Regime Dynamics in Overnight Funding Markets: Modeling the SOFR–EFFR Spread

Conrad Voigt

conradvoigt@ucsb.edu

March 2026

Abstract

We model the daily spread between the Secured Overnight Financing Rate (SOFR) and the Effective Federal Funds Rate (EFFR) over 1,986 trading days from April 4, 2018 to March 18, 2026. The spread is non-linear, driven by aggregate reserve scarcity, and exhibits pronounced calendar-driven spikes at month-end, quarter-end, and year-end dates alongside heavy-tailed innovations (excess kurtosis = 758.45). We estimate five models: a linear SARIMA(2,0,1) baseline, a SETAR model with GARCH(1,1) volatility overlay, a two-regime Markov-Switching AR (MS-AR), an MS-AR with time-varying transition probabilities (TVTP) conditioned on reserve balances, and a Heterogeneous Autoregressive (HAR) benchmark. Tsay's (1989) test rejects linearity ($F = 248.47$, $p < 0.0001$), and the SETAR model identifies a reserve threshold of \$1.641 trillion separating abundant and scarce liquidity regimes. Out-of-sample evaluation on a 63-day hold-out set shows the MS-AR model achieves the lowest RMSE of 0.0288 percentage points, outperforming SARIMA (0.0364%) at the 1% level under the Diebold–Mariano test (DM = 3.834, $p < 0.001$). Five-fold rolling cross-validation confirms this ranking (MS-AR mean RMSE: 0.0322% ± 0.0123%).

1 Introduction

The spread between SOFR and EFFR measures the relative cost of secured versus unsecured overnight borrowing in U.S. money markets. SOFR, published daily by the Federal Reserve Bank of New York (FRBNY), is the volume-weighted median rate on overnight Treasury repo transactions (FRBNY, 2018). EFFR is the corresponding unsecured rate at which depository institutions lend reserve balances to one another overnight (Board of Governors, 2021). Their difference reflects collateral scarcity, the relative demand for secured funding, and the efficiency of monetary policy transmission to short-term rates.

The September 2019 repo market disruption motivated this project. SOFR spiked to 5.25% and the spread widened by several hundred basis points in a single session (Copeland et al., 2021; Anbil et al., 2020). The episode demonstrated that overnight funding markets are acutely sensitive to the aggregate level of reserve balances at the Federal Reserve and raised the question of whether linear time series models are adequate for this spread. Afonso et al. (2020) and Logan (2020) have shown that reserve scarcity introduces a non-linear demand curve: once balances fall below an unknown threshold, funding rates become highly responsive to small changes in supply. Separately, calendar-driven balance sheet constraints—the GSIB surcharge, the Liquidity Coverage Ratio (LCR), and leverage ratio reporting—produce systematic spikes at month-end and quarter-end dates (Munyan, 2015; Correa et al., 2021).

We proceed through four methodological stages: (i) a SARIMA baseline estimated via Box–Jenkins identification with AIC-based grid search; (ii) a SETAR model using lagged reserve balances as the threshold variable, with a GARCH(1,1) overlay for conditional volatility; (iii) a Markov-Switching AR model and its TVTP extension, where reserve balances enter the transition probabilities through a logistic link; and (iv) a HAR benchmark following Corsi (2009). All five models are evaluated on a common 63-day hold-out via rolling one-step-ahead forecasts, with pairwise accuracy tested by the Diebold and Mariano (1995) test. The main results are: (1) the spread is stationary but strongly non-linear, with a reserve threshold near $1.641 trillion delineating abundant and scarce regimes; (2) the MS-AR model with probabilistic regime identification achieves the lowest out-of-sample RMSE; and (3) the SETAR regime structure absorbs nearly all conditional heteroscedasticity, rendering the GARCH overlay statistically insignificant.

2 Data

Three series are retrieved from the Federal Reserve Economic Data (FRED) database maintained by the Federal Reserve Bank of St. Louis (`https://fred.stlouisfed.org`):

1. **SOFR** (FRED code: `SOFR`). Daily, published by the FRBNY since April 3, 2018. The raw series contains 1,988 non-missing observations through March 19, 2026. SOFR is the volume-weighted median rate on overnight Treasury repo transactions (FRBNY, 2018).

2. **EFFR** (FRED code: `DFF`). Daily, published by the Board of Governors over the same range, with 2,908 observations. The higher count reflects weekends and Federal holidays on which EFFR is published but SOFR is not. EFFR is the volume-weighted median of overnight federal funds transactions reported via the FR 2420 (Board of Governors, 2021).

3. **Total Reserve Balances** (FRED code: `WRESBAL`). Weekly (Wednesday values) via the H.4.1 statistical release: 416 observations from April 4, 2018 to March 18, 2026, denominated in millions of U.S. dollars. The series measures total balances held by depository institutions at Federal Reserve Banks (Federal Reserve, 2026).

Frequency alignment. Weekly reserve balances are forward-filled to daily frequency, reflecting the assumption that aggregate system liquidity is approximately constant between reporting dates. After aligning all three series to a business-day calendar and dropping non-trading days, the dataset contains **1,986 observations** from April 4, 2018 to March 18, 2026.

The dependent variable is the daily spread $y_t = \mathrm{SOFR}_t - \mathrm{EFFR}_t$, measured in percentage points. Summary statistics appear in Table 1.

Table 1: Summary statistics: SOFR–EFFR daily spread (%).

Statistic	Value
Observations	1,986
Mean	−0.0020
Std. deviation	0.0844
Minimum	−0.1500
25th percentile	−0.0300
Median	−0.0100
75th percentile	0.0100
Maximum	2.9500
Skewness	22.49
Excess kurtosis	758.45

The skewness of 22.49 and excess kurtosis of 758.45 are driven by rare but large positive spikes at quarter-end and year-end dates. The Augmented Dickey–Fuller test yields −4.878 ($p < 0.0001$, 20 lags by AIC), rejecting the unit root null against critical values of −3.434 (1%), −2.863 (5%), and −2.568 (10%). The spread is $I(0)$; no differencing is applied.

3 Methodology

3.1 SARIMA$(p, d, q) \times (P, D, Q)_s$ Model

The Seasonal ARIMA model of Box and Jenkins (1970) serves as the linear baseline:

$$\Phi_P(B^s)\,\phi_p(B)\,(1-B)^d\,(1-B^s)^D\,y_t = \Theta_Q(B^s)\,\theta_q(B)\,\varepsilon_t, \qquad \varepsilon_t \sim \mathrm{WN}(0,\sigma^2), \tag{1}$$

where B is the backshift operator, $\phi_p(B)$ and $\theta_q(B)$ are the non-seasonal AR and MA polynomials, $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ are seasonal polynomials with period s, and d, D are differencing orders. Identification follows the Box–Jenkins procedure: the ACF and PACF motivate candidate orders, and a grid search over $p \in \{0, \ldots, 4\}$, $q \in \{0, \ldots, 4\}$, $P \in \{0, 1\}$, $Q \in \{0, 1\}$ with $s = 21$ trading days is conducted. Given the ADF stationarity result, $d = D = 0$. Selection is by AIC (Akaike, 1974). Estimation is by Gaussian QMLE, which yields consistent coefficient estimates under non-Gaussian innovations (Taniguchi and Kakizawa, 2000).

3.2 SETAR Model with GARCH(1,1) Volatility

The SETAR model of Tong (1983) permits the AR dynamics to switch based on whether a threshold variable exceeds a critical value. The threshold variable is the lagged reserve balance R_{t-d}:

$$y_t = \begin{cases} \mu_1 + \sum_{j=1}^{p} \phi_{1,j}\, y_{t-j} + \varepsilon_{1,t}, & \text{if } R_{t-d} > c \quad \text{(abundant regime)}, \\ \mu_2 + \sum_{j=1}^{p} \phi_{2,j}\, y_{t-j} + \varepsilon_{2,t}, & \text{if } R_{t-d} \le c \quad \text{(scarce regime)}, \end{cases} \tag{2}$$

where c is the threshold and d is the delay. The optimal (d^*, c^*) is found by grid search over $d \in \{1, \ldots, 5\}$ and 200 percentile-based candidates, minimizing the pooled SSR. The AR order is $p = 3$ from the PACF. Tsay's (1989) F-test (Tsay, 1989) formally tests for threshold non-linearity prior to estimation.

A GARCH(1,1) model (Bollerslev, 1986) is fitted to the SETAR residuals in a second step:

$$\sigma_t^2 = \omega + \alpha\,\varepsilon_{t-1}^2 + \beta\,\sigma_{t-1}^2, \tag{3}$$

estimated by QMLE under a Student-t distribution. This two-step procedure is convenient but introduces generated-regressor bias in the GARCH standard errors (Pagan, 1984).

3.3 Markov-Switching AR Model

The MS-AR model of Hamilton (1989) replaces the deterministic threshold with a latent Markov chain $\{s_t\}$:

$$y_t = \mu_{s_t} + \sum_{j=1}^{3} \phi_{j,s_t}\, y_{t-j} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2_{s_t}), \tag{4}$$

where $s_t \in \{0, 1\}$ follows a first-order Markov chain with transition probabilities $P(s_t = j \mid s_{t-1} = i) = p_{ij}$. Estimation is by EM with 20 random restarts. Smoothed regime probabilities are computed via the Kim (1994) smoother.

A TVTP extension models the transition probabilities as logistic functions of the lagged reserve balance:

$$P(s_t = 1 \mid s_{t-1} = 0, R_{t-1}) = \frac{\exp(\gamma_0 + \gamma_1 R_{t-1})}{1 + \exp(\gamma_0 + \gamma_1 R_{t-1})}, \tag{5}$$

allowing reserves to influence regime transition dynamics directly rather than through a static threshold.

3.4 Heterogeneous Autoregressive (HAR) Model

The HAR model of Corsi (2009) regresses the spread on its lagged daily, weekly, and monthly averages:

$$y_t = \beta_0 + \beta_1\, y_{t-1} + \beta_5\, \bar{y}_{t-5:t-1} + \beta_{21}\, \bar{y}_{t-21:t-1} + \varepsilon_t, \tag{6}$$

where $\bar{y}_{t-h:t-1} = h^{-1} \sum_{j=1}^{h} y_{t-j}$. The specification captures multi-horizon persistence through a simple OLS framework.

3.5 Out-of-Sample Evaluation

All five models are evaluated on a common **63-day hold-out set** via rolling one-step-ahead predictions with training-window-only estimation. Forecast accuracy is measured by RMSE. Pairwise comparisons use the Diebold and Mariano (1995) test at the 1-step horizon. A **5-fold rolling-window cross-validation** with 63-day non-overlapping windows assesses stability across different market conditions.

4 Results

4.1 Exploratory Analysis and Calendar Effects

Figure 1 plots the spread over the full sample. The series is centered near zero for most of the period, with isolated positive spikes that coincide with month-end and quarter-end dates. This pattern is consistent with the regulatory balance sheet dynamics documented by Munyan (2015): banks reduce repo exposure around reporting windows to manage their GSIB surcharge and LCR requirements.

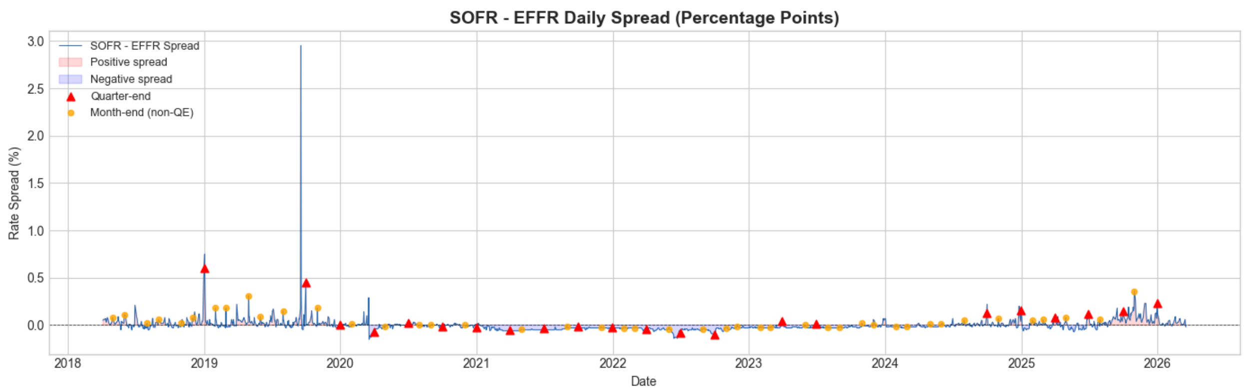

Figure 1: Daily SOFR–EFFR spread (percentage points), April 2018–March 2026. Red triangles: quarter-end dates. Orange circles: month-end (non-quarter-end) dates. Positive (negative) spread regions shaded red (blue).

Table 2 stratifies the spread by calendar day type. Normal days ($n = 1{,}497$) have a mean of -0.0077% and standard deviation of 0.0401%. Quarter-turn days ($n = 85$) exhibit a mean of 0.0287% with $2.7\times$ the normal-day volatility. Year-end dates ($n = 33$) are the most extreme, with a mean of 0.0712% and standard deviation of 0.1696%. FOMC windows ($n = 152$) have a moderate mean (0.0116%) but the highest single-day maximum (2.95%), attributable to the September 2019 episode. Tax dates ($n = 64$) are not statistically distinguishable from normal days.

Table 2: Calendar effect summary statistics for the SOFR–EFFR spread (%).

Day Type	Count	Mean	Std	Min	Max
Normal Days	1,497	−0.0077	0.0401	−0.1500	0.3000
Month-Turn	183	0.0126	0.0637	−0.0600	0.3600
Quarter-Turn	85	0.0287	0.1069	−0.1200	0.6000
FOMC Window	152	0.0116	0.2453	−0.1400	2.9500
Year-End	33	0.0712	0.1696	−0.0300	0.7500
Tax Date	64	−0.0091	0.0545	−0.1400	0.1800

The ACF (Figure 2, top panel) decays geometrically with spikes at lags 21 and 63, corresponding to one trading month and one trading quarter. The PACF (bottom panel) cuts off after lag 2–3. These patterns motivate the SARIMA specification and the AR order $p = 3$ used in the SETAR and MS-AR models.

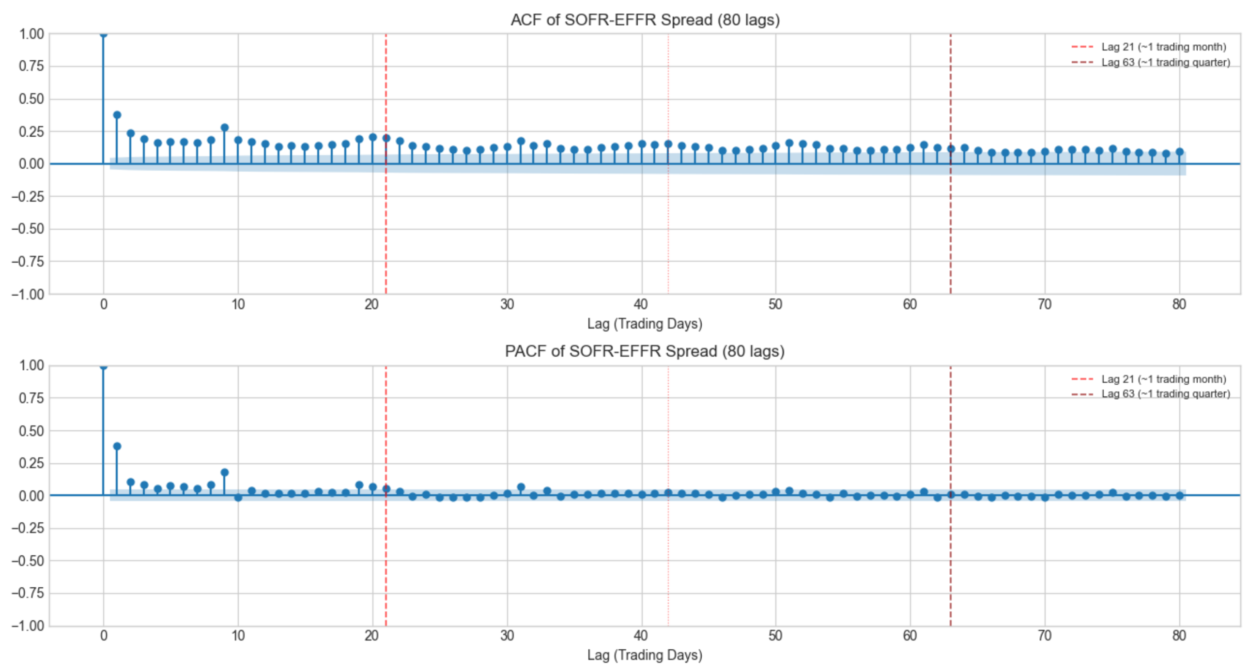

Figure 2: ACF (top) and PACF (bottom) of the SOFR–EFFR spread, 80 lags, 95% confidence bands. Red dashed lines at lags 21 and 63.

4.2 Results from SARIMA$(2, 0, 1) \times (0, 0, 0, 21)$

The grid search evaluates 99 candidate specifications (96 converge) and selects SARIMA$(2, 0, 1)\times$ $(0, 0, 0, 21)$ with AIC $= -4610.81$ and BIC $= -4588.43$. The optimal seasonal orders are $(P, D, Q) = (0, 0, 0)$: despite the visible seasonal spikes in the ACF, seasonal AR and MA terms do not improve the AIC. The model reduces to ARMA(2,1). Table 3 reports the top five specifications.

Table 3: Top five SARIMA specifications by AIC.

Rank	Order	AIC	BIC
1	$(2,0,1) \times (0,0,0,21)$	-4610.81	-4588.43
2	$(2,0,2) \times (0,0,0,21)$	-4605.64	-4577.67
3	$(3,0,2) \times (0,0,0,21)$	-4603.94	-4570.38
4	$(4,0,2) \times (0,0,0,21)$	-4602.89	-4563.75
5	$(3,0,3) \times (0,0,0,21)$	-4602.73	-4563.59

The estimated coefficients are $\hat{\phi}_1 = 1.2327$ (s.e. 0.015), $\hat{\phi}_2 = -0.2388$ (s.e. 0.010), $\hat{\theta}_1 = -0.9537$ (s.e. 0.015), and $\hat{\sigma}^2 = 0.0057$, all significant at the 1% level. The near-unit MA root ($|\hat{\theta}_1| \approx 0.95$) and the AR sum ($\hat{\phi}_1 + \hat{\phi}_2 \approx 0.994$) imply high persistence, consistent with the slow mean-reversion observed in the raw series.

Diagnostics. The Ljung–Box test rejects white noise at all lags ($Q(10) = 43.05$, $p < 0.0001$; $Q(20) = 53.87$, $p < 0.0001$; $Q(40) = 77.86$, $p = 0.0001$), indicating residual autocorrelation the linear model does not capture. The Jarque–Bera statistic is 9.67×10^7 ($p < 0.0001$; skewness $= 28.58$, kurtosis $= 1{,}082.3$). The coefficient estimates remain consistent under the QMLE property (Taniguchi and Kakizawa, 2000), but Gaussian-based confidence intervals are unreliable. Figure 3 displays the residual time series, histogram, and Q–Q plot, with severe tail departures from normality.

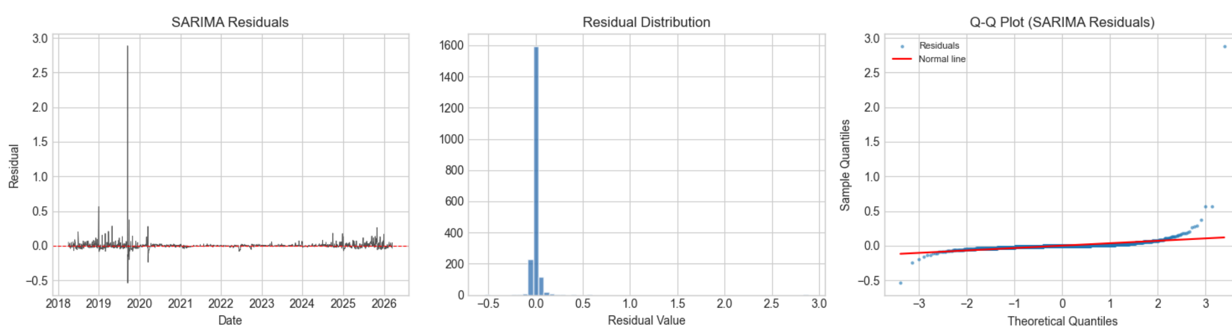

Figure 3: SARIMA(2,0,1) residual diagnostics. Left: residual time series. Center: histogram. Right: Q–Q plot against the standard normal.

The 12-day ahead forecast (Figure 4) converges to approximately 0.02% by day 3, with 95% confidence bands of $[-0.14\%, +0.18\%]$. These wide intervals reflect the high unconditional variance and underscore the limitations of the linear specification.

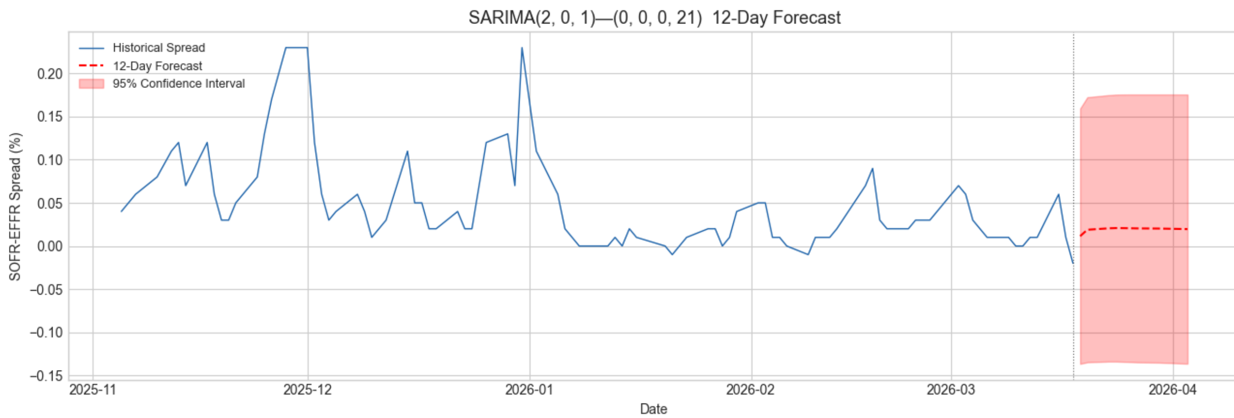

Figure 4: SARIMA(2,0,1) 12-day ahead forecast with 95% CI, against the last 90 trading days.

4.3 Results from SETAR + GARCH(1,1)

4.3.1 Threshold Non-Linearity Test

Tsay's F-test yields 248.47 ($p < 0.0001$, df $= (4, 1970)$), providing strong evidence for threshold non-linearity in the spread conditional on lagged reserve balances.

4.3.2 SETAR Estimation

Grid search identifies $d^* = 1$ and $c^* =$ 1,640,750 million (\$1.641 trillion), partitioning the sample into:

- **Regime 1—Abundant** ($R_{t-1} >$ \$1.641T): 1,754 observations (88.3%). Intercept $= -0.0006$; high AR persistence ($\hat{\phi}_1 = 0.895$, $\hat{\phi}_2 = -0.151$, $\hat{\phi}_3 = 0.148$).
- **Regime 2—Scarce** ($R_{t-1} \leq$ \$1.641T): 232 observations (11.7%). Intercept shifts to 0.0357; AR persistence collapses ($\hat{\phi}_1 = 0.156$, $\hat{\phi}_2 = 0.005$, $\hat{\phi}_3 = -0.016$), indicating faster mean-reversion with higher variance.

Table 4: SETAR regime-specific AR(3) coefficient estimates.

Parameter	Abundant ($R >$ \$1.641T)	Scarce ($R \leq$ \$1.641T)
Intercept	−0.0006	0.0357
AR(1)	0.8951	0.1562
AR(2)	−0.1506	0.0054
AR(3)	0.1481	−0.0158

When reserves exceed \$1.641 trillion, the repo market is well-supplied with liquidity and the spread follows a persistent near-zero process. Below the threshold, frictions in reserve distribution cause the spread to widen and become less forecastable from its own lags, consistent with the scarcity dynamics documented by Afonso et al. (2020). Figure 5 visualizes the regime partition over the sample.

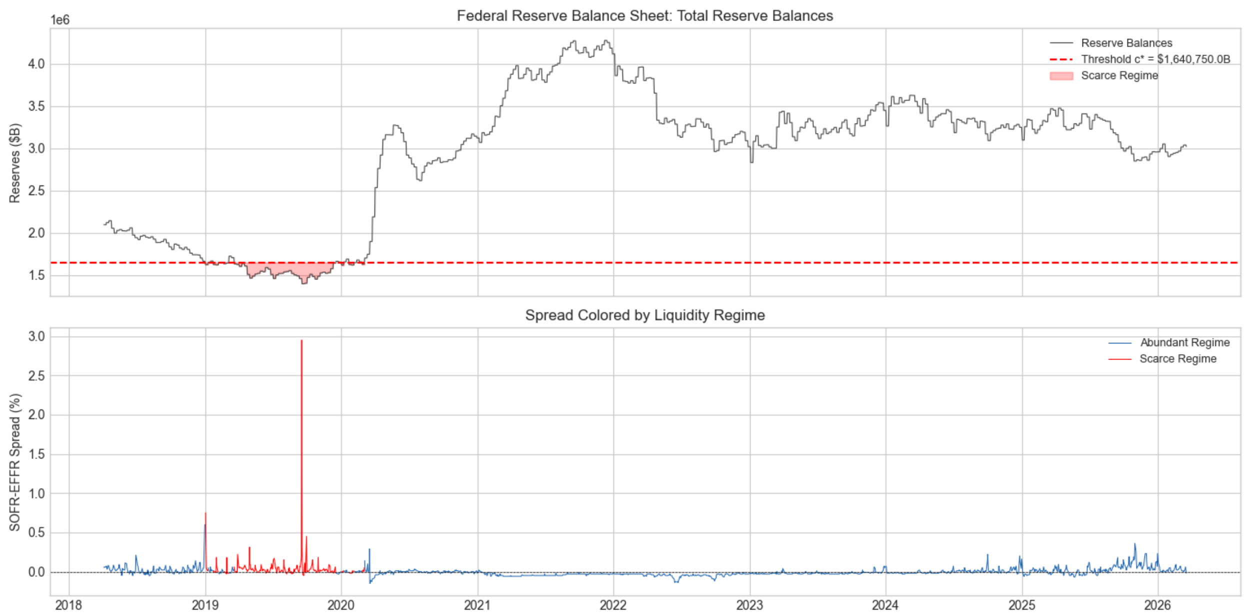

Figure 5: SETAR regime partition. Top: reserve balances with threshold $c^* = \$1.641\text{T}$ (red dashed); scarce periods shaded. Bottom: spread colored by regime.

4.3.3 GARCH(1,1) on SETAR Residuals

Engle's ARCH LM test on the SETAR residuals gives $\text{LM} = 0.902$ ($p = 0.9999$) and $F = 0.090$ ($p = 0.9999$): no significant conditional heteroscedasticity remains. The regime structure has already captured the primary source of time-varying variance.

A GARCH(1,1) with Student-t innovations is fitted for completeness. The estimates are $\hat{\omega} = 0.1697$ (s.e. 0.090, $p = 0.060$), $\hat{\alpha} = 0.3795$ (s.e. 0.091, $p < 0.001$), $\hat{\beta} = 0.6205$ (s.e. 0.111, $p < 0.001$), and $\hat{\nu} = 3.41$ (s.e. 0.174). The persistence $\hat{\alpha} + \hat{\beta} = 1.000$ is on the IGARCH boundary. The Student-t distribution is preferred over Normal by AIC (7589.39 vs. 7611.76). Figure 6 plots the residuals and conditional volatility.

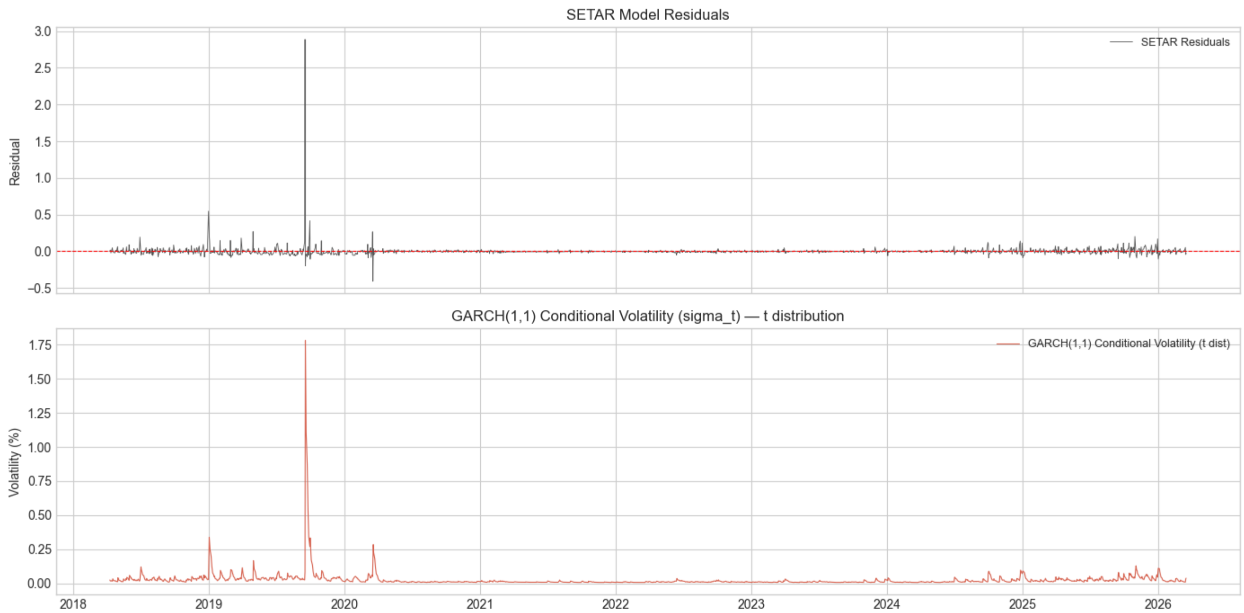

Figure 6: Top: SETAR residuals. Bottom: GARCH(1,1) conditional volatility under Student-t innovations.

4.4 Results from Markov-Switching AR

4.4.1 MS-AR Estimation

The MS-AR(3) model with 2 regimes (20 EM restarts) yields log-likelihood = 4,797.01, AIC = −9570.01, BIC = −9502.90. Table 5 reports the regime-specific coefficients. The scarce regime (regime 0) has variance $\hat{\sigma}_0^2 = 0.1247$, approximately 420× that of the abundant regime ($\hat{\sigma}_1^2 = 0.0003$).

Table 5: MS-AR(3) regime-specific parameter estimates.

Parameter	Abundant (regime 1)	Scarce (regime 0)
Intercept	−0.0095	0.1419
AR(1)	0.8205	−0.1989
AR(2)	−0.0413	−0.2363
AR(3)	0.0928	−0.2204
σ^2	0.0003	0.1247

The transition probabilities are $P(0 \to 0) = 0.7308$ and $P(1 \to 1) = 0.9871$, implying expected durations of 3.7 trading days (scarce) and 77.6 trading days (abundant). The scarce regime is transient but recurrent, consistent with short-lived funding pressures around reporting dates. Figure 7 displays the smoothed scarce-regime probability alongside the spread and reserve balances. The probability spikes align with periods when reserves approach the SETAR threshold.

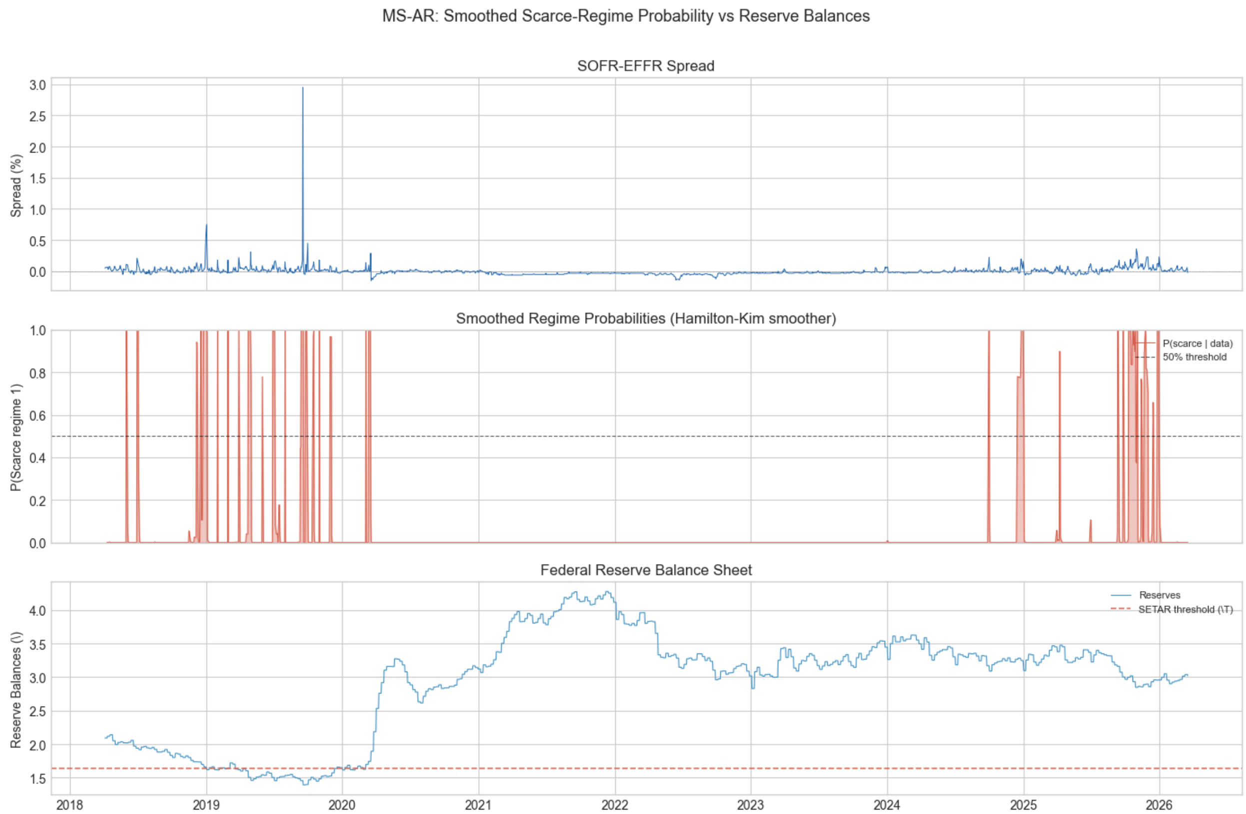

Figure 7: MS-AR regime analysis. Top: SOFR–EFFR spread. Middle: smoothed $P(\text{scarce} \mid \text{data})$ via the Hamilton–Kim smoother. Bottom: reserve balances with SETAR threshold for comparison.

The key distinction from SETAR is probabilistic regime assignment. SETAR assigns each observation to exactly one regime; MS-AR allows intermediate posterior probabilities, which is relevant during gradual reserve drawdowns where the transition between regimes is not instantaneous.

4.4.2 Time-Varying Transition Probabilities

The TVTP extension models transition probabilities as logistic functions of lagged reserves. The OOS RMSE is 0.0329%, compared to 0.0288% for constant-probability MS-AR. The Diebold–Mariano test yields DM $= -1.084$ ($p = 0.279$): the difference is not significant. The additional logistic parameters introduce estimation uncertainty that offsets the benefit of reserve-conditioned transitions over the 63-day evaluation window.

4.5 Results from the HAR Model

The HAR model is estimated on 1,965 observations with coefficients $\hat{\beta}_0 = -0.0006$, $\hat{\beta}_1 = 0.2584$ (daily), $\hat{\beta}_5 = 0.0021$ (weekly), and $\hat{\beta}_{21} = 0.5327$ (monthly), yielding $R^2 = 0.1957$. The sum of slope coefficients is 0.7931, indicating sub-unit total persistence. The monthly component dominates, contributing roughly twice the explanatory power of the daily lag, while the weekly term is negligible—suggesting that the 5-day horizon adds little beyond what the daily and monthly terms already capture. This weighting is consistent with the monthly balance-sheet cycle governing dealer repo activity (Munyan, 2015).

Figure 8 plots the fitted values against the realized spread (top) and decomposes fitted values into horizon contributions (bottom). The model tracks the spread adequately in normal conditions but cannot capture quarter-end and year-end spikes.

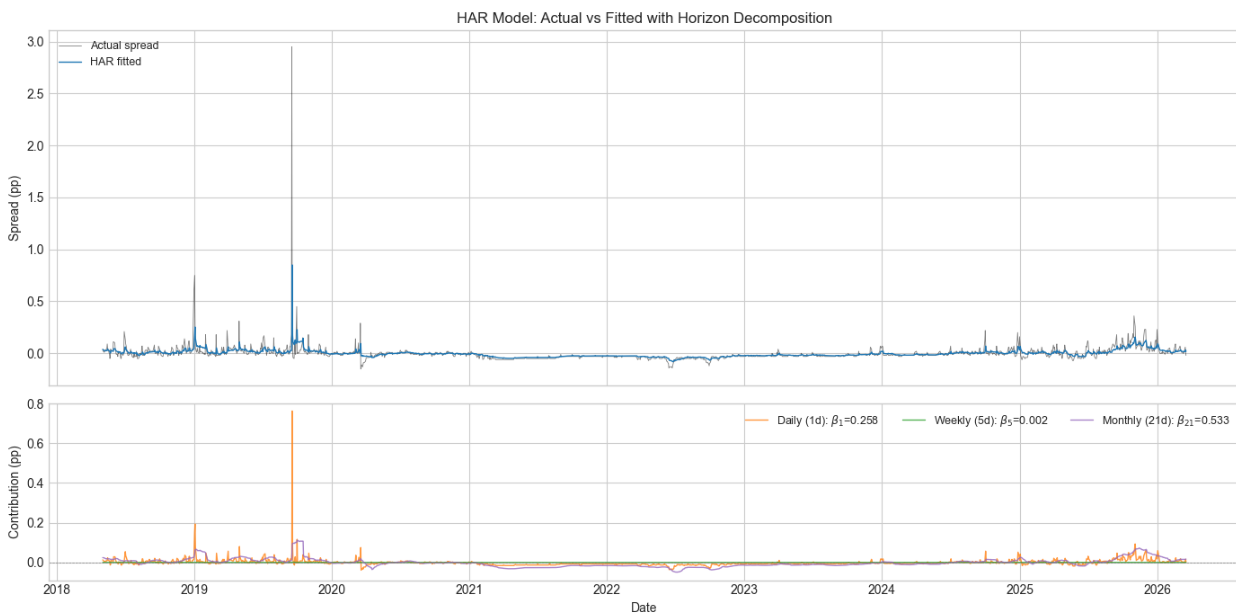

Figure 8: HAR model: actual versus fitted spread (top) and horizon decomposition into daily ($\hat{\beta}_1 = 0.258$), weekly ($\hat{\beta}_5 = 0.002$), and monthly ($\hat{\beta}_{21} = 0.533$) contributions (bottom).

The Ljung–Box test rejects white noise at all lags ($Q(10) = 40.73$, $p < 0.0001$; $Q(20) = 55.02$, $p < 0.0001$; $Q(40) = 72.35$, $p = 0.0013$), and the Jarque–Bera statistic is 9.43×10^7 ($p < 0.0001$; skewness $= 28.45$, kurtosis $= 1{,}074.5$), confirming the same heavy-tailed residual structure observed across all models. The out-of-sample RMSE of 0.0342% (Table 6) places HAR third among the five specifications—below both regime-switching models but above the SARIMA baseline.

4.6 Out-of-Sample Forecast Comparison

Table 6 reports rolling one-step-ahead RMSE on the 63-day hold-out. MS-AR achieves the lowest RMSE at 0.0288%, followed by MS-AR TVTP (0.0329%), HAR (0.0342%), SETAR (0.0347%), and SARIMA (0.0364%). Figure 9 overlays all five forecast paths against realized values; Figure 10 provides the bar chart comparison.

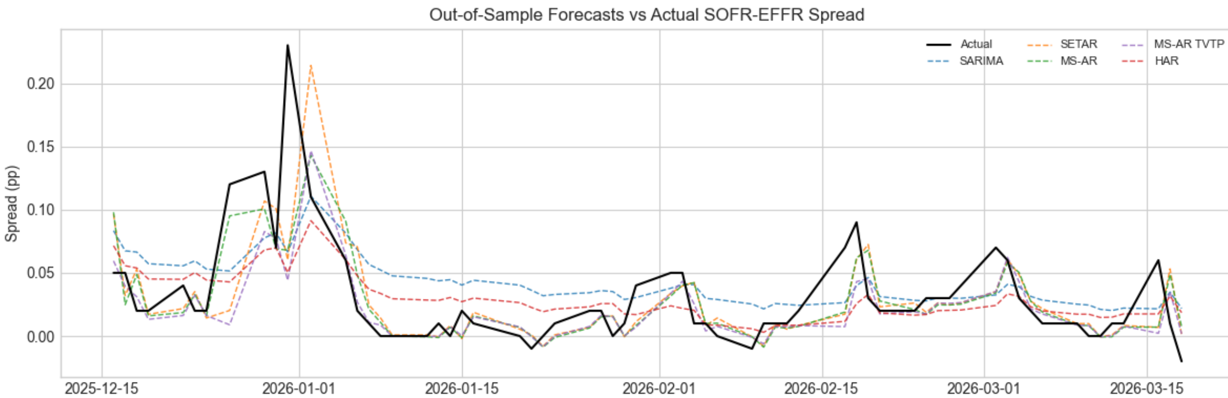

Figure 9: One-step-ahead out-of-sample forecasts from all five models versus the realized spread.

Table 6: Out-of-sample RMSE (63-day hold-out, rolling one-step-ahead).

Model	OOS RMSE (%)
MS-AR	0.0288
MS-AR TVTP	0.0329
HAR	0.0342
SETAR	0.0347
SARIMA	0.0364

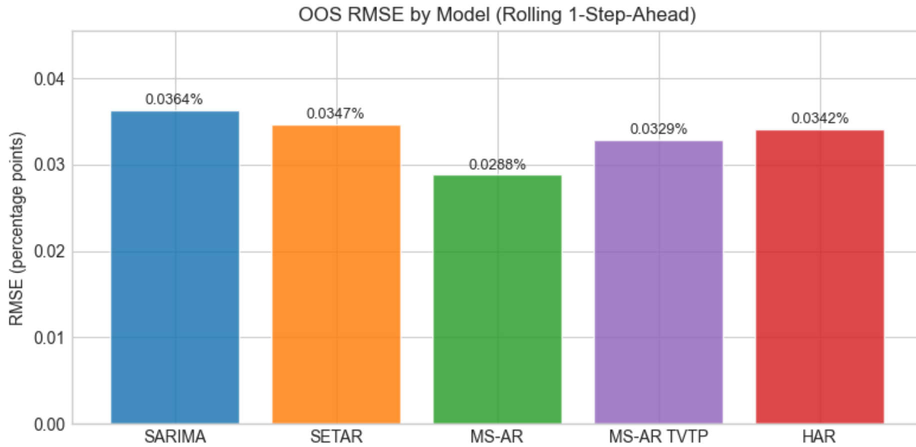

Figure 10: OOS RMSE by model. MS-AR: 0.0288%, a 20.9% improvement over SARIMA.

Table 7 reports pairwise Diebold–Mariano tests. SARIMA vs. MS-AR gives DM = +3.834 ($p < 0.001$): MS-AR is significantly more accurate at the 1% level. SETAR vs. MS-AR gives DM = +1.729 ($p = 0.084$): marginally significant evidence favoring probabilistic over threshold-based regime identification. All other pairs are insignificant at 10%.

Table 7: Diebold–Mariano pairwise test results (H_0: equal predictive accuracy).

Model 1	Model 2	DM stat	p-value
SARIMA	SETAR	+0.510	0.610
SARIMA	MS-AR	+3.834	$< 0.001^{***}$
SARIMA	MS-AR TVTP	+1.123	0.262
SARIMA	HAR	+1.249	0.212
SETAR	MS-AR	+1.729	0.084^{*}
SETAR	MS-AR TVTP	+0.644	0.520
MS-AR	MS-AR TVTP	−1.084	0.279
MS-AR TVTP	HAR	−0.632	0.528

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

4.7 Rolling Cross-Validation

To assess robustness, we conduct a 5-fold rolling cross-validation with non-overlapping 63-day test windows. Note that the MS-AR TVTP specification is omitted from this robustness check due to parameter convergence issues across shifting windows, as well as its underperformance in the primary hold-out test. Table 8 reports mean and standard deviation of RMSE across windows; Table 9 gives per-window results. MS-AR achieves the lowest mean RMSE

(0.0322%) and ranks first or second in every window. SETAR ranks second (0.0347%) with lower cross-window variance than SARIMA (0.0426%).

Table 8: Rolling cross-validation: mean ± std RMSE across 5 windows (%).

Model	Mean RMSE	Std RMSE	n
SARIMA	0.0426	0.0139	5
SETAR	0.0347	0.0116	5
MS-AR	0.0322	0.0123	5
HAR	0.0415	0.0131	5

Table 9: Per-window RMSE breakdown (%).

Window	Split Date	SARIMA	SETAR	MS-AR	HAR
1	2025-12-16	0.0364	0.0347	0.0288	0.0342
2	2025-09-15	0.0668	0.0534	0.0527	0.0641
3	2025-06-13	0.0358	0.0264	0.0234	0.0361
4	2025-03-14	0.0329	0.0239	0.0227	0.0316
5	2024-12-11	0.0410	0.0353	0.0335	0.0417

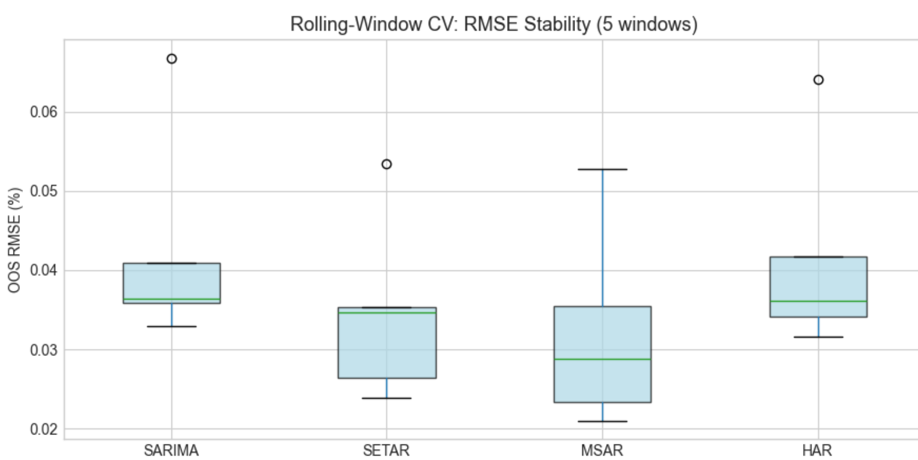

Figure 11: Rolling-window CV: RMSE distribution across 5 windows. MS-AR has the lowest median and tightest interquartile range.

Window 2 (split at 2025-09-15) produces elevated RMSE for all models, likely due to a quarter-end episode in the test period. The non-linear models (SETAR: 0.0534%, MS-AR: 0.0527%) still outperform the linear alternatives (SARIMA: 0.0668%, HAR: 0.0641%) by a substantial margin.

5 Conclusion and Future Study

We model the SOFR–EFFR spread with five specifications spanning linear, threshold, and regime-switching approaches. The results are as follows.

The spread is stationary but strongly non-linear. Tsay's test ($F = 248.47$, $p < 0.0001$) rejects linearity. The SETAR model identifies a reserve threshold near \$1.641 trillion separating an abundant regime (88.3% of observations, near-zero persistent spread) from a scarce regime (11.7%, elevated and rapidly mean-reverting spread), consistent with the reserve demand framework of Afonso et al. (2020) and Logan (2020).

Calendar effects are economically significant. Year-end dates produce a mean spread nearly $10\times$ that of normal days (0.0712% vs. -0.0077%), and quarter-turn dates exhibit $2.7\times$ the normal-day volatility. These patterns are attributable to GSIB surcharge and LCR reporting constraints (Munyan, 2015; Correa et al., 2021).

MS-AR achieves the best out-of-sample performance, with RMSE of 0.0288%—significantly lower than the SARIMA baseline of 0.0364% (DM = 3.834, $p < 0.001$). Rolling cross-validation confirms the ranking across all five windows. The advantage over SETAR is marginally significant ($p = 0.084$), suggesting that probabilistic regime identification provides incremental forecasting gains over a hard threshold.

The SETAR regime structure captures nearly all conditional heteroscedasticity in the data: the ARCH LM test on SETAR residuals is insignificant ($p \approx 1.0$). What appears as volatility clustering in the unconditional series is attributable to regime transitions rather than within-regime GARCH dynamics.

The TVTP extension does not significantly improve on constant-probability MS-AR ($p = 0.279$). The additional parameters in the logistic transition function introduce estimation noise that offsets the benefit of reserve-conditioned switching over the evaluation windows considered.

Future work. Incorporating calendar dummies into the regime-switching framework could reduce forecast errors around reporting dates. Joint SETAR–GARCH estimation would eliminate the generated-regressor bias of the current two-step procedure (Pagan, 1984). Higher-frequency intraday repo data could reveal dynamics not observable at the daily level. Expanding the threshold variable set to include TGA balances, SRF take-up, and ON RRP usage may sharpen regime identification, as these variables interact with aggregate reserve supply in determining funding conditions (Copeland et al., 2021).

References

Afonso, G., Cipriani, M., Copeland, A., Kovner, A., La Spada, G., and Martin, A. (2020). The market events of mid-September 2019. *Federal Reserve Bank of New York Staff Reports*, No. 918.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Anbil, S., Anderson, A., and Senyuz, Z. (2020). What happened in money markets in September 2019? *FEDS Notes*, Board of Governors of the Federal Reserve System.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Copeland, A., Duffie, D., and Yang, Y. (2021). Reserves were not so ample after all. *NBER Working Paper*, No. 29090.

Correa, R., Du, W., and Liao, G. (2021). U.S. banks and global liquidity. *NBER Working Paper*, No. 27491.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Federal Reserve Bank of New York. (2018). Secured Overnight Financing Rate data. `https://www.newyorkfed.org/markets/reference-rates/sofr`.

Board of Governors of the Federal Reserve System. (2021). Selected interest rates (daily)—H.15. `https://www.federalreserve.gov/releases/h15/`.

Board of Governors of the Federal Reserve System. (2026). Factors affecting reserve balances—H.4.1. `https://www.federalreserve.gov/releases/h41/`.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1–2):1–22.

Logan, L. K. (2020). Treasury market liquidity and the Federal Reserve during the COVID-19 pandemic. *Remarks at SIFMA webinar*, Federal Reserve Bank of New York.

Munyan, B. (2015). Regulatory arbitrage in repo markets. *Office of Financial Research Working Paper*, No. 15-22.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 25(1):221–247.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer, New York.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Springer, New York.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84(405):231–240.

Federal Reserve Bank of St. Louis. (2026). Federal Reserve Economic Data (FRED). `https://fred.stlouisfed.org`.