.79	94.795	7:29:49
29/02/2024	Interval 1	SR1N4	94.8361	94.8361	888	94.835	94.84	7:29:49
29/02/2024	Interval 1	SR1Q4	94.9661	94.9661	217	94.96	94.97	7:29:49
29/02/2024	Interval 1	SR1U4	95.0333	95.0333	6	95.03	95.035	7:29:49
29/02/2024	Interval 1	SR1V4	95.1263	95.1263	42	95.125	95.135	7:29:49
29/02/2024	Interval 1	SR1X4	95.2263	95.2265	17	95.225	95.235	7:29:49
29/02/2024	Interval 1	SR1Z4	95.3228	95.3228	9	95.32	95.33	7:29:49
29/02/2024	Interval 1	SR1F5	95.4325	95.4325	0	95.425	95.44	7:29:49
29/02/2024	Interval 1	SR1G5	95.555	95.555	0	95.545	95.565	7:29:49
29/02/2024	Interval 1	SR3Z3	94.645	94.645	1181	94.645	94.6475	7:29:49
29/02/2024	Interval 1	SR3H4	94.6812	94.6812	2869	94.68	94.6825	7:29:49
29/02/2024	Interval 1	SR3M4	94.8727	94.8727	17169	94.87	94.875	7:29:49
29/02/2024	Interval 1	SR3U4	95.155	95.154	10667	95.155	95.16	7:29:49
29/02/2024	Interval 1	SR3Z4	95.46	95.4566	23238	95.46	95.465	7:29:49
29/02/2024	Interval 2	SR1G4	94.6912	94.6912	0	94.69	94.6925	7:54:21
29/02/2024	Interval 2	SR1H4	94.69	94.69	685	94.69	94.6925	7:54:21
29/02/2024	Interval 2	SR1J4	94.6862	94.6861	1039	94.685	94.69	7:54:21
29/02/2024	Interval 2	SR1K4	94.7285	94.7285	2285	94.725	94.73	7:54:21
29/02/2024	Interval 2	SR1M4	94.81	94.8104	413	94.805	94.81	7:54:21
29/02/2024	Interval 2	SR1N4	94.8555	94.8555	1713	94.855	94.865	7:54:21
29/02/2024	Interval 2	SR1Q4	94.995	94.9936	901	94.995	95	7:54:21
29/02/2024	Interval 2	SR1U4	95.07	95.073	197	95.065	95.07	7:54:21
29/02/2024	Interval 2	SR1V4	95.1744	95.1744	423	95.17	95.175	7:54:21
29/02/2024	Interval 2	SR1X4	95.28	95.2809	540	95.275	95.28	7:54:21
29/02/2024	Interval 2	SR1Z4	95.38	95.3842	159	95.37	95.38	7:54:21
29/02/2024	Interval 2	SR1F5	95.485	95.4893	160	95.48	95.485	7:54:21
29/02/2024	Interval 2	SR1G5	95.62	95.62	100	95.605	95.62	7:54:21
29/02/2024	Interval 2	SR3Z3	94.645	94.645	169	94.645	94.6475	7:54:21
29/02/2024	Interval 2	SR3H4	94.6851	94.6851	24347	94.685	94.6875	7:54:21
29/02/2024	Interval 2	SR3M4	94.9	94.8919	50662	94.9	94.905	7:54:21
29/02/2024	Interval 2	SR3U4	95.2	95.1887	27608	95.2	95.205	7:54:21
29/02/2024	Interval 2	SR3Z4	95.515	95.4957	58954	95.515	95.52	7:54:21

Figure 2.1: Excerpt of CME Term SOFR Interval Input Data for 2024-02-29.

Date	Interval	Contract	Selected Price	Initial Price	Volume	Bid	Ask	Bid/Ask Snapshot Time
29/02/2024	Interval 1	SR3H4	94.6812	94.6812	2869	94.68	94.6825	7:29:49
29/02/2024	Interval 2	SR3H4	94.6851	94.6851	24347	94.685	94.6875	7:54:21
29/02/2024	Interval 3	SR3H4	94.6874	94.6872	9597	94.685	94.6875	8:23:25
29/02/2024	Interval 4	SR3H4	94.6875	94.6877	6955	94.6875	94.69	8:55:14
29/02/2024	Interval 5	SR3H4	94.6887	94.6887	858	94.6875	94.69	9:22:51
29/02/2024	Interval 6	SR3H4	94.6892	94.6891	6360	94.6875	94.69	9:52:00
29/02/2024	Interval 7	SR3H4	94.6875	94.6876	7480	94.685	94.6875	10:26:39
29/02/2024	Interval 8	SR3H4	94.6875	94.6874	10742	94.685	94.6875	10:59:17
29/02/2024	Interval 9	SR3H4	94.685	94.685	12293	94.6825	94.685	11:28:56
29/02/2024	Interval 10	SR3H4	94.6827	94.6829	1328	94.6825	94.685	11:51:58
29/02/2024	Interval 11	SR3H4	94.6845	94.6847	1699	94.6825	94.685	12:24:49
29/02/2024	Interval 12	SR3H4	94.685	94.685	3837	94.685	94.6875	12:53:29
29/02/2024	Interval 13	SR3H4	94.6871	94.6871	1348	94.685	94.6875	13:25:15
29/02/2024	Interval 14	SR3H4	94.6849	94.6849	6801	94.6825	94.685	13:53:59

Figure 2.2: Excerpt of Interval Data for Contract SR3H4 on 2024-02-29.

Date	Description	Contract	Value
29/02/2024	VWAP Final	SR1G4	94.6912
29/02/2024	VWAP Final	SR1H4	94.6896
29/02/2024	VWAP Final	SR1J4	94.6865
29/02/2024	VWAP Final	SR1K4	94.7299
29/02/2024	VWAP Final	SR1M4	94.8072
29/02/2024	VWAP Final	SR1N4	94.8571
29/02/2024	VWAP Final	SR1Q4	94.9963
29/02/2024	VWAP Final	SR1U4	95.0671
29/02/2024	VWAP Final	SR1V4	95.1708
29/02/2024	VWAP Final	SR1X4	95.2742
29/02/2024	VWAP Final	SR1Z4	95.3706
29/02/2024	VWAP Final	SR1F5	95.4806
29/02/2024	VWAP Final	SR1G5	95.6097
29/02/2024	VWAP Final	SR3Z3	94.6459
29/02/2024	VWAP Final	SR3H4	94.6859
29/02/2024	VWAP Final	SR3M4	94.8991
29/02/2024	VWAP Final	SR3U4	95.199
29/02/2024	VWAP Final	SR3Z4	95.5148
29/02/2024	FOMC		20/03/2024
29/02/2024	FOMC		01/05/2024
29/02/2024	FOMC		12/06/2024
29/02/2024	FOMC		31/07/2024
29/02/2024	FOMC		18/09/2024
29/02/2024	FOMC		07/11/2024
29/02/2024	FOMC		18/12/2024
29/02/2024	FOMC		29/01/2025
29/02/2024	FOMC		19/03/2025
29/02/2024	FOMC		30/04/2025
29/02/2024	FOMC		18/06/2025
29/02/2024	FOMC		30/07/2025
29/02/2024	FOMC		24/09/2025

Figure 2.3: Excerpt of CME Final Projection Inputs on 2024-02-29.

2.2 Calculation Methodology

Suppose that the CME Term SOFR Reference Rate is determined as of the date t_0, and t represents a business day after t_0, as defined by the SIFMA US Holiday Schedule. CME assumed that the SOFR is constant between 2 FOMC meetings. The overnight SOFR for the date t can be computed as:

$$SOFR(t; \mathbf{R}) = R_0 + \sum_{i=1}^{k} (\Delta R_i) \times \mathbf{1}_{\{t > FOMC_k\}} \tag{2.2}$$

where $FOMC_k$ is the date of the k-th FOMC meeting that occurs on or after t_0, R_0 is the initial SOFR on t_0, $\Delta R_i = R_i - R_{i-1}$ is the jump size in the SOFR on the day after the i-th FOMC policy date, R_i is the SOFR on the day between the i-th and the $(i+1)$-th FOMC policy date, $SOFR(t;\mathbf{R})$ is the SOFR on date t, $\mathbf{R} = (R_0, R_1, R_2, ..., R_k)$ and k is the index of the last relevant FOMC meeting date. And $\mathbf{1}_{\{\}}$ is the indicator function. The objective function is

$$\min_{\mathbf{R}} \left[\sum_{m=0}^{12} w_m^{(1)} \left(p_m^{(1)} - \hat{p}_m^{(1)}(\mathbf{R}) \right)^2 + \sum_{q=0}^{4} w_q^{(1)} \left(p_q^{(3)} - \hat{p}_q^{(3)}(\mathbf{R}) \right)^2 \right]^{\frac{1}{2}} + \lambda_C \left[\sum_k (\Delta R_k)^2 \right]^{\frac{1}{2}} \tag{2.3}$$

where the subscripts m and q are the reference month m and the quarter q of the $SR1$ and $SR3$ contracts, respectively, $p_m^{(1)}$ and $p_q^{(3)}$ are the blended observed prices of the $SR1$ and $SR3$ contracts, $\hat{p}_m^{(1)}(\mathbf{R})$ and $\hat{p}_q^{(3)}(\mathbf{R})$ are the implied values of the $SR1$ and $SR3$ contracts, $w_m^{(1)}$ and $w_q^{(3)}$ are the weighting parameters for the pricing errors of the $SR1$ and $SR3$ contracts. In this case, $w_m^{(1)} = w_q^{(3)} = 1/18$, and λ is the weighting parameter for the penalty function, $\lambda_C = \frac{0.01}{\sqrt{k}}$, k is the number of FOMC meetings scheduled in the period. CME Group (2025) uses the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization method to solve $\mathbf{R}$ which minimizes the objective function (2.3).

For SR1 contracts, if the current date is not in the reference month ($m > 0$), then the implied price will be calculated using the projected SOFR rates. And if the current date is in the reference month ($m = 0$), the implied price can be calculated using the published SOFR rates and the projected SOFR rates

$$\hat{p}_m^{(1)}(\mathbf{R}) = \begin{cases} 100 \times \left(1 - \frac{1}{N_m^{(1)}} \sum_{t \in T_m^{(1)}} SOFR(t;\mathbf{R}) \right), & m > 0 \\ 100 \times \left[1 - \frac{1}{N_0^{(1)}} \left(\sum_{t \in T_0^{(1-)}} r'_t + \sum_{t \in T_0^{(1+)}} SOFR(t;\mathbf{R}) \right) \right], & m = 0 \end{cases} \tag{2.4}$$

where $T_m^{(1)}$ is the set of calendar days for the m-th month, and N_m^1 is the total number of calendar days in the m-th month; $T_m^{(1+)} = \{t \in T_m^{(1)} | t \geq t_0\}$ is the set of business days that have passed in the current reference quarter after t_0 (included) during which the SOFR rates are published. $T_m^{(1-)} = \{t \in T_m^1 | t < t_0\}$ is the set of business days before t_0 (excluded) in the reference quarter, and r'_t is the published SOFR for the date t.

For SR3 contracts, whose reference quarter is not the current quarter ($q > 0$), the implied value depends only on the projected overnight SOFR rates. And if the reference quarter is the current quarter ($q = 0$), the implied value can be calculated using published SOFR rates and projected overnight SOFR rates

$$\hat{p}_q^{(3)}(\mathbf{R}) = \begin{cases} 100 \times \left(1 - \frac{360}{N_q^{(3)}} \left[\prod_{t \in T_q^{(3)}} \left(1 + \frac{SOFR(t;\mathbf{R}) \times d_t}{360}\right) - 1\right]\right), & q > 0 \\ 100 \times \left[1 - \frac{360}{N_0^{(3)}} \left(\prod_{t \in T_q^{(3-)}} \left(1 + \frac{r'_t \times d_t}{360}\right) \prod_{t \in T_q^{(3+)}} \left(1 + \frac{SOFR(t;\mathbf{R}) \times d_t}{360}\right) - 1\right)\right], & q = 0 \end{cases} \tag{2.5}$$

where $T_q^{(3)}$ is the set of business days for the q-th reference quarter, $N_q^{(3)}$ is the total number of calendar days in the q-th reference quarter, and d_t is the number of calendar days from date t to its next Business Day; $T_q^{(3+)} = \{t \in T_q^{(3)} | t \geq t_0\}$ is the set of business days that have passed in the current reference quarter after t_0 (included) during which the SOFR rates are published. $T_q^{(3-)} = \{t \in T_q^{(3)} | t < t_0\}$ is the set of business days before t_0 (excluded) in the reference quarter, and r'_t is the published SOFR for the date t.

The term rate from the projected SOFR is the compounded rate over tenors of one, three, six, and twelve months. The compounded term rate is

$$TR(T) = \frac{360}{T} \times \left[\prod_{t \in \tilde{T}(T)} \left(1 + \frac{SOFR(t;\mathbf{R}) \times d_t}{360}\right) - 1\right] \tag{2.6}$$

where T is the terminal date, $\tilde{T}(T)$ is the set of business days from the start date to the date T days in the future, t is a business day in the set $\tilde{T}(T)$, d_t is the number of calendar days from the date t to its next business day, and $SOFR(t, \mathbf{R})$ is the overnight SOFR as of date t.

2.3 Broyden-Fletcher-Goldfarb-Shanno (BFGS) Optimization

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, introduced by Broyden (1970), Fletcher (1970), Goldfarb (1970), and Shanno (1970) is an iterative optimization algorithm used to solve unconstrained non-linear optimization problems. It belongs to the family of quasi-Newton methods, which approximate the second-order derivative, the Hessian matrix, rather than computing it explicitly. The BFGS algorithm is presented in Algorithm 1 (Nocedal and Wright 2006).

There are several reasons why CME uses the BFGS method. Rather than directly computing the Hessian matrix, which is computationally expensive, BFGS iteratively approximates the inverse Hessian. The SOFR term rate estimation problem involves an input of $\mathbf{R}$, which leads to a multi-dimensional search space. First-order methods, such as gradient descent, converge too slowly. However, BFGS leverages curvature information to achieve a superlinear convergence rate. In contrast, optimization methods that do not use curvature information can be inefficient, particularly for ill-conditioned problems. The BFGS algorithm improves the search process by reducing unnecessary oscillations and maintaining a

Algorithm 1 BFGS Algorithm for Minimizing an Objective Function $f(x)$

Given: Objective function $f(x)$ to minimize.
Input: Initial guess x_0, initial inverse Hessian approximation $B_0 = I$, tolerance ϵ.
Initialize: Set iteration index $k = 0$.
while $||\nabla f(x_k)|| > \epsilon$ **do**
    Compute descent direction:

$$p_k = -H_k \nabla f(x_k)$$

    Perform line search to determine step size α_k.
    Update position:

$$s_k = \alpha_k p_k$$

$$x_{k+1} = x_k + s_k$$

    Define gradient difference:

$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

    Compute scaling factor:

$$\rho_k = \frac{1}{y_k^\top s_k}$$

    Update inverse Hessian approximation:

$$H_{k+1} = \left(I - \rho_k s_k y_k^\top\right) H_k \left(I - \rho_k y_k s_k^\top\right) + \rho_k s_k s_k^\top$$

    $k \leftarrow k + 1$
end while
Output: Approximate minimizer x^*.

positive definite approximation of the Hessian matrix. This property is crucial for reliably finding the minimum of a convex objective function, such as the one considered above.

2.4 Replication of the CME Methodology

The CME provides an example for estimating the SOFR term structure as of July 10, 2023. The data set consists of the settlement prices for all relevant SR1 and SR3 provided by the CME Group. In this set, the last day of the SR3 contract will be September 17, 2024, and there will be 9 FOMC meetings scheduled within this period. Using these future prices and FOMC dates, the implied SOFR rates between FOMC meeting dates are estimated by minimizing the objective function (2.3) with the BFGS optimization method. Then the forward SOFR term rates for various tenors are calculated by compounding the implied SOFR rates over the appropriate horizons by equation (2.6).

Table 2.1 presents a detailed comparison of the SOFR rates calculated by the CME and those obtained through the replication procedure (Algorithm 1 and 2) between the FOMC meeting dates. The rates are generally well aligned with some minor discrepancies, especially near the start of the sample period. For example, on the first period starting on the date of July 10, 2023, the replicated rate (5.04145%) is approximately 2 basis points lower than the CME rate (5.06%). A similar error is present for the next period, starting on July 26, 2023. The replicated rates become more aligned with the CME posted rates as they move along the sequence of FOMC dates, particularly after September 20, 2023. After this point, the absolute errors between the two rates are less than one basis point. These differences can be attributed to differences in numerical optimization. The replication procedure applies the standardized BFGS algorithm implemented in Python, which has default settings for convergence tolerance, step size, and internal numerical precision that may differ from the optimization method used by CME. Such algorithmic and computational differences can affect the estimation of the SOFR curve, particularly sensitive to the initial guess or the optimization setting. This may result in discrepancies of a few basis points in the final predicted rates.

Table 2.1: Comparison of the SOFR calculated by CME and by replication

FOMC/start Date	Rate in % (CME)	Rate in % (Replication)
2023-07-10	5.06	5.04145352
2023-07-26	5.29	5.28541776
2023-09-20	5.35	5.34927161
2023-11-01	5.41	5.41183808
2023-12-13	5.38	5.38447200
2024-01-31	5.31	5.31188167
2024-03-20	5.21	5.20856471
2024-05-01	5.05	5.05870893
2024-06-12	4.86	4.86954193
2024-07-31	4.66	4.65776014

Table 2.2 compares the term SOFR rates for various tenors, calculated by the CME and

the replication procedure. The result shows that the term rates obtained by the replication are very close to those by the CME method. The differences are less than one basis point across all maturities. These small errors arise directly from the differences in the SOFR rates in table 2.1.

Table 2.2: Comparison of term SOFR of different tenors by CME and by replication

Tenors	term SOFR in % (CME)	term SOFR n % (Replication)
1 month	5.19643143	5.18989469
3 months	5.30109311	5.29834987
6 months	5.40242570	5.40093802
12 months	5.38527007	5.38603116

Algorithm 2 CME Term SOFR Reference Rates Calculation

Input: quoted SR1 and SR3 futures prices $p^{(1)}$ and $p^{(3)}$, historical SOFR, FOMC dates.

Define: $\mathbf{R} = (R_0, R_1, \ldots, R_K)$, the SOFR during FOMC meetings.

for each observation interval **do**

    Compute VWAP, apply optimization to figure out the select prices.

end for

Calculate volume-weighted average price for each contract:

$$P_{SR} = \sum_i P_{\text{obs},i} \frac{V_{\text{obs},i}}{V_{\text{total}}}$$

Define the implied prices for SR1 and SR3 contracts, see equations (2.4), (2.5):

$$\hat{p}_m^{(1)}(\mathbf{R}), \quad \hat{p}_q^{(3)}(\mathbf{R})$$

Minimize objective function using BFGS method:

$$\min_{\mathbf{R}} \left[\sum_{m=0}^{12} w_m^{(1)} (p_m^{(1)} - \hat{p}_m^{(1)}(\mathbf{R}))^2 + \sum_{q=0}^{4} w_q^{(3)} (p_q^{(3)} - \hat{p}_q^{(3)}(\mathbf{R}))^2 \right]^{1/2} + \lambda_C \left[\sum_k (\Delta R_k)^2 \right]^{1/2}$$

Construct piecewise SOFR:

$$SOFR(t; \mathbf{R}) = R_0 + \sum_{i=1}^{k} (\Delta R_i) \cdot \mathbf{1}_{\{t > FOMC_k\}}$$

Compute term rates:

$$TR(T) = \frac{360}{T} \left[\prod_{t \in \tilde{T}(T)} \left(1 + \frac{SOFR(t; \mathbf{R}) \cdot d_t}{360} \right) - 1 \right]$$

Output: Piecewise SOFR curve, term SOFR rates.

Chapter 3

Predictive Machine Learning model with the Kalman Filter

This chapter develops a rigorous predictive modeling framework that integrates advanced machine learning techniques, specifically recurrent neural networks (RNNs) including Long Short-Term Memory (LSTM) architectures, with classical state-space estimation via the Kalman filter. The approach is based on the dynamic Nelson–Siegel (DNS) representation of the SOFR term structure, which allows for a parsimonious and expressive modeling of the yield curve's level, slope, and curvature over time. We first formalize the DNS model and outline its application to the SOFR curve data constructed following the procedures outlined in Chapter 2. Next, we construct an RNN architecture composed of LSTM layers to model the temporal evolution of the Nelson–Siegel factors. The Kalman filter, integrated as the filter layer of the network, is then employed to perform the final prediction and optimal state estimation of these latent factors (Kalman 1960). The structure and function of each layer, as well as the overall training methodology, are described in detail in the following sections.

3.1 Dynamic Nelson-Siegel Term Structure

The Nelson-Siegel (NS) model, first introduced by Nelson and Siegel (1987), has become a foundational approach for modeling the term structure of interest rates due to its parsimonious formulation and ability to flexibly fit observed yield curves with three interpretable factors: level, slope, and curvature. The NS model is widely adopted by central banks, financial institutions, and market data providers for curve fitting, risk management, and bond pricing.

To capture the dynamic evolution of the yield curve over time, Diebold and Li (2006) extended the model to a dynamic Nelson-Siegel (DNS), allowing the factors to evolve as stochastic processes, commonly modeled as auto-regressive or state-space models. The DNS model is now a standard tool for both fitting and forecasting the yield curve. Further extensions include arbitrage-free versions described by Christensen, Diebold, and Rudebusch (2011), which impose no-arbitrage constraints on the model to ensure theoretical consistency for pricing and risk management. These arbitrage-free frameworks are especially important when the Nelson-Siegel approach is applied to the valuation of interest rate derivatives or

for use in risk-sensitive applications.

The DNS model here expresses the yield $y(t, \tau)$ with maturity τ at time t as:

$$y(t, \tau) = b_\tau X_t \tag{3.1}$$

where b_τ is the loading parameter

$$b_\tau = \left(1, \frac{1 - e^{-\lambda\tau}}{\lambda\tau}, \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right), \tag{3.2}$$

and $X_t = (L_t, S_t, C_t)$, where $L(t)$ is the level factor that captures long-term interest rate trends, $S(t)$ is the slope factor that captures short-term rate movements, $C(t)$ is the curvature factor that influences medium-term dynamics, and λ is the decay parameter that determines the maturity at which the curvature term reaches its maximum. This model structure allows for a smooth and flexible fit to observed yield curves. At each time point t, the factors L_t, S_t, and C_t are estimated by fitting the model to the observed forward rates, while λ can be fixed or jointly estimated.

In this study, we apply the DNS model to the SOFR curve implied by SOFR futures contracts. To construct the term structure, we extract the implied SOFR rates from SOFR futures contracts on a daily basis. These curves are piecewise constant and aligned with the FOMC meeting dates. For each date, the Nelson-Siegel model is fitted to the resulting SOFR curve, producing a daily time series of parameters $\{L_t, S_t, C_t\}$, which captures the evolution of the SOFR term structure over time.

A key practical detail is that the SOFR is published only on U.S. business days. On weekends and holidays, the applicable rate is carried over from the most recent business day. Therefore, when fitting the curve, we assign a weight w_i to each rate f_i proportional to the number of calendar days to which the rate applies. This weighting scheme ensures that multi-day rates (e.g., weekends and holidays) are appropriately accounted for in the fitting process.

To capture the time evolution of the Nelson-Siegel factors in a flexible and tractable manner, we model $X_t = (L_t, S_t, C_t)$ using the extended Vašíček process. The classic Vašíček model described by Vašíček (1977) is a widely used mean-reverting Gaussian process in fixed income modeling, notable for its analytical tractability and closed-form solutions for bond pricing. The extended Vašíček model, as formalized by Hull and White (1990), allows the mean level, mean reversion speed, and volatility to vary over time or depend on the current state. This flexibility makes the model suitable for the SOFR market, where regime changes and policy interventions are frequent.

The conditional expected value and variance are obtained by proposition 3.1.1.

Proposition 3.1.1. *Suppose X_t follows the extended Vašíček process*

$$dX_t = \kappa_t \left(\theta_t - X_t\right) dt + \sigma_t dW_t, \tag{3.3}$$

where κ_t, θ_t, and σ_t are functions that depend on t. The expected value and the variance are

$$\mathbb{E}[X_T \mid \mathcal{F}_t] = e^{-\int_t^T \kappa_u du} X_t + \int_t^T e^{-\int_u^T \kappa_v dv} \kappa_u \theta_u du, \tag{3.4}$$

$$\mathrm{Var}[X_T \mid \mathcal{F}_t] = \int_t^T e^{-\int_u^T \kappa_v dv} \Sigma_u e^{-\int_u^T \kappa_v^\top dv} du. \tag{3.5}$$

where $\mathcal{F}_t$ denotes the filtration representing all information available up to time t. All random variables and processes (e.g., X_t, W_t) are adapted to $\{\mathcal{F}_t\}_{t\geq 0}$, and $\Sigma_t = \sigma_t \sigma_t^\top$ is the covariance matrix of the process X_t.

The explicit formulas for the conditional mean and variance of the extended Vašíček process are fundamental for the estimation of the dynamic Nelson-Siegel factors. In particular, they form the basis for the prediction and update steps in the Kalman filter, which we employ in the empirical analysis to extract the latent factors from observed SOFR futures prices. Overall, this section develops a comprehensive modeling and stochastic framework, combining the dynamic Nelson-Siegel (DNS) model for the SOFR curve with the extended Vašíček process for the evolution of its latent factors. This foundation supports the term structure estimation and forecasting methods implemented in the following chapters.

Filipović (1999) considers the consistency of the Nelson Siegel family and HJM models in context of arbitrage-free modeling of forward rates and bonds. The consistency of the Nelson Siegel family with respect to the SOFR term-structure, especially considering the strong evidence of jumps at scheduled FOMC meeting dates, suggests modifications to our framework following Fontana, Grbac, and Schmidt (2024) and Skov and Skovmand (2021).

3.2 Recurrent Neural Network Modeling

In this section, we will describe the application of Recurrent Neural Networks (RNNs) to predict the time evolution of the SOFR curve factors, following the approach and the notation introduced by Gao (2021). In particular, we focus on modeling the dynamic Nelson-Siegel (DNS) factors (L_t, S_t, C_t) associated with the SOFR term structure. We will show how RNNs can be trained to learn the way these factors change over time.

A fundamental component of neural networks is the dense layer, which consists of a linear transformation followed by an activation function. Throughout this section, we denote a dense layer by D, with the following formulation:

$$y = D(x) := a(\mathfrak{W}x + \mathfrak{b}), \tag{3.6}$$

where x denotes the input, $\mathfrak{W}$ represents the weights, $\mathfrak{b}$ is the bias, and the operation between $\mathfrak{W}$ and x is typically a matrix product or a tensor product. The function $a(\cdot)$ stands for the activation function, which determines the activation state of each neuron. Activation functions play an important role in neural networks since they enable non-linear modeling capabilities by allowing certain neurons to be activated or inactivated.

In our modeling framework, data are processed through a sequence of modules that connect repeatedly as an RNN. Each module consists of four main layers: the input layer, the state layer, the residual layer, and the filter layer.

3.2.1 Long Short-Term Memory (LSTM)

Traditional recurrent neural networks are the natural modeling choice for time series data because they can encode information from previous time steps into a hidden state that is recursively updated. However, conventional RNNs often suffer from the problem of vanishing or exploding gradients, making it difficult for them to learn long-term dependencies in sequential data (Hochreiter 1991). This limitation can be especially problematic for financial time series, where persistent economic or policy regimes may cause latent factors (such as Nelson-Siegel parameters) to evolve over extended periods. To overcome these challenges, Hochreiter and Schmidhuber (1997) introduced the Long Short-Term Memory (LSTM) network. The LSTM augments the standard RNN with a memory cell and a system of gating mechanisms, specifically the forget, input, and output gates. These gates enable the LSTM to learn when to retain or discard information, allowing the network to effectively capture both short-range and long-range dependencies in the data.

3.2.2 Input layer

The input layer captures the time dynamics of the Nelson-Siegel factors using an LSTM cell. The linear model will be trained with data in a 3-dimensional tensor of size $S' \times L' \times F'$ and at each time step, the input data is a $1 \times F'$ vector. S' is the number of samples, L' is the sequential length, and F' is the size of the feature.

At each time t, the LSTM cell receives the current input x and the previous states (c, h), where c is the output and h is the hidden state. Then it will output the updated cell state c_t and the hidden state h_t. We denote the LSTM cell by $L(\cdot, \cdot)$:

$$(c_t, h_t) = L\big(x, (c_{t-1}, h_{t-1})\big).$$

The calculation within the LSTM cell at each t follows:

$$f_t = a_g\left(\mathcal{W}_f x + \mathcal{W}'_f h_{t-1} + b_f\right), \tag{3.7}$$
$$i_t = a_g\left(\mathcal{W}_i x + \mathcal{W}'_i h_{t-1} + b_i\right), \tag{3.8}$$
$$o_t = a_g\left(\mathcal{W}_o x + \mathcal{W}'_o h_{t-1} + b_o\right), \tag{3.9}$$
$$\tilde{c}_t = a_c\left(\mathcal{W}_c x + \mathcal{W}'_c h_{t-1} + b_c\right), \tag{3.10}$$
$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t, \tag{3.11}$$
$$h_t = o_t \circ a_h(c_t), \tag{3.12}$$

where the operator $\circ$ is the Hadamard product (element-wise product).

There are three gates in the LSTM structure: the forget gate (f_t), the input gate (i_t). The forget gate f_t determines which information from the previous cell state c_{t-1} should be retained or discarded, allowing the network to selectively "forget" irrelevant historical data. The input gate i_t controls how much of the new cell state $\tilde{c}_t$, which encodes the current input and recent hidden information, is written in the cell state. The output gate o_t determines which parts of the updated memory c_t are passed on to the next layer as the hidden state h_t. The initial values are $c_0 = \bar{0}$ and $h_0 = \bar{0}$. For an input $X_t \in \mathbb{R}^N$, N is the number of input

features and H is the number of hidden units in the model, the model's weight matrices and bias vectors are predefined as:

$$\mathcal{W}, \mathcal{W}' \in \mathbb{R}^{H \times N}, \quad \text{and} \quad b \in \mathbb{R}^{H}.$$

The activation functions used in the LSTM are given by:

$$a_g(x) = \frac{1}{1 + e^{-x}}, \quad a_c(x) = \tanh(x), \quad a_h(x) = \tanh(x).$$

The function a_g returns a value between 0 and 1, which is often used between layers in the neuron or to predict probabilities. The hyperbolic tangent function tanh is similar to the sigmoid but produces outputs from -1 to 1, and is commonly used in output layers or binary classification.

In the linear model, the input layer consists of two connected LSTM cells, L_1 and L_2. Given the input at the time step t, denoted Y_t, the LSTM layer processes the data as follows:

$$\begin{aligned}
\left(c_t^{I_1},\ h_t^{I_1}\right) &= L_1\left(Y_t,\ \left(c_{t-1}^{I_1},\ h_{t-1}^{I_1}\right)\right), \\
\left(c_t^{I_2},\ h_t^{I_2}\right) &= L_2\left(c_t^{I_1},\ \left(c_{t-1}^{I_2},\ h_{t-1}^{I_2}\right)\right), \\
c_t^I &= c_t^{I_2},
\end{aligned}$$

where c_t^I is the output of the input layer.

3.2.3 State layer

Given the output $c_t^I \in \mathbb{R}^{1 \times H}$ from the input layer, it will be the input to the state layer, which generates three dense layers: κ, θ, and σ

$$\begin{aligned}
\kappa(c_t^I) &: [0, T] \times \mathbb{R}^{1 \times H} \to \mathbb{R}^{d \times d}, \\
\theta(c_t^I) &: [0, T] \times \mathbb{R}^{1 \times H} \to \mathbb{R}^{d}, \\
\sigma(c_t^I) &: [0, T] \times \mathbb{R}^{1 \times H} \to \mathbb{R}^{d \times d}.
\end{aligned}$$

The computations in the state layer are expressed as

$$\begin{aligned}
\kappa &= a_\kappa\left(\mathcal{W}_\kappa \cdot c_t^I + b_\kappa\right), \\
\theta &= a_\theta\left(\mathcal{W}_\theta \cdot c_t^I + b_\theta\right), \\
\sigma &= a_\sigma\left(\mathcal{W}_\sigma \cdot c_t^I + b_\sigma\right),
\end{aligned}$$

where $\cdot$ represents the tensor product, and the parameters $\mathcal{W}_\kappa$, $\mathcal{W}_\theta$, $\mathcal{W}_\sigma$ are the layer weights, with b_κ, b_θ, b_σ as the corresponding bias terms. The dimensions of these parameters are as follows:

$$\begin{aligned}
\mathcal{W}_\kappa &\in \mathbb{R}^{H \times d \times d}, \quad b_\kappa \in \mathbb{R}^{d \times d}, \\
\mathcal{W}_\theta &\in \mathbb{R}^{H \times d}, \quad b_\theta \in \mathbb{R}^{d}, \\
\mathcal{W}_\sigma &\in \mathbb{R}^{H \times d \times d}, \quad b_\sigma \in \mathbb{R}^{d \times d}.
\end{aligned}$$

The activation functions in the state layer are chosen as

$$\begin{aligned} a_\kappa(x) &= x, \\ a_\theta(x) &= \tanh(x), \\ a_\sigma(x) &= \tanh(x), \end{aligned}$$

3.2.4 Residual layer

After obtaining the predictions $\hat{Y}_t$ and the observed values Y_t, the residual layer processes the modeling error and estimates the residual uncertainty. We first calculate the residual e_t, and the batch normalization $\bar{e}_t$ (see Ioffe and Szegedy (2015)) of the residual. Then we feed the $\bar{e}_t$ into an LSTM L_R to generate a noise matrix u_t by applying a dense layer D_R. The equations are given by

$$\begin{aligned} e_t &= |Y_t - \hat{Y}_t|, \\ \bar{e}_t &= \mathrm{BN}(e_t), \\ (c_t^R, h_t^R) &= L_R\left(\bar{e}_t, (c_{t-1}^R, h_{t-1}^R)\right), \\ u_t &= D_R(c_t^R), \end{aligned}$$

After we have the model parameters $(\kappa_t, \theta_t, \sigma_t, u_t)$, the state variables $\hat{X}_t$, the predicted values $\hat{Y}_t$, and the actual observed values Y_t, we apply a filter to predict the next state $\hat{X}_{t+1}$ and the next prediction $\hat{Y}_{t+1}$.

3.2.5 Kalman filter layer

Consider the SOFR rates $y_t = (y_1, ..., y_m)$ at time t with fixed tenors $\tau_1, ..., \tau_m$. We assume that the noise between the observation and the model (3.1) is Gaussian with a mean of zero and variance U_t,

$$y(t, \tau) = b_\tau X_t + \varepsilon_t, \tag{3.13}$$

where $\mathbb{E}[\varepsilon_t] = 0$, $\mathrm{Var}[\varepsilon_t] = \mathbb{E}[\varepsilon_t \varepsilon_t^\top] = U_t$, and the matrix U_t is diagonal, whose diagonal entries are the elements of the vector u_t obtained from the residual layer.

It is hard to compute the expectation and variance of X_t using the proposition when the state variable is non-scalar and the parameters κ_t, θ_t, and σ_t are matrices. Hence, we use a discretization grid $[t_0, t_1, \ldots, t_k, t_{k+1}, \ldots]$ with a constant time increment $\Delta t_k = t_{k+1} - t_k$ and the matrix κ_t, θ_t, and σ_t as constant within each interval $t \in [t_k, t_{k+1})$. We write X_k for X_{t_k} and $\mathcal{F}_k$ for $\mathcal{F}_{t_k}$. Hence, we have the estimations by Proposition 3.1.1,

$$\mathbb{E}[X_{k+1} \mid \mathcal{F}_k] = e^{-\kappa_k \Delta t} X_k + \left(I - e^{-\kappa_k \Delta t}\right) \theta_k, \tag{3.14}$$

$$\mathrm{Var}[X_{k+1} \mid \mathcal{F}_k] = \int_{t_k}^{t_k + \Delta t} e^{-\kappa_k t} \Sigma_k e^{-\kappa_k^\top t}\, dt. \tag{3.15}$$

We denote $Q_k = \mathrm{Var}[X_{k+1} \mid \mathcal{F}_k]$, and the computation of Q_k can be simplified using the spectral decomposition of the matrix κ_k

$$\kappa_k = E_k V_k E_k^{-1},$$

where E_k is the eigenvector matrix of κ_k, and V_k is a diagonal matrix with the eigenvalues ζ of κ on the diagonal. Then the conditional variance can be written as

$$Q_k = E\left(\int_{t_k}^{t_k+\Delta t} e^{-Vt}\Omega_k e^{-V^\top t}\, dt\right) E^\top,$$

where $\Omega_k = E_k^{-1}\Sigma_k E_k^{-\top} = (\omega_{i,j})_{i,j}$ and $\Sigma_k = \sigma_k \sigma_k^\top$. The integral inside Q_k can be computed element-wise as

$$\begin{aligned}\int_{t_k}^{t_k+\Delta t} \left(e^{-Vt}\Omega_k e^{-V^\top t}\right)_{i,j} dt &= \int_{t_k}^{t_k+\Delta t} e^{-\zeta_i t}\omega_{i,j} e^{-\zeta_j t}\, dt \\ &= \omega_{i,j}\int_{t_k}^{t_k+\Delta t} e^{-(\zeta_i+\zeta_j)t} dt \\ &= \frac{\omega_{i,j}}{\zeta_i+\zeta_j}\left(1 - e^{-(\zeta_i+\zeta_j)\Delta t}\right).\end{aligned}$$

Hence, we can obtain the approximation of Q_k

$$Q_k = E_k\left(\frac{\omega_{i,j}}{\zeta_i+\zeta_j}\left(1 - e^{-(\zeta_i+\zeta_j)\Delta t}\right)\right)_{i,j} E_k^\top.$$

The state and the observation equations are

$$\begin{aligned}X_{k+1} &= e^{-\kappa_{t_k}\Delta t} X_k + \left(I - e^{-\kappa_{t_k}\Delta t}\right)\theta_{t_k} + w_k, \\ y_{k+1} &= M X_{k+1} + \varepsilon_k,\end{aligned}$$

and M is defined as

$$M = \begin{bmatrix} b_1(\tau_1) & b_2(\tau_1) & b_3(\tau_1) \\ \vdots & \vdots & \vdots \\ b_1(\tau_m) & b_2(\tau_m) & b_3(\tau_m)\end{bmatrix},$$

where τ_m is the maximum tenor among all observations. The noise terms w_k and ε_k are assumed to be Gaussian with a zero mean and covariance Q_k and U_k

$$\begin{pmatrix} w_k \\ \varepsilon_k\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} Q_k & 0 \\ 0 & U_k\end{pmatrix}\right). \tag{3.16}$$

Following Date and Ponomareva (2011), the prediction step of the Kalman filter is given by

$$\hat{X}_{k|k-1} = e^{-\kappa_{t_{k-1}}\Delta t}\hat{X}_{k-1|k-1} + \left(I - e^{-\kappa_{t_{k-1}}\Delta t}\right)\theta_{t_{k-1}}, \tag{3.17}$$

$$\hat{P}_{k|k-1} = e^{-\kappa_{t_{k-1}}\Delta t}\hat{P}_{k-1|k-1}\left[e^{-\kappa_{t_{k-1}}\Delta t}\right]^\top + Q_{k-1}, \tag{3.18}$$

$$\hat{y}_k = M\hat{X}_{k|k-1}. \tag{3.19}$$

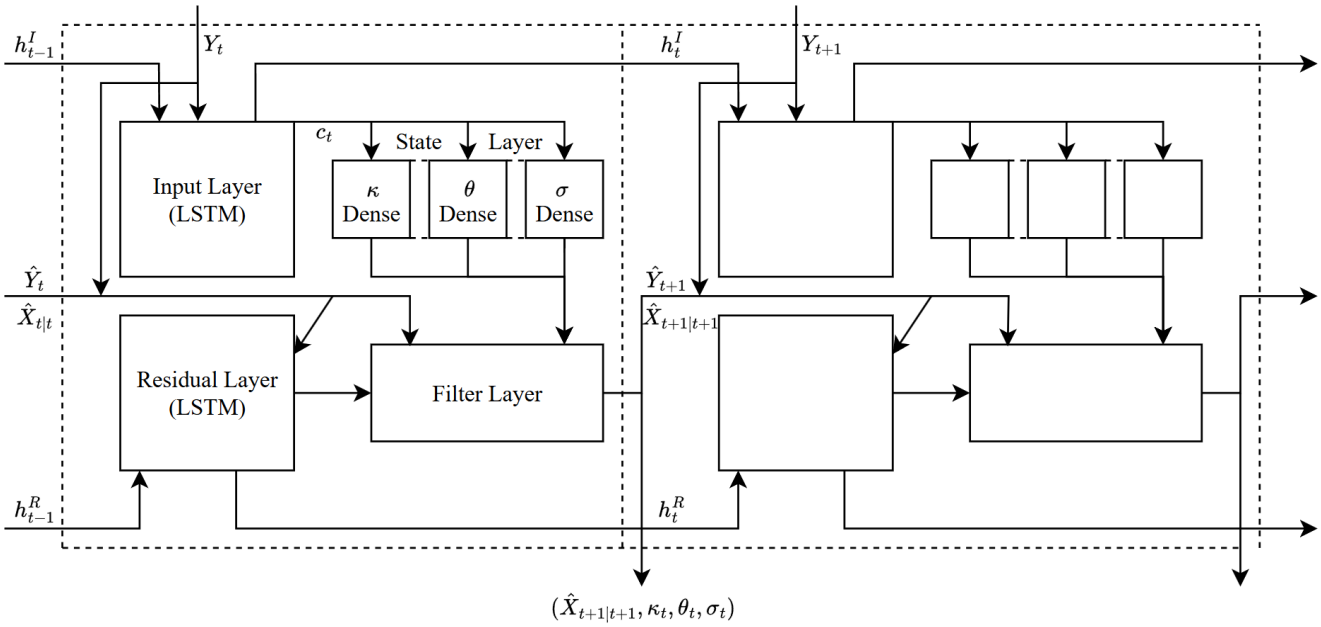


Figure 3.1: Recurrent Neural Networks

The measurement (update) step of the Kalman filter is given by

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k \nu_k, \tag{3.20}$$

$$\hat{P}_{k|k} = \hat{P}_{k|k-1} - K_k M \hat{P}_{k|k-1}, \tag{3.21}$$

$$\nu_k = y_k - \hat{y}_k, \tag{3.22}$$

$$F_k = M \hat{P}_{k|k-1} M^\top + U_k, \tag{3.23}$$

$$K_k = \hat{P}_{k|k-1} M^\top F_k^{-1}. \tag{3.24}$$

The filter layer serves as the output cell, applying filtering operations, updating the SOFR rates, and generating the final predicted values. In particular, the filter layer uses state variables (κ, θ, σ), which are obtained in the state layer using weights $\mathfrak{W}$, bias terms b, and outputs from the input layer. All model parameters, including weights $\mathfrak{W}$ and biases $\mathfrak{b}$ in each layer, are trained to minimize the prediction loss function.

The basic module of the RNN consists of four fully connected layers, and the overall RNN architecture is constructed by stacking these fundamental modules. Figure 3.1 illustrates the structure of both the basic module and the complete RNN.

3.2.6 Loss function

The choice of loss function is critical for training neural networks on interest rate data. In this study, we consider both the mean squared error (MSE), also known as the L_2 norm, and

mean absolute error (MAE), also called the L_1 norm losses:

$$L_2(\vartheta) = \frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - \hat{Y}_i\right\|_2^2,$$

$$L_1(\vartheta) = \frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - \hat{Y}_i\right\|,$$

where ϑ denotes the model weights, Y_i is the observed value, and $\hat{Y}_i$ is the predicted value. The MSE is sensitive to larger errors and thus emphasizes outliers, while the MAE is more robust and particularly suitable for modeling small-magnitude errors, a common feature of interest rate series. By comparing the performance under both loss functions, we assess not only predictive accuracy but also robustness to outliers and suitability for the characteristics of interest rate data.

The learning objective is to find the optimal parameter vector $\hat{\vartheta}$ that minimizes the loss function:

$$\hat{\vartheta} = \arg\min_{\vartheta} L(\vartheta), \tag{3.25}$$

where L can denote either L_2 or L_1. We apply the following gradient descent update step with a learning rate $\alpha_k \in (0, 1)$:

$$\vartheta_k = \vartheta_{k-1} - \alpha_k \nabla_{\vartheta} L(\vartheta_{k-1}), \tag{3.26}$$

where the learning rate α_k is a function of the loss value in each iteration.

The process of training a neural network is based on an optimization procedure. In the process, an optimizer updates the model's weights by computing gradients through back-propagation and applying gradient descent. This adjusts the model weights in proportion to the derivative of the error with respect to the corresponding weights. The optimizer determines the specific strategy for updating the weights, thus playing a critical role in both the speed and effectiveness of model training.

The ADAM algorithm, short for Adaptive Moment Estimation, introduced by Kingma and Ba (2015), is a widely used optimization method in machine learning and deep learning tasks. The parameters β_1, β_2 are exponential decay rates for the moment estimates, which are hyperparameters typically set to 0.9 and 0.999, respectively. ADAM maintains exponentially decaying averages of biased first moment estimate m_t and biased second raw moment estimate v_t, which are updated as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \tag{3.27}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \tag{3.28}$$

where $g_t = \nabla_{\vartheta} L(\vartheta)$ is the gradient of the loss function and g_t^2 indicates the elementwise square $g_t \odot g_t$. Here, $g_t \odot g_t$ denotes the elementwise (Hadamard) product of g_t, that is, $(g_t \odot g_t)_i = (g_{t,i})^2$. To correct for initialization bias, bias-corrected moment estimates are computed:

$$\hat{m}_t = \frac{m_t}{1 - (\beta_1)^t}, \tag{3.29}$$

$$\hat{v}_t = \frac{v_t}{1 - (\beta_2)^t}. \tag{3.30}$$

The parameter update rule for Adam is given by:

$$\vartheta_t = \vartheta_{t-1} - \alpha_t \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \tag{3.31}$$

where ϵ is a small constant, fixed as 10^{-8}, to prevent division by zero. The learning rate α_t is dynamically adjusted based on the progress in minimizing the loss function.

$$\alpha_t = \left(1 - 0.9 \times \mathbf{1}_{\left\{\left|\frac{L(\vartheta_t) - L(\vartheta_{t-1})}{L(\vartheta_{t-1})}\right| < 0.001\right\}}\right) \cdot \alpha_{t-1}, \tag{3.32}$$

where

$$\mathbf{1}_{\left\{\left|\frac{L(\vartheta_t) - L(\vartheta_{t-1})}{L(\vartheta_{t-1})}\right| < 0.001\right\}} = \begin{cases} 1, & \text{if } \left|\frac{L(\vartheta_t) - L(\vartheta_{t-1})}{L(\vartheta_{t-1})}\right| < 0.001, \\ 0, & \text{otherwise.} \end{cases} \tag{3.33}$$

This adaptive strategy ensures that the learning rate decreases automatically when the improvement of the loss function stagnates. This will reduce the risk of overfitting, speed up convergence to the optimal solution, and minimize the need for manual learning rate adjustment.

To summarize, in this Chapter, we apply the CME methodology to 1-month and 3-month SOFR futures data, collected from the CME Group to construct a daily piecewise-constant SOFR curve. Each day, we fit this curve to the Nelson-Siegel model, obtaining dynamic level, slope, and curvature parameters. Assuming these parameters follow an extended Vašíček process, we model their evolution and use recurrent neural networks (RNNs) for multi-step prediction of the term structure (see Algorithm 3). The next Chapter provides a numerical implementation using actual market data.

Algorithm 3 Full SOFR Term Structure Forecasting Algorithm

Input: CME SR1 and SR3 futures prices $\{p_m^{\mathrm{SR1}}, p_q^{\mathrm{SR3}}\}$, historical SOFR rates $\{r'_t\}$, FOMC dates $\{FOMC_k\}$.

CME Methodology: Estimate SOFR curve using the CME optimization (see Algorithm 2)

Dynamic Nelson–Siegel:

$$y(t,\tau) = L_t + S_t \frac{1-e^{-\lambda\tau}}{\lambda\tau} + C_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$$

For each t, fit Nelson–Siegel factors (L_t, S_t, C_t) to the SOFR curve.

RNN-Kalman Forecasting Layers:

Input (LSTM) Layer:

$$(c_t, h_t) = L\big(X_t, (c_{t-1}, h_{t-1})\big)$$

State Layer:

$$\kappa_t = a_\kappa\left(\mathcal{W}_\kappa \cdot c_t + b_\kappa\right), \quad \theta_t = a_\theta\left(\mathcal{W}_\theta \cdot c_t + b_\theta\right), \quad \sigma_t = a_\sigma\left(\mathcal{W}_\sigma \cdot c_t + b_\sigma\right)$$

Residual Layer:

$$e_t = |Y_t - \hat{Y}_t|, \quad \bar{e}_t = \mathrm{BN}(e_t), \quad (c_t^R, h_t^R) = L_R\left(\bar{e}_t, (c_{t-1}^R, h_{t-1}^R)\right), \quad u_t = D_R(c_t^R)$$

Kalman Filter Layer: Apply the Kalman filter prediction and measurement steps (see Eqs. (3.17)–(3.24))

RNN-Kalman Forecasting process:
Initialize: $\mathbf{H}_t = \{c_t^I, h_t^I, c_t^R, h_t^R\}$, set $(\hat{X}_{t_0|t_0}, P_{t_0|t_0}, \mathbf{H}_{t_0}, \mathbf{T})$.
for $t \in \mathbf{T}$ **do**
    $\hat{X}_{t+1|t+1}, \hat{Y}_{t+1}, \mathbf{H}_t = \mathrm{RNN}(\hat{X}_{t|t}, \hat{Y}_t, Y_t, \mathbf{H}_{t-1})$
end for

Forecasting and Evaluation:
Generate 1-day and 5-day ahead forecasts $\{\hat{Y}(t+1,\tau)\}$ and $\{\hat{Y}(t+5,\tau)\}$
Optimize parameters ϑ (using ADAM method) by minimizing

$$L(\vartheta) = \frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - \hat{Y}_i\right\|^2 \quad \text{(MSE)}, \qquad L(\vartheta) = \frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - \hat{Y}_i\right\| \quad \text{(MAE)}$$

Output: Forecasted SOFR curves $\{\hat{Y}(t+1,\tau)\}$ and $\{\hat{Y}(t+5,\tau)\}$

Chapter 4

Numerical Implementation

The data set consists of daily observations of the 1-month and 3-month SOFR futures prices, obtained from the CME Term SOFR Constituent File provided by CME Group, covering the period from November 20, 2023, to June 17, 2025. This data set is selected because the CME Group is the leading and most liquid exchange for SOFR futures, and its published prices are widely regarded as the market standard. Comparable data from other sources is either unavailable or does not provide the same level of reliability and completeness. We first apply the CME methodology to construct a daily piecewise-constant SOFR forward rate curve. This curve is then fitted to the Nelson–Siegel model, yielding a time series of the parameters: level (L_t), slope (S_t), and curvature (C_t) that characterize the shape of the SOFR forward curve over time. We assume that the dynamics of the Nelson–Siegel factors follow an extended Vašíček process, and we apply the recurrent neural network (RNN) on the model. The RNN is trained to capture the temporal patterns in the latent factors and forecast their future values. We integrate a Kalman filtering layer that aligns the predicted latent states with the realized market data.

To evaluate forecasting performance, we partition the dataset based on the number of CME business days into training and test sets using an 80/20 chronological split. Specifically, the first 80% of the data (from November 17, 2023 to February 24, 2025) is used for training, and the remaining 20% (from February 25, 2025 to June 17, 2025) is used for testing. This time-based split ensures a realistic forecasting framework in which the model is trained strictly on historical data and evaluated on future, unseen observations. We perform both one-day ahead and five-day ahead forecasting. For the 5-day forecasting horizon, we adopt the batching strategy proposed by Gao (2021), in which the data are divided into 5 non-overlapping sequences, and the sequences are assumed to represent the observations on 5 days in a week. Instead of training them separately, we mix the 5 distinct sequences as one training set.

The decay parameter λ in the Nelson-Siegel model governs the speed at which the influence of the curvature factor diminishes with maturity and thereby determines the location of the "hump" in the yield curve. As noted by Diebold and Li (2006), the curvature loading reaches its maximum. In this study, as the focus is on modeling short-term interest rates using SOFR futures with maturities up to 15 months, we set $\lambda = 1$, which positions the maximum curvature loading at $\tau^* = \ln 2/1 \approx 0.693$ years (approximately 8.3 months). This ensures that the model is most flexible in the region where the majority of the data lies. Fig-

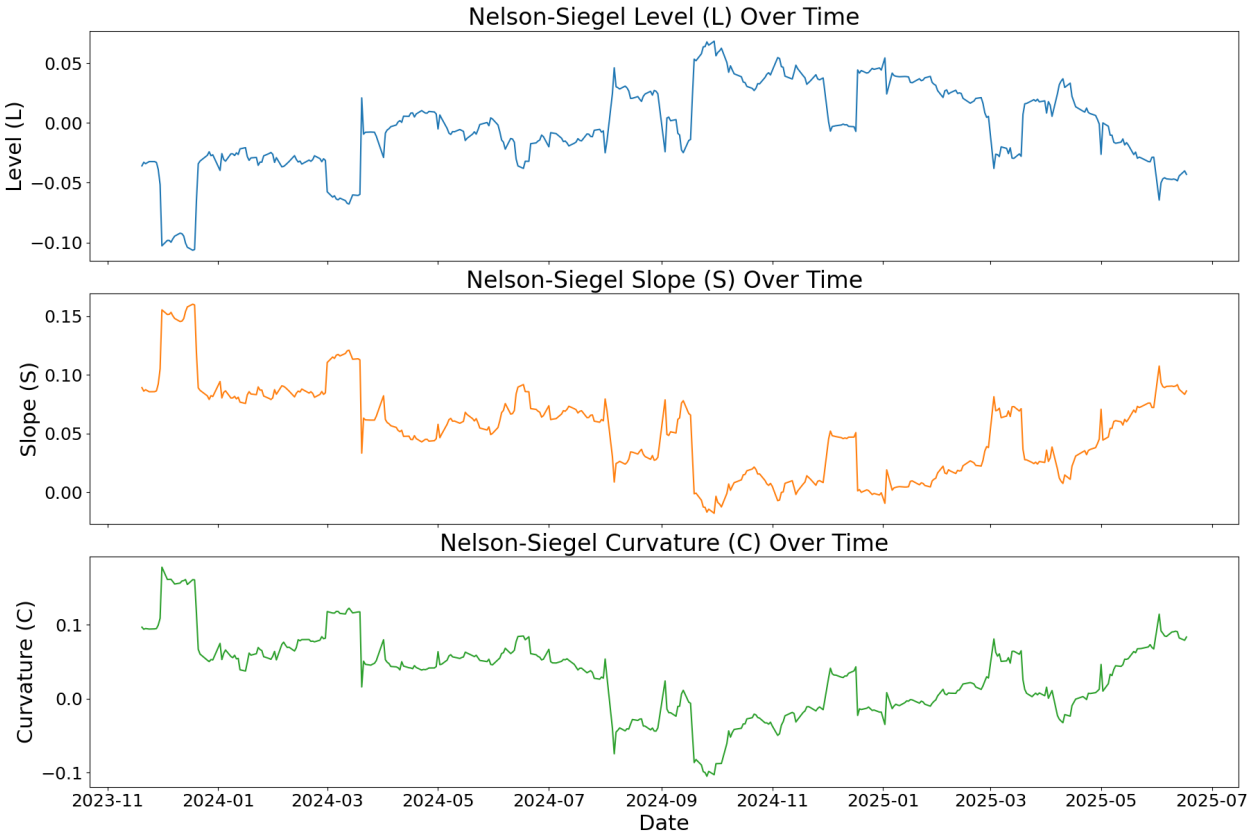


Figure 4.1: Estimated Nelson-Siegel parameters with fixed $\lambda = 1$

ure 4.1 illustrates the dynamic evolution of the Nelson-Siegel factors (L_t, S_t, C_t) over time as $\lambda = 1$. And Figure 4.2 shows the curve of Nelson-Siegel along with the time during the entire data set.

Each training sample consists of a sequence of $S' = 50$ days, where the feature dimension is $F' = 13$, which is the number of tenors: 1 week, 1 month, 2 months, ..., 12 months. The model uses this input to predict the term structure for the next 1-day or 5-day horizon. The input dimension is $d = 3$, corresponding to the three DNS factors. The main LSTM input layer employs $H = 64$ hidden units, a size chosen in line with standard practices for deep learning on financial time series. Batch normalization is applied to the residual errors before feeding them to the residual LSTM, which uses a hidden size of 16. All weights and biases in the LSTM and subsequent linear layers are initialized using PyTorch defaults (PyTorch 2024). Specifically, the weights and biases within the LSTM cells are initialized uniformly in $\left[-\sqrt{\frac{1}{H}}\sqrt{\frac{1}{H}}\right]$, while those in the linear state layers are initialized uniformly in $\left[-\sqrt{\frac{1}{F'}}, \sqrt{\frac{1}{F'}}\right]$, where F' is the input feature dimension for the layer.

Due to the amount of available data, systematic hyperparameter tuning on a test set was not undertaken. The hyperparameters used in the model include a learning rate of 1×10^{-3}, a batch size of 64, and 50 epochs of training. The Adam optimizer is used for all parameters, with a default $(\beta_1, \beta_2) = (0.9, 0.999)$, and the learning rate decay is triggered when the validation loss plateaus.

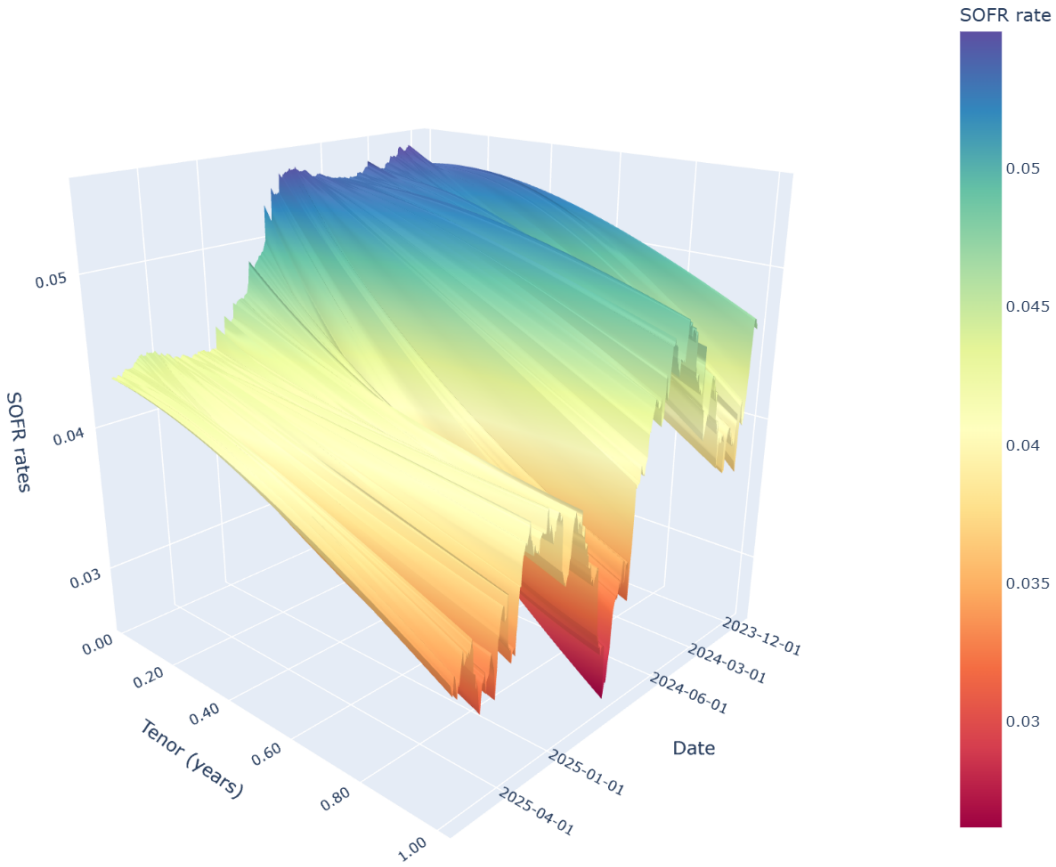


Figure 4.2: Dynamic evolution of the SOFR term structure surface over time.

We run our models for 50 epochs and plot the pattern of the loss function during training to demonstrate the sufficiency of the training process. The dataset is partitioned into training and test set, as introduced above, and we compare the forecast performance by reporting the loss function for the four model variants: 1-day-ahead and 5-day-ahead forecasts, each under both L_2 and L_1 loss functions. In all 4 configurations, after 30 epochs, the values of the loss function are stable, which shows that the number of epochs is sufficient (see Figure 4.3). Importantly, the training and test loss curves remain closely aligned and ultimately converge to similar levels across all settings. This parallel behavior between training and test loss suggests that the models do not exhibit signs of overfitting.

Figure 4.4 illustrates the root mean squared error (RMSE) for the four variants. Each plot shows the daily RMSE in basis points (bps), which enables a direct comparison of the predictive accuracy and stability across different modeling choices. For 1-day-ahead forecasts, under both L_2 and L_1 loss, the predictive errors are modest, with most RMSE values ranging between 3 and 10 bps throughout the test period. There are two visible spikes in early March and early April. These results indicate that the model can reliably capture the daily dynamics of the SOFR term structure for short-horizon forecasts. The 5-day-ahead forecasting results display higher RMSE values: the typical level is around 8 to 14 bps, and the worst-case error in early April can exceed 35 bps. The L_2 and L_1 loss functions produce similar trends, with L_2 tending to attenuate extremes and yield a slightly lower median error. Overall, the model provides a reasonable approximation of the SOFR curve in both 1-day and 5-day predictions, with performance naturally degrading as the horizon lengthens.

Figure 4.5 presents the daily RMSE for 5-day-ahead SOFR term structure forecasts over the test period, now enhanced with vertical dashed lines that indicate the timing of major policy events. These event markers allow for a clearer visual alignment between the spikes in RMSE and the occurrence of two FOMC meetings (March 19 and May 7, 2025, marked in red) as well as the administration's tariff announcement (April 2, 2025, marked in blue). Each of these events corresponds to a pronounced local maximum in the RMSE series, highlighting the model's difficulty in forecasting abrupt, policy-driven shifts in the SOFR curve. This pattern highlights the challenge for data-driven models in anticipating the impact of monetary policy and unexpected political events on the short-rate term structure.

Figure 4.6 and Figure 4.7 provide visual comparisons of the observed and predicted SOFR term structures for two days from the test set, under different forecast horizons and loss functions. Figure 4.6a shows a typical "good fit" day for the 1-day-ahead forecast using L_1 (2025-03-28, RMSE = 2.19 bps), where the predicted curve closely tracks the observed SOFR rates across all tenors. Figure 4.6b displays a typical "median fit" case for the 1-day-ahead forecast under L_1 (2025-03-20, RMSE = 6.87 bps), where the model still captures the overall shape of the observed term structure, with a mild level bias at the short-mid tenors. Figure 4.7 shows analogous results for the 5-day-ahead forecast using the L_2 loss. Figure 4.7a presents a "good fit" example (2025-04-11, RMSE = 4.17 bps), with the predicted curve closely matching the observed values and only a small short-end offset. Figure 4.7b illustrates a "median fit" case (2025-06-03, RMSE = 10.49 bps), where the main discrepancy is curvature: the forecast shows an inverted shape relative to the observed curve.

Figure 4.8 illustrates the time-varying dynamics of the state factor parameters. The estimated state variable parameters, κ, θ, σ remain stable and well-behaved throughout the

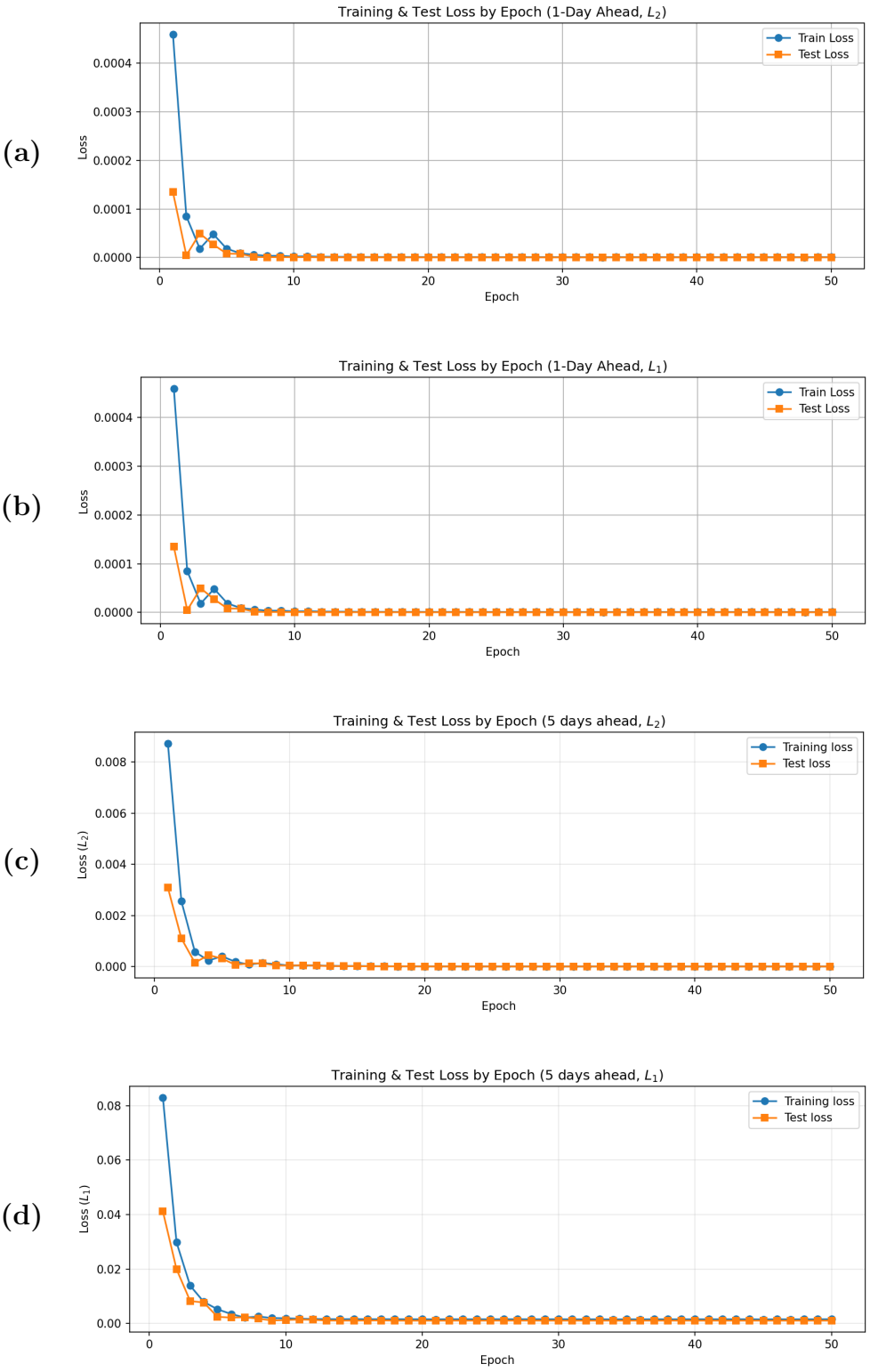


Figure 4.3: Training and test loss over 40 epochs under all forecasting settings: (a) L_2 loss, 1-day-ahead; (b) L_1 loss, 1-day-ahead; (c) L_2 loss, 5-day-ahead; (d) L_1 loss, 5-day-ahead.

Figure 4.4: Test set RMSE by day for all forecasting settings: (a) L_2 loss, 1-day-ahead; (b) L_1 loss, 1-day-ahead; (c) L_2 loss, 5-day-ahead; (d) L_1 loss, 5-day-ahead.

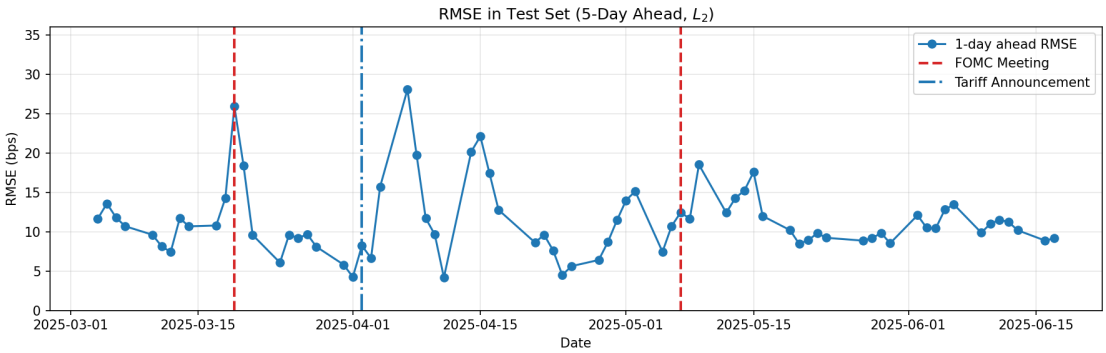


Figure 4.5: Daily RMSE for 5-day-ahead SOFR term structure forecasts in the test set (L_2 loss). Vertical dashed lines indicate the time of key policy events: two FOMC meetings (red) and the tariff announcement (blue).

sample period, indicating the effectiveness of the Kalman filtering approach in our framework. In contrast, Gao (2021) reports that before arbitrage regularization, the estimated state variable parameters by the Kalman filter exhibit significant fluctuations and distinct regime shifts, with abrupt changes in parameter values. After applying arbitrage regularization, their trajectories become much smoother and more stable. The stability observed in our results suggests that explicit arbitrage regularization may not be necessary for our modeling approach.

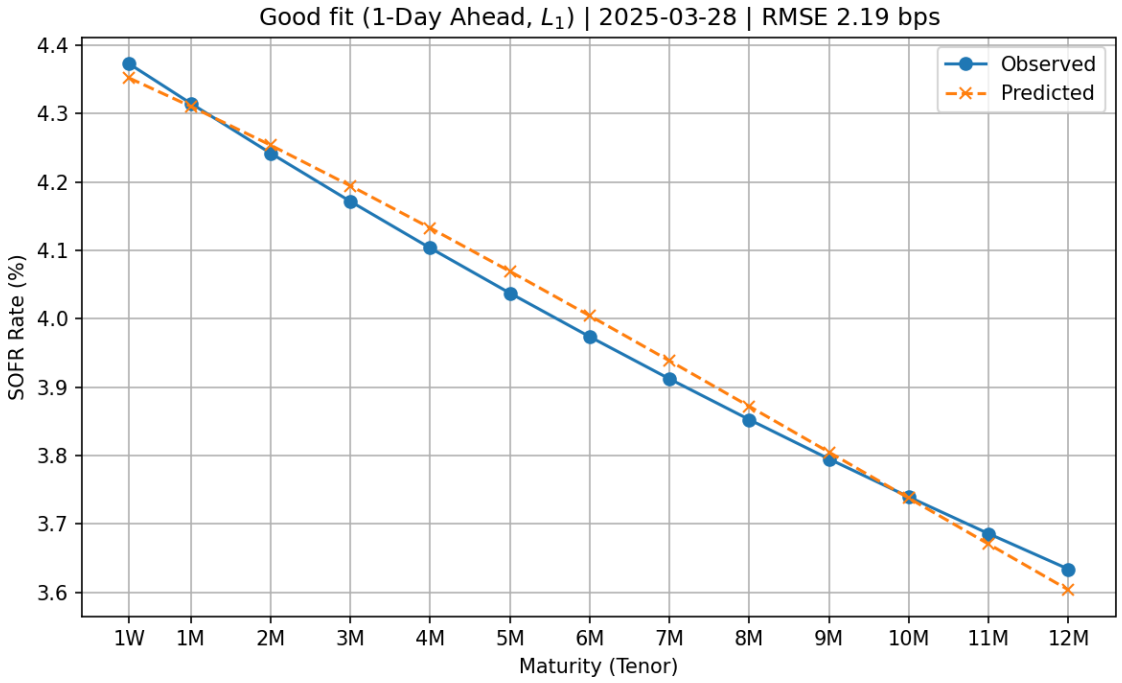
A persistence forecast is a naive benchmark that assumes the future curve will be identical to the most recently observed curve. The results presented in Figure 4.9 shows the comparative forecasting accuracy of the RNN model versus the curve persistence (no-change) benchmark for both one-day-ahead and five-day-ahead SOFR term structure forecasts. Across the test set, for the 1-day horizon the model's RMSE is generally higher than the persistence RMSE since the 1-day persistence benchmark is exceptionally strong. For the 5-day horizon, the model and persistence method exhibit similar trend of error. Both methods reach their peak RMSE on the same calendar dates.

To complement level errors in basis points, we report the Mean Absolute Percentage Error (MAPE) and the Root Mean Square Percentage Error (RMSPE) for observations Y_i^{obs} and forecasts $\hat{Y}_i$:

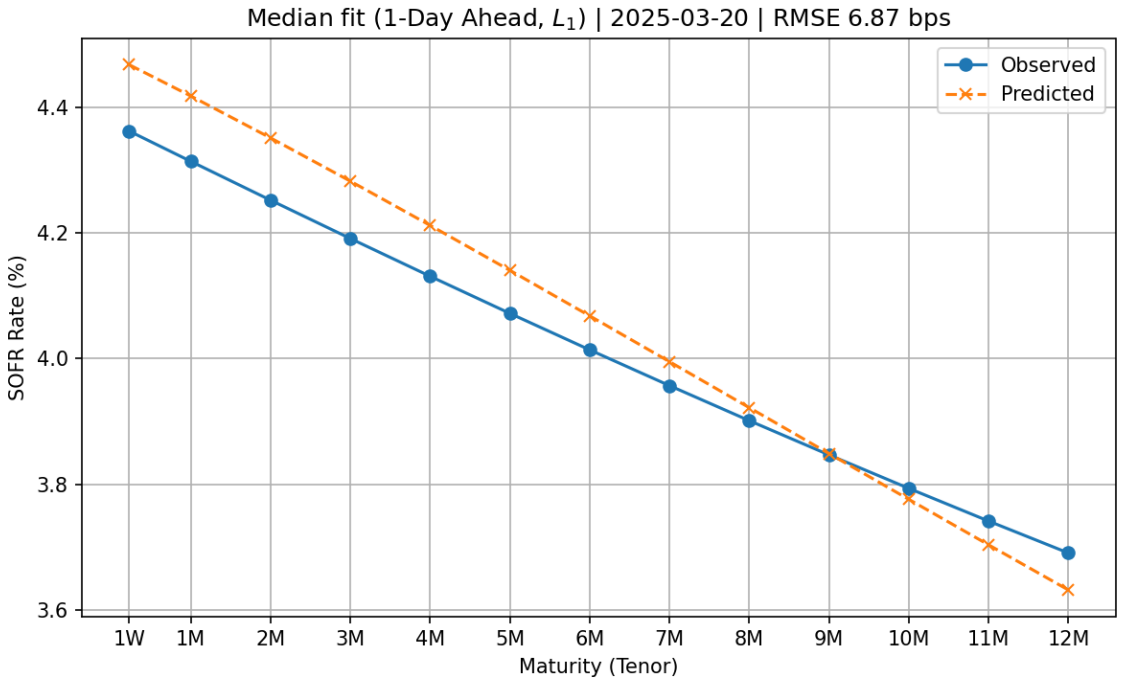
$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i^{\text{obs}} - \hat{Y}_i}{Y_i^{\text{obs}}}\right| \times 100\%, \qquad \text{RMSPE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i^{\text{obs}} - \hat{Y}_i}{Y_i^{\text{obs}}}\right)^2} \times 100\%.$$

Both metrics are scale-independent and thus comparable across tenors with different rate levels. MAPE averages absolute percentage deviations and is easy to interpret, while RMSPE squares percentage errors and therefore places more weight on larger misses.

Table 4.1 summarizes the forecast accuracy of the RNN model across different tenors and forecast horizons, measured using both MAPE and RMSPE under L_2 and L_1 loss functions. Figure 4.10 shows that both MAPE and RMSPE exhibit a U-shaped pattern: errors decrease from 1 week to a minimum at intermediate maturities (3 months to 4 months), then rise again toward 12 months. As shown in Figure 4.11, both the persistence benchmark and the

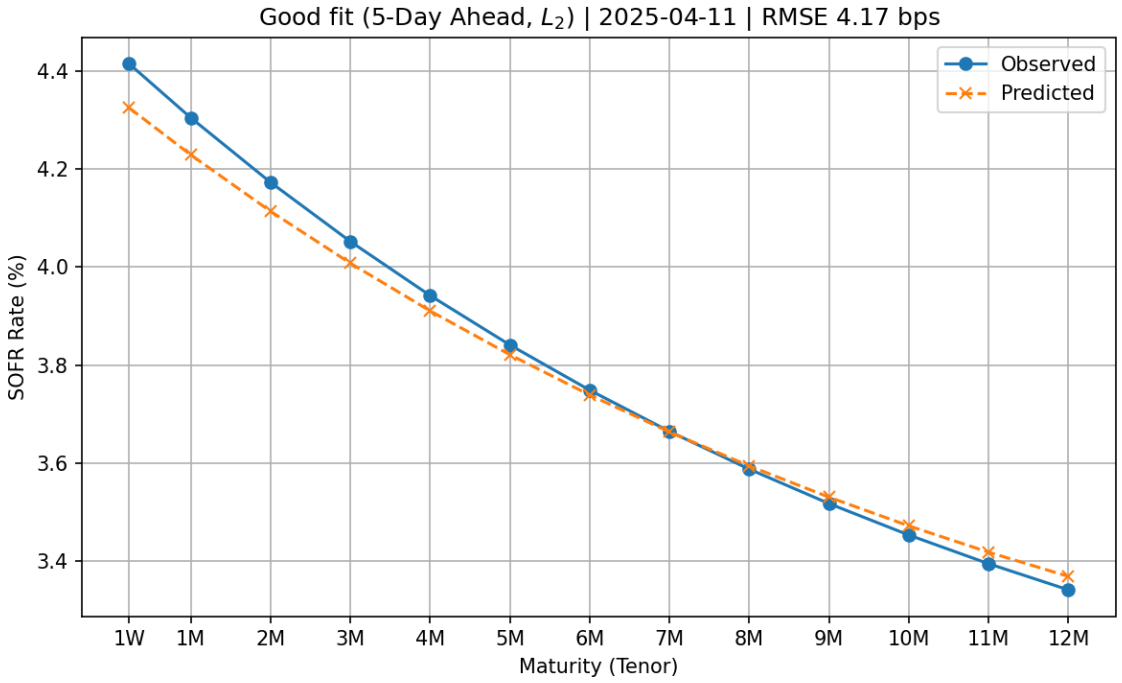


(a) A typical "good fit" day (RMSE = 2.19 bps, 2025-03-28).

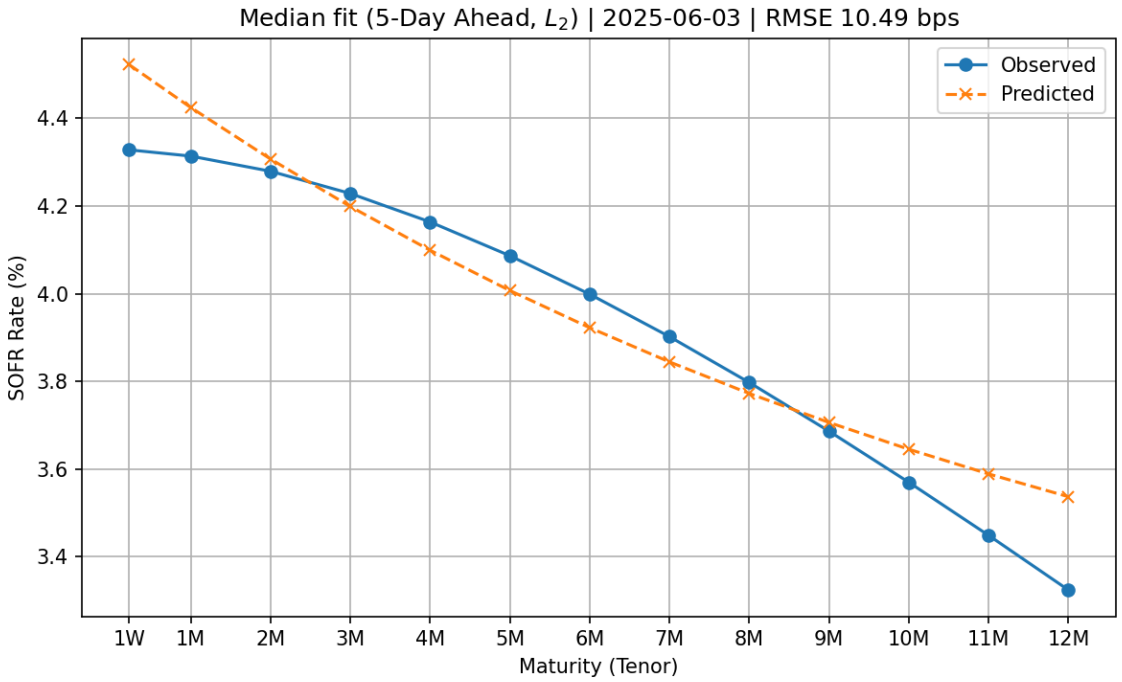


(b) A typical "median fit" day (RMSE = 6.87 bps, 2025-03-20).

Figure 4.6: Observed vs. predicted SOFR term structure for two days from the 1-day-ahead test set (L_2 loss).



(a) A typical "good fit" day (RMSE = 4.17 bps, 2025-04-11).



(b) A typical "median fit" day (RMSE = 10.49 bps, 2025-06-03).

Figure 4.7: Observed vs. predicted SOFR term structure for two 5-day-ahead forecasts from the test set (L_1 loss).

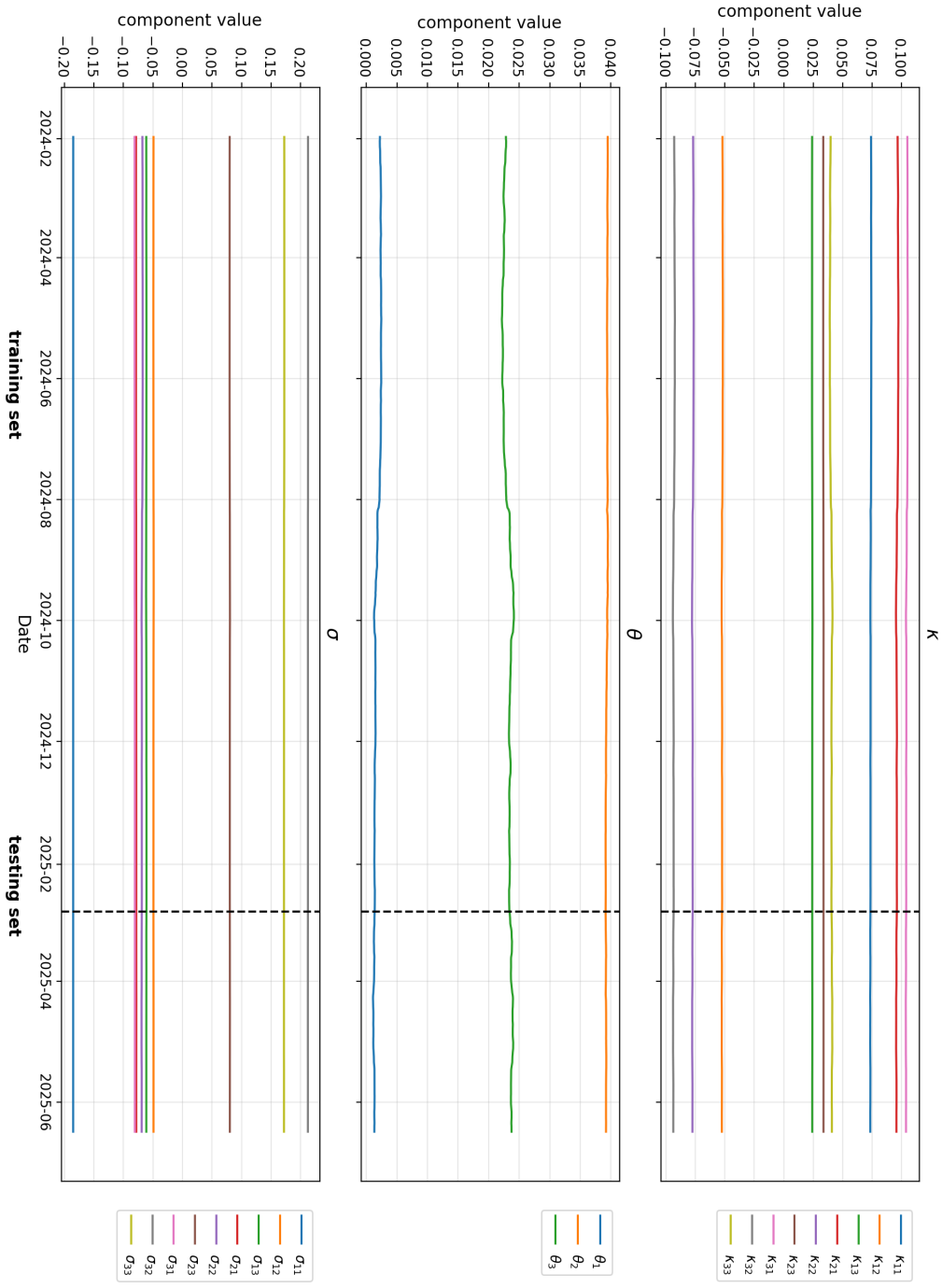
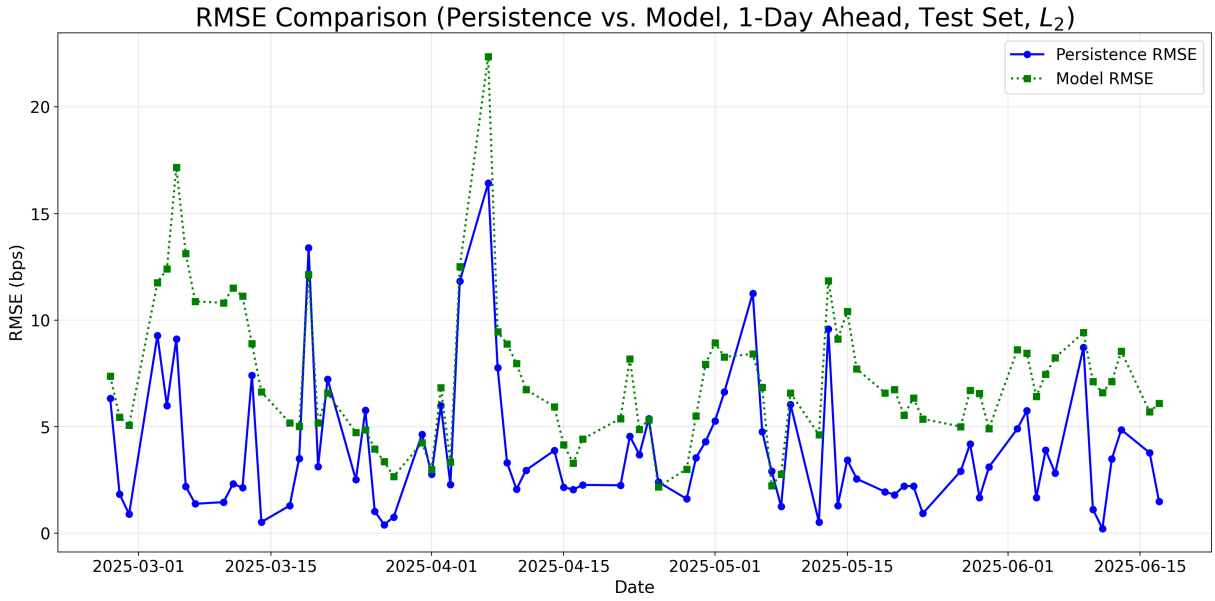
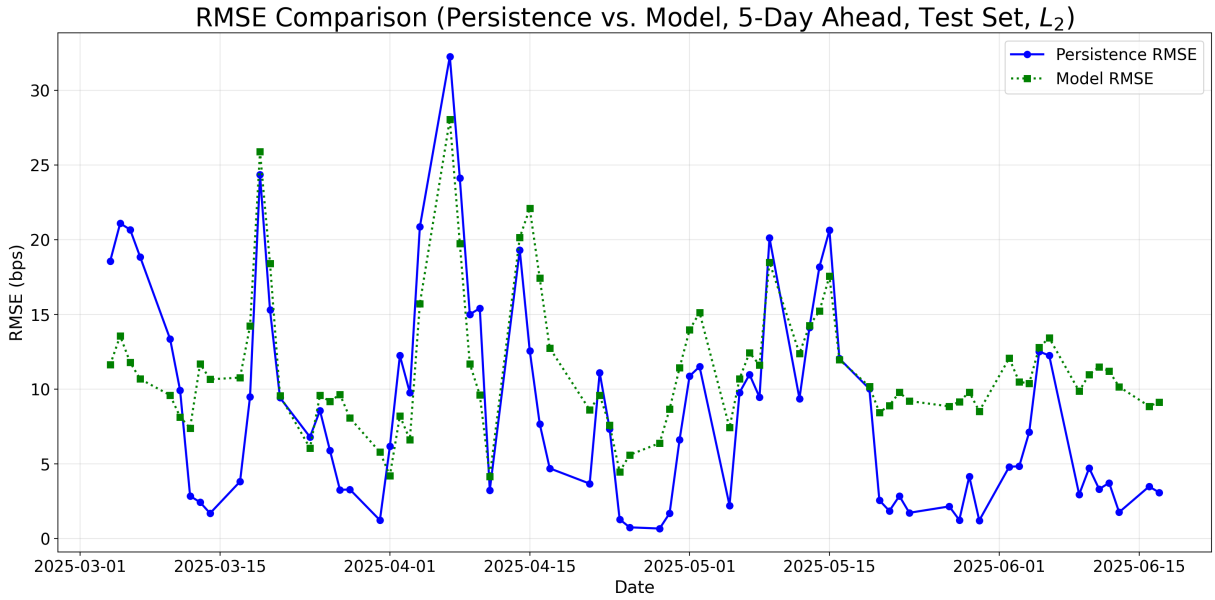


Figure 4.8: Dynamic evolution of Extended Vašíček model state parameters



(a) 1-Day Ahead Forecasts



(b) 5-Day Ahead Forecasts

Figure 4.9: Curve Persistence RMSE vs. Model RMSE

model's predictions tend to replicate the overall shape of the most recently observed SOFR curve. This behavior is especially evident when market conditions remain stable, resulting in accurate forecasts and low errors across all. However, when the actual SOFR curve undergoes a structural change, such as a shift in level, slope, or curvature, the predicted curve often retains the previous shape. As a result, the predicted curve can deviate substantially from the observed target. These deviations are most pronounced at the short and long ends of the curve, where shape changes tend to be amplified. Consequently, forecast errors increase at these, contributing to the U-shaped error pattern.

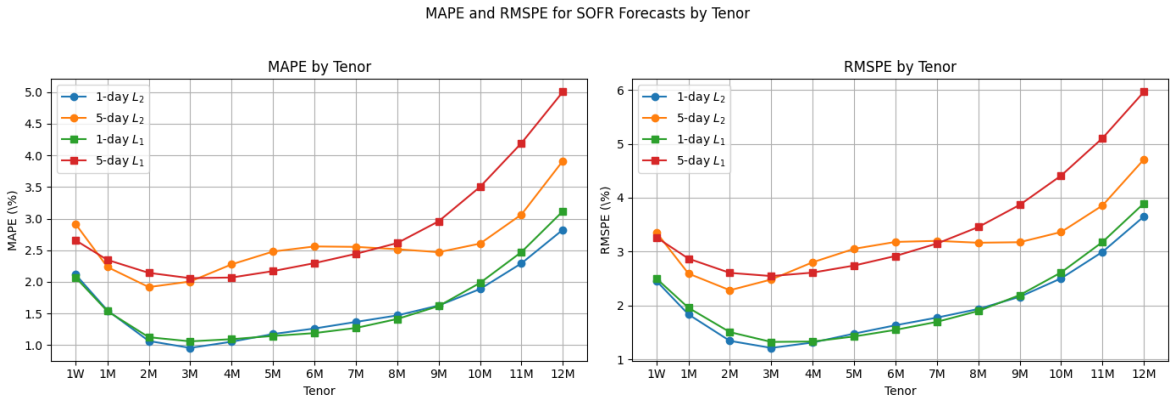


Figure 4.10: MAPE and RMSPE for SOFR Forecasts by Tenor.

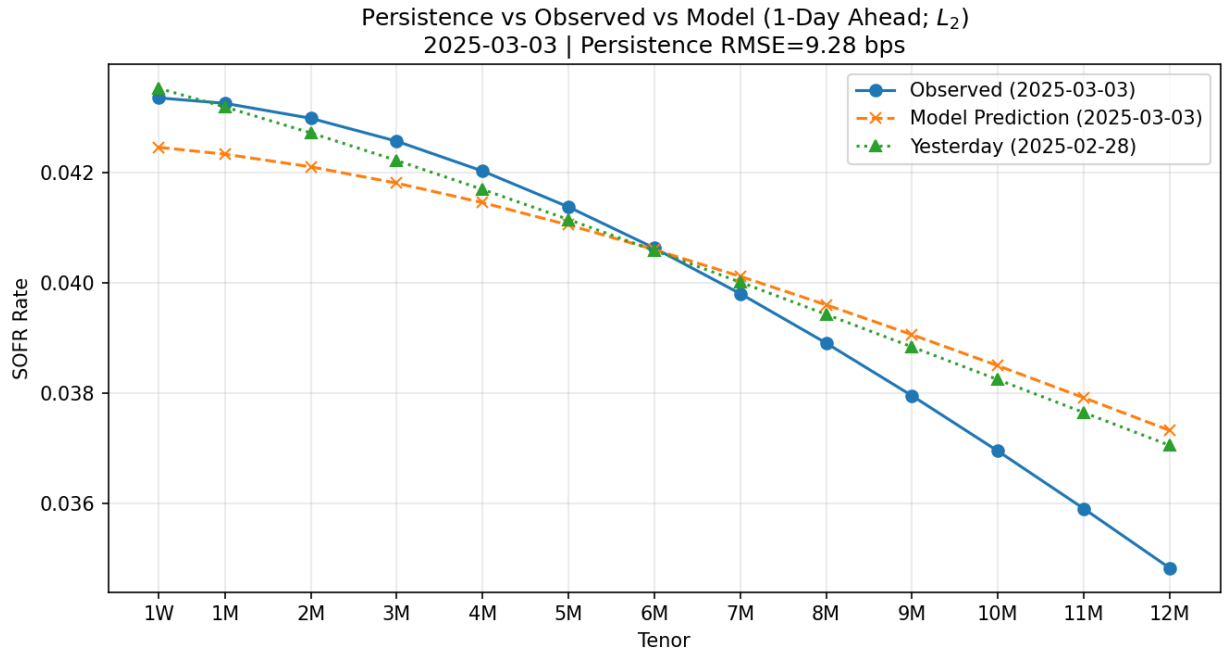


Figure 4.11: Comparison of curve from persistence, observe and model

Tables 4.2, 4.3, 4.4, and 4.5 reports errors across tenor and datasets (train vs. test). The

Table 4.1: SOFR Model Test Errors: MAPE and RMSPE for 1-day and 5-day Ahead Forecasts, L_2 loss (top) and L_1 loss (bottom)

Tenors	MAPE (%)		RMSPE (%)	
	1-day	5-day	1-day	5-day
L_2 norm (MSE-based)				
1W	2.1174622	2.9165883	2.4509470	3.3503340
1M	1.5496132	2.2351022	1.8333876	2.5913348
2M	1.0657535	1.9176404	1.3422042	2.2825112
3M	0.9563673	2.0061310	1.2116110	2.4816730
4M	1.0542395	2.2766452	1.3118292	2.8017845
5M	1.1756610	2.4796734	1.4764854	3.0511074
6M	1.2630296	2.5601325	1.6319817	3.1790707
7M	1.3667045	2.5532744	1.7740269	3.1986547
8M	1.4701335	2.5167053	1.9338800	3.1643650
9M	1.6263856	2.4696805	2.1603003	3.1740480
10M	1.8868936	2.6051104	2.5024720	3.3626610
11M	2.2930540	3.0644370	2.9946244	3.8560443
12M	2.8246295	3.9119318	3.6518680	4.7109860
L_1 norm (MSE-based)				
1W	2.0717068	2.6579041	2.4960900	3.2611735
1M	1.5368519	2.3477770	1.9560751	2.8687215
2M	1.1255457	2.1426282	1.5047694	2.6044817
3M	1.0600957	2.0587358	1.3226506	2.5454469
4M	1.0944982	2.0691526	1.3302700	2.6093936
5M	1.1454170	2.1703910	1.4237878	2.7404892
6M	1.1904736	2.2961960	1.5467414	2.9179263
7M	1.2713852	2.4430583	1.6952789	3.1483536
8M	1.4136631	2.6138140	1.8972000	3.4548514
9M	1.6172191	2.9573705	2.1917415	3.8665745
10M	1.9839088	3.5015604	2.6122966	4.4102900
11M	2.4714074	4.1934500	3.1779175	5.1057363
12M	3.1163316	5.0038843	3.8953786	5.9652386

Table 4.2: SOFR Model Test Errors: MAPE and RMSPE for 1-day Forecasts, L_2 loss

Tenor	MAPE (%)		RMSPE (%)	
	Train	Test	Train	Test
1W	1.6655287	2.1174622	2.1017737	2.4509470
1M	1.3391687	1.5496132	1.7047370	1.8333876
2M	1.0678953	1.0657535	1.3945222	1.3422042
3M	0.9632064	0.9563673	1.2847438	1.2116110
4M	0.9636767	1.0542395	1.3152254	1.3118292
5M	1.0069411	1.1756610	1.4353986	1.4764854
6M	1.1010287	1.2630296	1.6294873	1.6319817
7M	1.3123723	1.3667045	1.9068304	1.7740269
8M	1.6135497	1.4701335	2.2832015	1.9338800
9M	2.0033264	1.6263856	2.7695446	2.1603003
10M	2.5053920	1.8868936	3.3696290	2.5024720
11M	3.1447709	2.2930540	4.0823383	2.9946244
12M	3.8739913	2.8246295	4.9044495	3.6518680

lowest errors are consistently observed at intermediate tenor (roughly 3–6M), with larger errors at both the front and back ends of the curve. Comparing loss functions, the 1-day models under L_1 and L_2 are similar. L_2 is modestly lower across mid-to-long tenors, while L_1 can be slightly better at the very short end. For the 5-day horizon, L_2 generally yields lower MAPE and RMSPE, whereas L_1 is competitive near the front end. As expected, 5-day-ahead forecasts are less accurate than 1-day-ahead forecasts across all tenor. Overall, the pattern of errors is stable across settings: intermediate tenor are the most predictable, with persistent challenges at the curve extremities. Future methodological enhancements could focus on integrating dynamic weighting across tenor or specialized regularization at the endpoints to further reduce the error and enhance the model robustness.

Let MAE_i denote the mean absolute error of the predicted SOFR curve on day i for $i = 1, \ldots, N$, and let ε be a fixed threshold (in basis points), typically, we choose as $\varepsilon \in \{5, 8, 11\}$. The hit rate at threshold ε is defined as

$$\text{Hit Rate}(\varepsilon) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{\text{MAE}_i \leq \varepsilon\}},$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function that equals 1 if the condition is true and 0 otherwise, and N is the total number of test days.

Table 4.6 demonstrates that predictive accuracy is higher for the one-day-ahead models. At the 8 bps threshold, both L_2 and L_1 achieve a hit rate of 76.9%. At a tighter threshold (5 bps), L_1 is marginally higher in the 1-day case, whereas at 11 bps L_2 is higher. In the five-day-ahead case, hit rates are lower overall. L_1 exceeds L_2 at 5 bps, but L_2 leads at 8 bps and 11 bps. Overall, short-horizon forecasts are strongest, and neither loss dominates across thresholds: L_1 helps at tight tolerances, while L_2 performs better at moderate/looser tolerances, especially for 5-day forecasts.

Table 4.3: SOFR Model Test Errors: MAPE and RMSPE for 1-day Forecasts, L_1 loss

Tenor	MAPE (%)		RMSPE (%)	
	Train	Test	Train	Test
1W	1.58996320	2.0717068	2.0916188	2.4960900
1M	1.21977620	1.5368519	1.6653379	1.9560751
2M	0.99263840	1.1255457	1.3878578	1.5047694
3M	0.96294034	1.0600957	1.3531432	1.3226506
4M	1.03244030	1.0944982	1.4273131	1.3302700
5M	1.11093690	1.1454170	1.5190018	1.4237878
6M	1.14448770	1.1904736	1.6069325	1.5467414
7M	1.16750300	1.2713852	1.7183949	1.6952789
8M	1.27482900	1.4136631	1.9068005	1.8972000
9M	1.52453390	1.6172191	2.2250018	2.1917415
10M	1.90420990	1.9839088	2.7013760	2.6122966
11M	2.44240710	2.4714074	3.3373165	3.1779175
12M	3.10567400	3.1163316	4.1202884	3.8953786

Table 4.4: SOFR Model Test Errors: MAPE and RMSPE for 5-day Forecasts, L_2 loss

Tenor	MAPE (%)		RMSPE (%)	
	Train	Test	Train	Test
1W	5.3497530	2.9165883	6.2979380	3.3503340
1M	4.0803237	2.2351022	5.0400350	2.5913348
2M	2.7809563	1.9176404	3.6677103	2.2825112
3M	1.8957523	2.0061310	2.6826117	2.4816730
4M	1.6450083	2.2766452	2.3023686	2.8017845
5M	1.9472384	2.4796734	2.6184068	3.0511074
6M	2.5098710	2.5601325	3.3572133	3.1790707
7M	3.1826143	2.5532744	4.2670590	3.1986547
8M	3.8640625	2.5167053	5.2482820	3.1643650
9M	4.5328640	2.4696805	6.2734230	3.1740480
10M	5.2190840	2.6051104	7.3422100	3.3626610
11M	5.9448013	3.0644370	8.4652950	3.8560443
12M	6.7917824	3.9119318	9.6580920	4.7109860

Table 4.5: SOFR Model Test Errors: MAPE and RMSPE for 5-day Forecasts, L_1 loss

Tenor	MAPE (%)		RMSPE (%)	
	Train	Test	Train	Test
1W	5.1293800	2.6579041	7.0383472	3.2611735
1M	4.2546880	2.3477770	5.889863	2.8687215
2M	3.2888076	2.1426282	4.5363700	2.6044817
3M	2.5343082	2.0587358	3.4084842	2.5454469
4M	2.0213807	2.0691526	2.6768537	2.6093936
5M	1.9397024	2.1703910	2.5819680	2.7404892
6M	2.2091317	2.2961960	3.1050968	2.9179263
7M	2.7010145	2.4430583	3.9614365	3.1483536
8M	3.2892764	2.6138140	4.9515500	3.4548514
9M	3.9606876	2.9573705	5.9939423	3.8665745
10M	4.6798124	3.5015604	7.0614870	4.4102900
11M	5.4537973	4.1934500	8.1499090	5.1057363
12M	6.3099220	5.0038843	9.2654740	5.9652386

Table 4.6: Hit Rate (%) for Test set and MAE Thresholds

Configuration	**5 bps** (%)	**8 bps** (%)	**11 bps** (%)
1-Day L_2	39.74358974	76.92307692	96.15384615
1-Day L_1	41.02564103	76.92307692	92.30769231
5-Day L_2	5.405405405	37.83783784	72.97297297
5-Day L_1	12.16216216	35.13513514	60.81081081

Tables 4.6, 4.7, 4.8, 4.9, and 4.10 report hit rates by tenor, error threshold, and configuration. Consistent with expectations, model with shorter prediction horizons yield higher hit rates, especially at stricter thresholds. Within the 1-day horizon, L_1 and L_2 are broadly similar: L_1 is slightly higher at very short tenor and tight tolerances (3 to 5 bps), while L_2 matches or exceeds L_1 at mid to long tenor and at looser thresholds (8 to 11 bps). For the 5-day horizon, hit rates are lower overall. L_1 often leads at 3 to 5 bps, whereas L_2 is comparable or higher at 8 to 14 bps. Across both horizons, the best performance concentrates around 2M to 6M, with the short and long tenors remaining the most challenging.

Figures 4.12a, 4.12b illustrate the worst-fit curve for the SOFR term structure forecasts using four configurations. Notably, both the 1-day-ahead and 5-day-ahead forecasts on April 7, 2025, exhibit significant errors, with RMSE values that are among the highest in the test set. This poor fit can be directly attributed to the consequence of the sudden tariff hike announced on April 2, which triggered a substantial policy-driven shift in the SOFR curve. Both the model forecasts and the persistence benchmark struggle in this environment, as their predictions largely reflect the pre-shock regime. Only the last two days of the model's input window (April 3 and April 4) contain information about the initial market reaction to the policy change, which was still evolving. This example shows how abrupt policy shifts can induce discontinuities in the term structure, undermining the effectiveness of both naive

Table 4.7: Hit Rate (%) by Tenor for Different RMSE Thresholds (1-Day L_2 Model)

Tenor	**3 bps** (%)	**5 bps** (%)	**8 bps** (%)	**11 bps** (%)
1W	12.82051282	24.35897436	42.30769231	61.53846154
1M	24.35897436	37.17948718	69.23076923	82.05128205
2M	47.43589744	62.82051282	82.05128205	93.58974359
3M	43.58974359	65.38461538	89.74358974	97.43589744
4M	46.15384615	58.97435897	88.46153846	96.15384615
5M	37.17948718	62.82051282	80.76923077	97.43589744
6M	33.33333333	57.69230769	76.92307692	93.58974359
7M	37.17948718	55.12820513	80.76923077	92.30769231
8M	33.33333333	53.84615385	76.92307692	92.30769231
9M	32.05128205	51.28205128	75.64102564	88.46153846
10M	30.76923077	48.71794872	70.51282051	78.20512821
11M	23.07692308	44.87179487	64.10256410	74.35897436
12M	20.51282051	41.02564103	51.28205128	69.23076923

Table 4.8: Hit Rate (%) by Tenor for Different MAE Thresholds (1-Day L_1 Model)

Tenor	**3 bps** (%)	**5 bps** (%)	**8 bps** (%)	**11 bps** (%)
1W	16.66666667	25.64102564	48.71794872	70.51282051
1M	26.92307692	46.15384615	64.10256410	82.05128205
2M	47.43589744	61.53846154	74.35897436	92.30769231
3M	43.58974359	62.82051282	84.61538462	96.15384615
4M	38.46153846	61.53846154	83.33333333	96.15384615
5M	37.17948718	60.25641026	85.89743590	96.15384615
6M	42.30769231	57.69230769	83.33333333	94.87179487
7M	44.87179487	55.12820513	80.76923077	93.58974359
8M	39.74358974	56.41025641	76.92307692	89.74358974
9M	33.33333333	52.56410256	73.07692308	85.8974359
10M	32.05128205	44.87179487	67.94871795	76.92307692
11M	15.38461538	37.17948718	61.53846154	71.79487179
12M	11.53846154	25.64102564	50	64.10256410

Table 4.9: Hit Rate (%) by Tenor for Different RMSE Thresholds (5-Day L_2 Model)

Tenor	**3 bps** (%)	**5 bps** (%)	**8 bps** (%)	**11 bps** (%)	**14 bps** (%)
1W	8.108108108	16.21621622	31.08108108	45.94594595	60.81081081
1M	14.86486486	27.02702703	41.89189189	54.05405405	81.08108108
2M	16.21621622	29.72972973	55.40540541	77.02702703	89.18918919
3M	24.32432432	37.83783784	51.35135135	70.27027027	79.72972973
4M	18.91891892	36.48648649	52.70270270	64.86486486	71.62162162
5M	18.91891892	35.13513514	47.29729730	59.45945946	64.86486486
6M	21.62162162	33.78378378	45.94594595	58.10810811	70.27027027
7M	21.62162162	35.13513514	47.29729730	56.75675676	70.27027027
8M	20.27027027	29.72972973	52.70270270	63.51351351	78.37837838
9M	17.56756757	36.48648649	52.70270270	70.27027027	79.72972973
10M	21.62162162	35.13513514	58.10810811	63.51351351	79.72972973
11M	17.56756757	28.37837838	41.89189189	60.81081081	71.62162162
12M	9.459459459	16.21621622	33.78378378	45.94594595	60.81081081

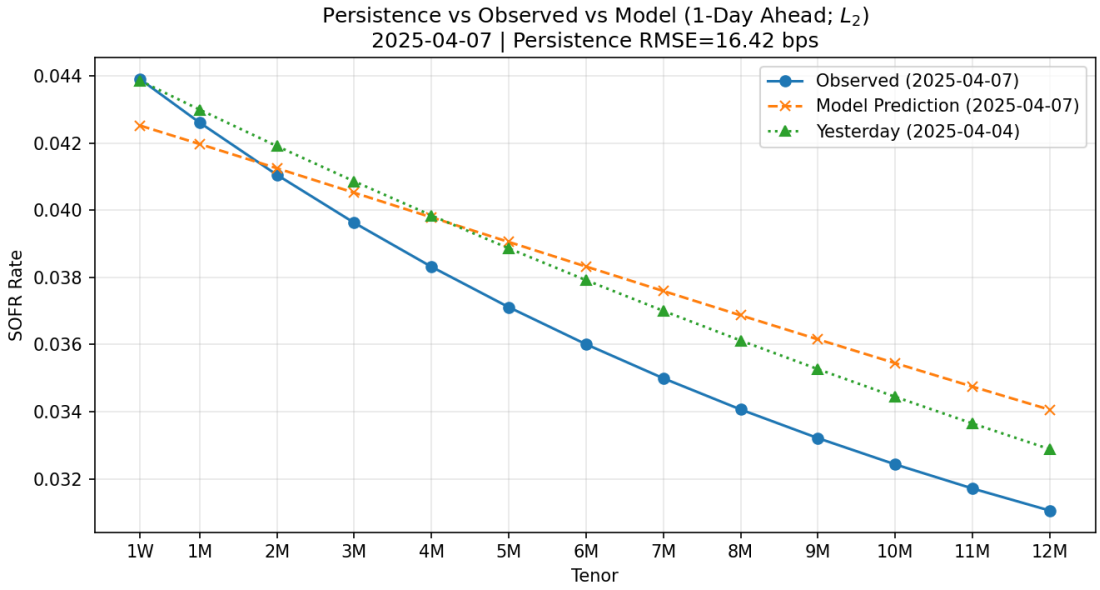
Table 4.10: Hit Rate (%) by Tenor for Different MAE Thresholds (5-Day L_1 Model)

Tenor	**3 bps** (%)	**5 bps** (%)	**8 bps** (%)	**11 bps** (%)	**14 bps** (%)
1W	13.51351351	22.97297297	37.83783784	58.10810811	70.27027027
1M	12.16216216	21.62162162	47.29729730	66.21621622	77.02702703
2M	16.21621622	28.37837838	48.64864865	68.91891892	83.78378378
3M	21.62162162	36.48648649	51.35135135	68.91891892	78.37837838
4M	22.97297297	39.18918919	50	64.86486486	81.08108108
5M	25.67567568	37.83783784	48.64864865	71.62162162	82.43243243
6M	20.27027027	27.02702703	51.35135135	70.27027027	82.43243243
7M	16.21621622	33.78378378	52.7027027	67.56756757	85.13513514
8M	20.27027027	33.78378378	51.35135135	63.51351351	75.67567568
9M	20.27027027	32.43243243	47.2972973	62.16216216	71.62162162
10M	12.16216216	24.32432432	40.54054054	55.40540541	63.51351351
11M	9.459459459	17.56756757	28.37837838	45.94594595	52.7027027
12M	12.16216216	16.21621622	22.97297297	32.43243243	40.54054054

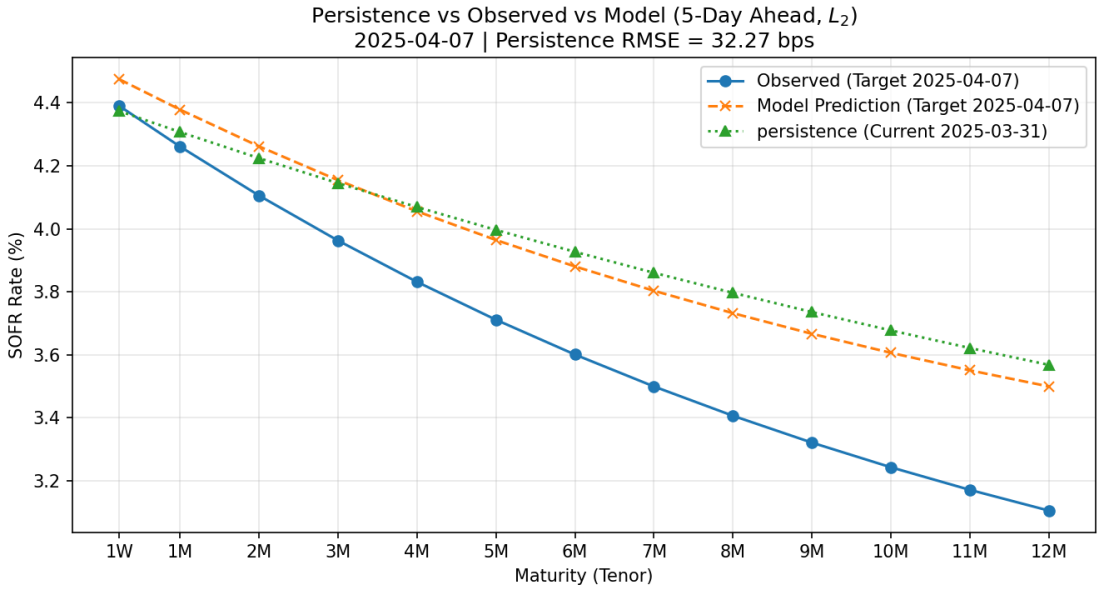
and more advanced forecasting approaches.

In this chapter, we implemented a data-driven framework for modeling and forecasting the SOFR term structure. Starting from CME SOFR futures data, we constructed daily forward rate curves and estimated Nelson–Siegel factors, which were then modeled with a recurrent neural network and Kalman filtering. The empirical results show that the model achieves low RMSE values for one-day-ahead forecasts, typically between 3 and 10 basis points across the test set, with median errors around 8 basis points. For five-day-ahead forecasts, the model's performance declines as expected. The RMSE values are noticeably higher, often ranging from 10 to 20 basis points and reaching up to 35 basis points during episodes of market volatility or structural curve changes.

Across all forecast horizons and loss functions, a pronounced U-shaped error pattern emerges: errors are lowest at intermediate tenor (3 to 4 months) and higher at the short and long ends of the curve. The reason for the U-shaped error is that after sudden shifts in the SOFR curve, when seeking to minimize overall RMSE, the model tends to match the middle section of the curve more closely, resulting in larger errors at the short and long ends. The model provides reliable fits but shows slightly higher errors than persistence during stable periods. The model struggles to capture abrupt structural shifts, particularly at the short and long ends of the curve. These findings highlight both the strengths and the limitations of the methodology and suggest direction for further enhancing the model's ability to adapt to regime shifts in the term structure.



(a) Worst fit curve for 1-day-ahead forecasting (L_2 loss)



(b) Worst fit curve for 5-day-ahead forecasting (L_2 loss)

Figure 4.12: Observed vs predicted SOFR curves on April 7, 2025, showing the worst fit in the test set for (a) 1-day-ahead and (b) 5-day-ahead L_2 forecasting.

Chapter 5

Conclusion and Future Research

This thesis has developed and evaluated a comprehensive machine learning framework for modeling and forecasting the SOFR term structure, with a particular focus on CME SOFR futures. The main contribution of this work is to provide a transparent and systematic modeling pipeline for SOFR, integrating well-established financial methods with modern machine learning approaches. By constructing a daily, piecewise-constant SOFR forward curve using the official CME methodology and fitting it to the dynamic Nelson–Siegel model, we extract time series of the key SOFR curve factors: level, slope, and curvature. The subsequent application of recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) architectures integrated with Kalman filtering, enables the modeling of the temporal evolution of these factors and provides robust short-term SOFR curve forecasts.

The results highlight the importance of robust training strategies. The training and test loss curves are closely aligned and converge after 30 epochs, indicating that the models do not suffer from overfitting. Both the L_2 and L_1 loss functions yield similar trends, with L_1 providing slightly more robustness to outliers. The proposed approach demonstrates strong out-of-sample performance under both mean squared error (MSE) and mean absolute error (MAE) loss functions. Results show that the model reliably captures the daily dynamics of the SOFR term structure for both one-day and five-day-ahead forecasts, achieving root mean squared errors (RMSE) that generally remain below those of the persistence benchmark. The advantage of the RNN-based model is especially pronounced during periods of structural change or discontinuity in the SOFR curve, where the persistence approach fails to anticipate abrupt shifts. The findings confirm that the dynamic machine learning model provides measurable improvements in adaptability and predictive accuracy relative to naive persistence benchmarks, even though fully capturing abrupt policy-driven changes remains a challenge.

A significant and timely direction for future research is the explicit modeling of discontinuities in SOFR term structures, particularly those induced by scheduled monetary policy decisions such as FOMC meetings. Traditional continuous-time models often struggle to accurately reflect the pronounced jumps in overnight rates that occur in response to policy actions, leading to potential misestimation of both the risk-free curve and associated derivative prices. As observed in the analysis, the persistence model in particular tends to underperform when the term structure undergoes marked changes in curvature or slope, which is frequently linked to policy-driven discontinuities. Gellert and Schlögl (2021) ad-

dressed this gap by introducing frameworks that integrate both continuous diffusion dynamics and deterministic jumps. The piecewise-continuous short rate model allows the SOFR to evolve smoothly between FOMC meetings while incorporating a deterministic, known jump at each meeting date to reflect the immediate and policy-driven nature of rate changes. This modeling choice is motivated by the institutional structure of U.S. monetary policy and the backward-looking construction of the SOFR, which makes it particularly sensitive to scheduled policy announcements. Building on these insights, Schlögl, Skov, and Skovmand (2024) demonstrated that standard continuous models are clearly misspecified around FOMC meetings and fail to jointly fit overnight and futures rates during recent periods of monetary tightening, whereas a jump-augmented model accurately captures the observed dynamics. The importance of scheduled jumps is further formalized by Fontana, Grbac, and Schmidt (2024), who extended the HJM framework to accommodate stochastic discontinuities at pre-specified calendar dates and derive explicit pricing formulas for bonds and caplets. Collectively, these studies highlight that faithfully modeling policy-driven jumps is essential for accurate pricing and risk management of SOFR-linked products in the post-LIBOR era.

In addition to better modeling discontinuities, another research avenue is to improve the way models capture changes in the curvature and slope of the SOFR curve by introducing a more flexible, period-specific weighting scheme during training or loss calculation. Rather than treating all maturities equally, future models could assign distinct weights to different segments of the curve—such as 1 week to 3 months, 4–6 months, 7–9 months, and 10–12 months—allowing the learning process to more effectively address changes in different parts of the curve. This approach could help the model adapt to localized changes in the shape of the SOFR curve and improve overall performance, particularly during periods of structural transition or when policy shocks disproportionately affect specific maturities.

Another area of future research is the integration of arbitrage-free regularization into machine learning models of the term structure. As machine learning and deep learning architectures become increasingly common in financial modeling, it is crucial to ensure that their outputs remain consistent with the fundamental economic principle of no-arbitrage. Without such constraints, data-driven models may inadvertently produce SOFR curves or forward rates that violate arbitrage conditions, potentially leading to unrealistic forecasts and unreliable risk measures. Gellert and Schlögl (2021) emphasized the challenges associated with arbitrage violations in the presence of discontinuities, highlighting the need for model structures or regularization terms that enforce economic coherence. Gao (2021) showed that adding an arbitrage-free penalty during neural network training improves both the robustness and economic plausibility of yield curve predictions. Incorporating these approaches into deep learning architectures or state-space models may further enhance out-of-sample forecasting accuracy and produce yield and forward curves that are both realistic and arbitrage-free, which is crucial for pricing, risk management, and regulatory compliance in fixed income markets.

The SOFR market is still in a period of transition and is expected to grow significantly as market participants continue to adapt to the new benchmark. As more liquid SOFR-linked products, including SOFR interest rate swaps and options on these swaps (swaptions), become available and widely traded, the quality and quantity of data will continue to improve. This ongoing development will create new opportunities for research and further refinement of SOFR term structure models. As a result, pricing, risk management, and forecasting in

the evolving fixed income market will become more accurate in the future.

References

Bianchi, Daniele, Matthias Büchner, and Andrea Tamoni (2020). "Bond risk premiums with machine learning". In: *Review of Financial Studies* 34.2, pp. 1046–1089.

Board of Governors of the Federal Reserve System (2025). *Federal Open Market Committee (FOMC)*. URL: https://www.federalreserve.gov/monetarypolicy/fomc.htm.

Bolder, David Jamieson (2015). *Fixed-income portfolio analytics: a practical guide to implementing, monitoring and understanding fixed-income portfolios*. 1st. Springer International Publishing.

Brace, Alan, Dariusz Gatarek, and Marek Musiela (1997). "The market model of interest rate dynamics". In: *Mathematical Finance* 7.2, pp. 127–155.

Brigo, Damiano and Fabio Mercurio (2007). *Interest Rate Models - Theory and practice: with smile, inflation and credit*. 2nd. Springer.

Broyden, Charles George (1970). "The convergence of a class of double-rank minimization algorithms". In: *IMA Journal of Applied Mathematics* 6.1, pp. 76–90.

Christensen, Jens Henrik Eggert, Francis X Diebold, and Glenn David Rudebusch (2011). "The affine arbitrage-free class of Nelson–Siegel term structure models". In: *Journal of Econometrics* 164.1, pp. 4–20.

CME Group (2025). *CME Term SOFR Reference Rates Benchmark Methodology*. URL: https://www.cmegroup.com/market-data/files/cme-term-sofr-reference-rates-benchmark-%20methodology.pdf.

Cox, John Carrington, Jonathan Edwards Ingersoll Jr, and Stephen Alan Ross (1985). "A theory of the term structure of interest rates". In: *Econometrica* 53.2, pp. 385–407.

Date, Paresh and Ksenia Ponomareva (2011). "Linear and non-linear filtering in mathematical finance: a review". In: *IMA Journal of Management Mathematics* 22.3, pp. 195–211.

Diebold, Francis X and Canlin Li (2006). "Forecasting the term structure of government bond yields". In: *Journal of Econometrics* 130.2, pp. 337–364. ISSN: 0304-4076.

Federal Reserve Bank of New York (2025). *Secured Overnight Financing Rate (SOFR)*. URL: https://www.newyorkfed.org/markets/reference-rates/sofr.

Filipović, Damir (1999). "A Note on the Nelson–Siegel Family". In: *Mathematical Finance* 9.4, pp. 349–359.

Fletcher, Roger (1970). "A new approach to variable metric algorithms". In: *The Computer Journal* 13.3, pp. 317–322.

Fontana, Claudio, Zorana Grbac, and Thorsten Schmidt (2024). "Term Structure Modelling with Overnight Rates beyond Stochastic Continuity". In: *Mathematical Finance* 32.1, pp. 151–189.

Gao, Xiang (2021). "Stochastic control, numerical methods, and machine learning in finance and insurance". Ph.D. Thesis. Concordia University.

Gellert, Karol and Erik Schlögl (2021). "Short Rate Dynamics: A Fed Funds and SOFR Perspective". In: *SSRN Electronic Journal*.

Goldfarb, Donald (1970). "A family of variable-metric methods derived by variational means". In: *Mathematics of Computation* 24.109, pp. 23–26.

Heath, David, Robert Jarrow, and Andrew Morton (1992). "Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation". In: *Econometrica* 60.1, pp. 77–105.

Hochreiter, Sepp (1991). *Untersuchungen zu dynamischen neuronalen Netzen*.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long short-term memory". In: *Neural Computation* 9.8, pp. 1735–1780.

Huggins, Doug and Christian Schaller (2022). *SOFR Futures and Options*. Wiley. ISBN: 9781119888956.

Hull, John and Alan White (1990). "Pricing interest-rate-derivative securities". In: *The Review of Financial Studies* 3.4, pp. 573–592.

Ioffe, Sergey and Christian Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *CoRR* abs/1502.03167.

Kalman, Rudolf E. (1960). "A new approach to linear filtering and prediction problems". In: *Journal of Basic Engineering* 82.1, pp. 35–45.

Kingma, Diederik Pieter and Jimmy Ba (2015). "Adam: A Method for Stochastic Optimization". In: *ICLR (Poster)*.

Kratsios, Anastasis and Cody Hyndman (2019). "Deep Arbitrage-Free Learning in a Generalized HJM Framework via Arbitrage-Regularization". In: *Risks* 8.2, p. 40.

Nelson, Charles R and Andrew F Siegel (1987). "Parsimonious modeling of yield curves". In: *Journal of Business* 60.4, pp. 473–489.

Nocedal, Jorge and Stephen J Wright (2006). *Numerical Optimization*. 2nd. Springer.

PyTorch (2024). *torch.nn.LSTM*. https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html.

Schlögl, Erik, Jacob Bjerre Skov, and David Skovmand (2024). "Term Structure Modeling of SOFR: Evaluating the Importance of Scheduled Jumps". In: *International Journal of Theoretical and Applied Finance* 27.02.

Shanno, David Francis (1970). "Conditioning of quasi-Newton methods for function minimization". In: *Mathematics of Computation* 24.111, pp. 647–656.

Skov, Jacob Bjerre and David Skovmand (2021). "Dynamic Term Structure Models for SOFR Futures". In: *Journal of Futures Markets* 41.10, pp. 1520–1544.

Vašíček, Oldřich (1977). "An equilibrium characterization of the term structure". In: *Journal of Financial Economics* 5.2, pp. 177–188.