risks

MDPI

Article

Optimized Moving Average Smoothing for Volatility Forecasting in Futures Markets

Pijus Zlatkus [1], Aistis Raudys [2,*], Linas Lazaravičius [1], Linas Žvirblis [1], Tomas Plankis [2], Vytautas Valaitis [2], Julius Andrikonis [2] and Rimantas Vaicekauskas [2]

[1] Independent Researcher, LT-08303 Vilnius, Lithuania
[2] Institute of Computer Science, Faculty of Mathematics and Informatics, Vilnius University, Didlaukio g. 47, LT-08303 Vilnius, Lithuania
* Correspondence: aistis.raudys@mif.vu.lt

Abstract

Common moving averages (MAs) used for volatility forecasting rely on fixed heuristic weights. We propose a Custom Moving Average (CMA) whose weights are learned to forecast the next-day True Range (TR). Using daily OHLC data for 55 futures contracts with a chronological 60/20/20 train/validation/test split, CMA weights are optimized for each look-back period p by projected subgradient descent against Mean Absolute Error (MAE), with validation-based early stopping. We benchmark CMA against nine standard MAs over 62 look-back periods, selecting the period on validation and reporting accuracy on the held-out test set. Learned weights concentrate on an effective horizon of 15–20 observations regardless of p: about 85% of the mass sits in lags 0–14 for both $p = 36$ and $p = 120$. EMA is the validation-best method at very short windows; CMA dominates from $p \approx 16$ onwards, winning on 53 of 55 instruments by $p = 120$. Under a single-global-period rule, CMA at $p^* = 36$ attains the lowest test-set geometric-mean MAE, with EMA at $p^* = 7$ the closest competitor (0.87% higher); per-instrument validation selection does not overturn the ranking, with EMA again closest at 0.53%. Higher-order smoothers (T3, TEMA, DEMA) do not improve on CMA under either rule; CMA's advantage is robust to the choice of selection granularity.

Keywords: volatility forecasting; true range; moving averages; learned weighting; custom moving average; projected subgradient descent; mean absolute error; cross-asset evaluation

JEL Classification: C53; C58; G17

Check for updates

Academic Editor: Young Shin Aaron Kim

Received: 21 May 2026
Revised: 25 June 2026
Accepted: 2 July 2026
Published: 6 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Volatility, which represents the magnitude and frequency of price fluctuations, is a central concept in financial theory and practice (Poon and Granger 2003). It is not just a measure of dispersion, but also a proxy for market uncertainty, risk, and information flow (Chua and Tsiaplias 2019). Producing accurate volatility forecasts is essential for a wide spectrum of financial activities. For example, in risk management, volatility estimates are critical inputs for calculating Value at Risk (VaR) and Expected Shortfall (ES), which are regulatory and internal benchmarks for market risk exposure (Brooks and Persand 2003; Yamai and Yoshiba 2005). Portfolio managers use volatility forecasts for strategic and tactical asset allocation to optimize risk–return profiles (Brière et al. 2010). Furthermore, traders, particularly those employing algorithmic and quantitative strategies, depend on precise

short-term volatility predictions to set profit targets and stop-loss levels, adjust position sizes, and develop strategies that explicitly trade volatility, such as straddles, strangles, or variance swaps (Joiner et al. 2022). Given this extensive utility, even marginal improvements in the accuracy of volatility forecasts can yield substantial economic advantages and contribute to greater market stability and efficiency.

The broader volatility-forecasting literature includes several well-established modeling traditions. In econometrics, ARCH and GARCH models remain foundational because they model conditional variance directly through past shocks and past volatility (Bollerslev 1986; Engle 1982). A related realized-volatility literature uses higher-frequency return information to construct observed volatility measures; HAR-type models, for example, represent persistence through daily, weekly, and monthly volatility components (Corsi 2009). Recent work confirms that this area remains active: modern realized-volatility forecasting has been surveyed through the machine-learning and deep-learning period (Leushuis and Petkov 2026), futures-market volatility has been forecast using machine-learning and deep-learning models (Luo et al. 2024; Yang et al. 2024), and HAR-type realized-volatility models have been extended with graph neural networks (Hu et al. 2025). These approaches are important reference points, but the present study addresses a narrower and practitioner-oriented question within technical analysis: conditional on using daily OHLC-based moving-average forecasts of True Range, can the weighting rule be learned from data rather than fixed heuristically? Accordingly, CMA is evaluated against alternative moving-average specifications rather than against structurally different volatility-model families.

1.1. Moving Averages: Utility and Inherent Limitations in Volatility Analysis

Moving averages (MAs) are among the oldest and most widely adopted tools in technical analysis, valued for their simplicity of calculation and interpretation. Their primary function is to smooth time-series data, thereby filtering out high-frequency noise and making the underlying trends or cyclical patterns more discernible. Despite their widespread use, conventional MAs face significant challenges when applied to forecasting financial volatility (Poon and Granger 2003). The lagging nature of MAs, a direct consequence of their reliance on historical data, can impede the timely detection of shifts in volatility regimes (Marcucci 2005). This is particularly critical, as volatility is known to exhibit clustering, where periods of high turmoil are followed by further turbulence, and tranquil periods tend to persist (Cont 2007). A lagging indicator may fail to capture the onset of such clusters promptly. For instance, the equal weighting of an SMA may underreact to recent shocks, while the fixed decay factor of an EMA might not be appropriate across all market conditions or for all assets (Griffioen 2003). This structural inflexibility can lead to suboptimal forecast accuracy. While adaptive MAs such as KAMA offer some improvement by adjusting responsiveness, their adaptation rules are themselves heuristic and may not be universally optimal.

1.2. Research Motivation, Objectives, and Contributions

Recognizing the limitations of existing moving averages for volatility forecasting, this research is motivated by the hypothesis that a moving average with empirically optimized, data-driven weights can offer better predictive performance. The primary objective of this study is to adapt and evaluate a Custom Moving Average (CMA), with weights specifically optimized to predict the next day's True Range (TR), a comprehensive measure of daily price volatility. The optimization is performed using projected batch subgradient descent to minimize the Mean Absolute Error (MAE) between the CMA's one-step-ahead TR forecast and the subsequently realized actual TR. In our implementation, for each defined look-back

window, we learn a single set of global weights from all instruments' training data and apply them to each instrument rather than imposing a fixed analytic weighting formula.

This study extends two related lines of previous research. First, optimized custom moving-average weights have been proposed for financial price-series smoothing, where the goal is to improve the smoothness–accuracy trade-off relative to fixed-form moving averages (Raudys and Pabarškaitė 2018). Second, recent work has examined transformed financial price and volatility-risk series using recurrent neural networks and XGBoost, including experiments on True Range volatility forecasting for futures data (Raudys and Goldstein 2022). We connect these directions by learning moving-average weights directly for one-step-ahead TR forecasting: the proposed model retains the transparent form of a moving average, but its weights are trained for the volatility-prediction objective.

The main contributions of this paper are multifaceted:

- Methodological Innovation: This study extends the optimized CMA framework to volatility (True Range) forecasting, using a flexible, learned weighting scheme instead of predefined or simple adaptive rules.
- Comprehensive Empirical Evaluation: The CMA's performance is benchmarked against a diverse set of nine established MAs (SMA, EMA, WMA, SWMA, DEMA, TEMA, T3, KAMA, and ZMA) across 55 financial instruments and 62 distinct daily look-back periods.
- Deployable, Validation-Based Selection: All look-back periods are selected on the validation segment of the chronological 60/20/20 split, with final accuracy reported on the test segment. The study evaluates CMA against benchmarks at two selection granularities—a single global period per method, and a per-instrument period chosen by validation—so that the reported numbers correspond to operational rules rather than to oracle upper bounds.

2. Materials and Methods

This section details the financial data used, the computation of True Range as the target volatility measure, the suite of benchmark moving averages used for comparison, the formulation of the proposed Custom Moving Average (CMA), the projected batch subgradient descent algorithm for optimizing CMA weights, and the Mean Absolute Error (MAE) metric for performance assessment.

2.1. Data Specification and Pre-Processing

The empirical analysis used daily Open, High, Low, and Close (OHLC) data for 55 distinct continuous-contract futures series spanning 3 January 2000, to 7 April 2025. The universe comprised nine equity-index futures, nine currency futures, nine fixed-income futures, seven energy futures, five metal futures, six grain and oilseed futures, and ten softs/livestock and minor-commodity futures traded across multiple major exchanges. The set was constructed to span the actively traded segments of each asset class, providing broad cross-sectional coverage of the global futures market; the complete list of instrument identifiers is available from the corresponding author on request (see the Data Availability Statement). For each instrument, we derived the True Range (TR) series from $\{H_t, L_t, C_{t-1}\}$ (i.e., High, Low, and previous close).

For CMA optimization, the True Range values from all instruments in the training segment were pooled and standardized using a single scaler fitted only on that pooled training segment; the same transformation was then applied to each instrument's validation and test segments. The OHLCV data were obtained from TradeStation as continuous futures series and read from the vendor's native binary format. TradeStation links individual contracts at rollover points and applies back-adjustment to improve price continuity. For the standard

continuous symbols, rollovers follow TradeStation's market-specific continuous-contract rules, which are generally based on a modified open-interest interpretation. Custom continuous symbols were used only where TradeStation did not provide a standard continuous contract for the required market, allowing all instruments in the sample to be represented as adjusted continuous daily series; these custom symbols used explicit rollover parameters and constant back-adjustment. True Range was computed from the adjusted continuous OHLC series, using the previous adjusted close from the same series. We treated the vendor's continuous-contract series as the underlying price input and did not re-construct rollovers. The start date varied by instrument because individual contracts had different listing histories, while all series ended on 7 April 2025. For each instrument, we used a chronological 60/20/20 split: the first 60% of the TR series formed the training segment, the next 20% the validation segment, and the final 20% the held-out test segment. The validation segment was used exclusively for early stopping and model selection during CMA weight optimization, while all final out-of-sample metrics were reported on the test segment.

Because residual roll-related discontinuities are possible in continuous futures data even after vendor adjustment, we performed a gap-driven TR diagnostic. We flagged observations satisfying three conditions: TR was in the top 1% for that instrument, TR was dominated by the previous-close gap component rather than the intraday high–low range, and the excess gap component exceeded five times the instrument's median TR. This rule identified 48 observations out of 299,121 daily observations (0.016%). As an aggregation sensitivity check using the already selected models, excluding the instrument with the highest concentration of flagged observations did not change the method ranking.

2.2. Volatility: True Range

The chosen measure of daily price volatility in this study is the True Range (TR), developed by Wilder (1978). TR is favored for its comprehensiveness, as it accounts not only for the intraday price movement but also for price gaps that can occur between the previous day's close and the current day's open. This makes it a more robust volatility measure than the simple range or standard deviation of closing prices, especially in markets prone to overnight or weekend gaps. For any given trading day t, with High H_t, Low L_t, and the previous day's Close C_{t-1}, the True Range is defined as:

$$\mathrm{TR}_t = \max[(H_t - L_t), |H_t - C_{t-1}|, |L_t - C_{t-1}|]$$

The objective of the moving averages in this study is to forecast the next day's True Range, TR_{t+1}, using a series of past True Range values $\mathrm{TR}_t, \mathrm{TR}_{t-1}, \ldots, \mathrm{TR}_{t-p+1}$, where p is the look-back period of the moving average.

2.3. Benchmark Moving Averages

The predictive performance of the proposed CMA is benchmarked against nine widely recognized moving averages. Each MA is applied to the historical True Range series to generate forecasts. For an MA of period p, the forecast for TR_{t+1} is generated using data up to time t, denoted as $\mathrm{MA}_p(X)_t$.

1. **Simple Moving Average (SMA)**: The arithmetic average of the last p observations of X. It assigns equal weight $1/p$ to each observation.

$$\mathrm{SMA}_p(X)_t = \frac{1}{p}\sum_{i=0}^{p-1} X_{t-i}$$

Its simplicity is its main appeal, but it is often criticized for its slow response to recent changes and for the "drop-off" effect, whereby the SMA can shift abruptly when an old observation rolls out of the window even if no notable new data has arrived.

2. **Exponential Moving Average (EMA)**: Assigns exponentially decreasing weights to past observations of X, making it more responsive to recent data than the SMA.

$$\mathrm{EMA}_p(X)_t = \alpha X_t + (1-\alpha)\,\mathrm{EMA}_p(X)_{t-1}$$

where $\alpha = \dfrac{2}{p+1}$.

3. **Double Exponential Moving Average (DEMA)**: Introduced by Mulloy (1994), DEMA attempts to reduce lag by applying an EMA to another EMA and adjusting the result.

$$\mathrm{DEMA}_p(X) = 2 \times \mathrm{EMA}_p(X) - \mathrm{EMA}_p(\mathrm{EMA}_p(X))$$

4. **Triple Exponential Moving Average (TEMA)**: Introduced in the same article as DEMA (Mulloy 1994), this MA involves a three-fold application of exponential smoothing, designed for even greater responsiveness.

$$\mathrm{EMA}_p^2(X) = \mathrm{EMA}_p(\mathrm{EMA}_p(X))$$

$$\mathrm{EMA}_p^3(X) = \mathrm{EMA}_p(\mathrm{EMA}_p^2(X))$$

$$\mathrm{TEMA}_p(X) = 3 \times \mathrm{EMA}_p(X) - 3 \times \mathrm{EMA}_p^2(X) + \mathrm{EMA}_p^3(X)$$

5. **Tillson's Triple Exponential Moving Average (T3)**: A low-lag, high-smoothness filter introduced by Tillson (1998). T3 applies a generalized DEMA (GD) three times, controlled by a volume factor $v \in [0,1]$. Larger v increases responsiveness while keeping the curve smooth compared with EMA/DEMA/TEMA. Empirically, $v \approx 0.7$ is a common default in practice (Tillson 1998).

$$\mathrm{GD}_p(X)_t = (1+v)\,\mathrm{EMA}_p(X)_t - v\,\mathrm{EMA}_p\left(\mathrm{EMA}_p(X)_t\right)_t$$

$$\mathrm{T3}_p(X)_t = \mathrm{GD}_p(\mathrm{GD}_p(\mathrm{GD}_p(X)))_t$$

6. **Weighted Moving Average (WMA)**: Assigns linearly increasing weights to past observations of X, with the most recent value receiving the highest weight. For a period p, the implementation uses weights $1, 2, \ldots, p$ (oldest to newest data point in the window).

$$\mathrm{WMA}_p(X)_t = \frac{2}{p(p+1)} \sum_{k=1}^{p} k\, x_{t-p+k}$$

7. **Sinusoidal Weighted Moving Average (SWMA)**: Assigns weights based on a sine function. It uses weights $w_j = \sin(\pi \times j/(p+1))$ for $j = 1, \ldots, p$ applied to the window of p values.

$$\mathrm{SWMA}_p(X)_t = \frac{\sum_{j=1}^{p} \mathrm{X}_{t-p+j} \times \sin(\frac{\pi j}{p+1})}{\sum_{k=1}^{p} \sin(\frac{\pi k}{p+1})}$$

8. **Kaufman's Adaptive Moving Average (KAMA)**: Developed by Kaufman (2013), KAMA adjusts its smoothing period based on market volatility, measured by an "Efficiency Ratio" (ER). Given X_t and a window p, define

$$D_t = \sum_{k=0}^{p-1} \left| X_{t-k} - X_{t-k-1} \right| \qquad N_t = \left| X_t - X_{t-p} \right|$$

The efficiency ratio is clamped to $[0, 1]$:

$$\mathrm{ER}_t = \begin{cases} 1 & D_t \le N_t \text{ or } D_t = 0 \\ \dfrac{N_t}{D_t} & \text{otherwise} \end{cases}$$

Let

$$\alpha_f = \frac{2}{f+1} \qquad \alpha_s = \frac{2}{s+1} \qquad \alpha_t = \left(\mathrm{ER}_t(\alpha_f - \alpha_s) + \alpha_s \right)^2$$

KAMA is updated recursively by

$$\mathrm{KAMA}_p(X)_t = \mathrm{KAMA}_p(X)_{t-1} + \alpha_t \left(X_t - \mathrm{KAMA}_p(X)_{t-1} \right)$$

with initialization $\mathrm{KAMA}_p(X)_{t_0} = X_{t_0-1}$ at the first computable time. The fast and slow EMA periods follow Kaufman's defaults of $f = 2$ and $s = 30$, giving $\alpha_f = \frac{2}{3} \approx 0.66667$ and $\alpha_s = \frac{2}{31} \approx 0.06452$.

9. **Zero-Lag Moving Average (ZMA)**: Conceptualized by Ehlers (2001), ZMA aims to eliminate lag. The implementation uses $\alpha = 2/(p+1)$.

$$\mathrm{ZMA}_p(X)_t = \alpha \cdot (\mathrm{X}_{t-1} + \beta \cdot (\mathrm{X}_t - \mathrm{ZMA}_p(X_{t-1})_{t-1})) + (1-\alpha) \cdot \mathrm{ZMA}_p(X_{t-1})_{t-1}$$

where β is a parameter set to 0.2.

For convenience, Table 1 summarizes all nine benchmark filters together with the proposed CMA.

Table 1. Summary of moving-average filters compared in this study.

Acronym	Name	Key Parameter	Reference
SMA	Simple MA	period p	—
EMA	Exponential MA	$\alpha = 2/(p+1)$	—
DEMA	Double Exponential MA	period p	Mulloy (1994)
TEMA	Triple Exponential MA	period p	Mulloy (1994)
T3	Tillson's Triple EMA	p, volume factor $v = 0.7$	Tillson (1998)
WMA	Weighted MA	linear weights $1, \ldots, p$	—
SWMA	Sinusoidal-Weighted MA	sine weights	—
KAMA	Kaufman's Adaptive MA	p, $f = 2$, $s = 30$	Kaufman (2013)
ZMA	Zero-Lag MA	$\alpha = 2/(p+1)$, $\beta = 0.2$	Ehlers (2001)
CMA	Custom MA (this work)	learned weights $w_0, \ldots, w_{p-1}$	this paper

2.4. The Proposed Custom Moving Average (CMA)

The Custom Moving Average (CMA) proposed in this study is a data-driven, unit-sum linear moving-average filter with empirically optimized weights. For a calculation period p, the CMA forecast for X_{t+1} based on information up to time t (using weights w_j for X_{t-j}) is:

$$\mathrm{CMA}_p(\mathrm{w})_t = \sum_{j=0}^{p-1} w_j \cdot \mathrm{X}_{t-j}$$

where $\mathrm{w} = (w_0, w_1, \ldots, w_{p-1})$ is the vector of p weights. The weights are constrained:

$$\sum_{j=0}^{p-1} w_j = 1$$

This constraint enforces unit-sum weights; after each gradient step the weight vector is projected back onto the affine hyperplane that satisfies the constraint. Non-negativity is not imposed, so the optimization permits signed coefficients. Thus, CMA is a unit-sum linear moving-average filter, but not necessarily a convex weighted average whose output must lie within the range of the observations in the window. Unlike the nine fixed-form benchmarks summarized in Table 1, whose only free parameter is the look-back period p, the CMA weight vector w is the only quantity in this study that is fit to data.

2.5. CMA Weight Optimization via Projected Batch Subgradient Descent

The weight vector w is optimized by minimizing the Mean Absolute Error (MAE) on the training dataset. Because MAE is convex but non-smooth, we use projected batch subgradient descent with a sum-to-one projection after each update. The loss function $L(\mathrm{w})$ is:

$$L(\mathrm{w}) = \frac{1}{N_{\text{train}} - p} \sum_{t=p+1}^{N_{\text{train}}} |\mathrm{TR}_{\text{actual},t} - \mathrm{CMA}_p(\mathrm{w})_{t-1}|$$

where N_{train} is the number of observations in the training set, $\mathrm{TR}_{\text{actual},t}$ is the actual True Range at time t, and $\mathrm{CMA}_p(\mathrm{w})_{t-1}$ is the forecast made at $t-1$ for time t. The sum starts at $t = p+1$ because the CMA forecast requires p prior observations and is therefore undefined for the first p time indices. MAE is chosen because daily TR is heavy-tailed and contains episodic volatility spikes. Unlike MSE, the MAE subgradient depends on the sign rather than the magnitude of the forecast residual, limiting the extent to which an extreme forecast error can dominate a weight update while retaining a convex objective for the linear CMA filter.

The optimization uses batch gradient descent. Key steps in implementation are:

1. **Initialization:** Weights w_j are initialized uniformly as $w_j = 1/p$.
2. **Subgradient Computation:** Using the feature matrix X (lagged TR values) and target y (actual TR), the MAE subgradient is $X^\top \cdot \mathrm{sgn}(\text{predictions} - y)/N_{\text{train}}$, where $\mathrm{sgn}(0) = 0$.
3. **Weight Update:** Weights are updated using a learning rate η:

$$w_{j,\text{new}} = w_j - \eta \cdot g_j$$

   where g_j denotes the j-th component of the batch subgradient of the MAE loss with respect to the weight vector **w**.
4. **Projection onto Feasible Set:** After each update, weights are projected to satisfy the sum-to-one constraint. For any updated vector v, the affine projection is

$$\Pi(v) = v - \frac{\sum_{j=0}^{p-1} v_j - 1}{p} \mathbf{1}_p$$

   where $\mathbf{1}_p$ denotes the p-dimensional vector of ones.
5. **Early stopping and model selection:** Steps 2–4 are repeated for a predefined number of epochs while monitoring validation MAE. The final weights are selected from the epoch with the lowest validation MAE (with patience-based early stopping).

This optimization procedure was performed to learn a single set of global weights for each period p using scaled True Range data. These learned global weights are then evaluated on the test set for each instrument.

Hyperparameters. The optimizer uses a constant learning rate $\eta = 0.01$, a maximum of 3000 epochs, and an early-stopping patience of 10 epochs measured on validation MAE; the weights returned are those of the best-validation epoch. The weight vector is initialized uniformly at $w_j = 1/p$. Within each epoch, the optimizer cycles through the 55 per-instrument training matrices in sequence; for each instrument, one batch subgradient step is taken over that instrument's full lagged-feature matrix, followed by the sum-to-one projection (yielding 55 weight updates per epoch). After training, weights with absolute value below 0.005 are set to zero and the resulting vector is re-projected onto the sum-to-one hyperplane. Because this projection applies a common correction to all components to restore the unit sum, coefficients set to zero can become very small non-zero values. The thresholding should therefore be understood as numerical shrinkage rather than strict sparsification. This procedure accounts for the near-zero long-lag tail observed in the optimized CMA weights reported in the Section 3. Optimization across the 62 look-back periods is parallelized via a Python 3.11.7 multiprocessing pool, with one process per period.

Targeted optimizer sensitivity. As a targeted sensitivity check, rather than a full re-estimation of all 62 look-back periods, we re-estimated CMA weights for representative short, selected, and long look-backs p = 736,120. We varied the learning rate $(0.005, 0.01, 0.02)$, initialization method (uniform, linearly recent-weighted, EMA-like, and random simplex), and early-stopping patience (10 and 25 epochs without validation improvement). The resulting validation and test MAE values were materially stable. For the selected CMA period $p = 36$, test MAE varied by less than 0.1% across these variants, suggesting that the reported results are not materially driven by a particular starting point or hyperparameter value.

2.6. Performance Evaluation Metrics

Mean Absolute Error (MAE) is the primary metric and remains the criterion used for CMA optimization, validation-based early stopping, and selection of the look-back period for every method. Forecast accuracy measures emphasize different properties of the forecast-error distribution (Hyndman and Koehler 2006), and volatility-forecast comparisons are known to be sensitive to the chosen loss function and proxy construction (Patton 2011). Primary out-of-sample comparisons are reported as MAE on the held-out test segment:

$$\mathrm{MAE} = \frac{1}{N_{\text{test}}} \sum_{t=1}^{N_{\text{test}}} \left| \mathrm{TR}_{\text{actual},t} - \mathrm{Forecast}_{t-1} \right|$$

where N_{test} is the number of valid forecasts in the test set, and $\mathrm{Forecast}_{t-1}$ is the MA's one-step-ahead prediction for $\mathrm{TR}_{\text{actual},t}$. A lower MAE signifies a more accurate model.

To assess whether the ranking depends on the error measure, we additionally score the unchanged, validation-MAE-selected test forecasts using Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) on the original, unstandardized TR scale:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\text{test}}} \sum_{t=1}^{N_{\text{test}}} \left(\mathrm{TR}_{\text{actual},t} - \mathrm{Forecast}_{t-1} \right)^2}$$

$$\mathrm{MAPE} = \frac{100}{N_+} \sum_{t:\, \mathrm{TR}_{\text{actual},t} > 0} \left| \frac{\mathrm{TR}_{\text{actual},t} - \mathrm{Forecast}_{t-1}}{\mathrm{TR}_{\text{actual},t}} \right|$$

Here, N_+ is the number of valid test observations with positive realized TR. No zero realized-TR observations occurred among the valid forecast–actual pairs used in the selected test comparisons. The trained weights and selected periods are not changed for these robustness checks: all p^* values remain fixed from validation MAE, thereby avoiding metric-specific retuning after inspecting test performance. As in the primary analysis, cross-instrument comparisons are based on the geometric mean of per-instrument error ratios.

3. Results

This section presents the empirical findings from a comparative analysis of the Custom Moving Average (CMA) against a suite of benchmark moving averages in predicting daily True Range. The performance of all indicators was evaluated out of sample across 55 financial instruments and 62 calculation periods drawn from five ranges with specified step sizes: 2–10 (step by 1), 12–30 (step by 2), 33–60 (step by 3), 65–150 (step by 5), and 160–300 (step by 10).

3.1. Cross-Instrument Predictive Performance

We compare ten moving averages including CMA at two selection granularities: a single global period per method (cross-instrument selection, Table 2), and a per-instrument period selected separately for each contract (per-instrument selection, Table 3). At both granularities the look-back period p^* is chosen on the validation segment of the chronological 60/20/20 split, and final accuracy is reported on the held-out test segment. The cross-instrument aggregator throughout this section is the geometric mean of per-instrument MAE, both for selecting p^* and for comparing methods, so that the period that determines selection is the same one that governs the headline ranking.

Cross-instrument selection. For each method we first compute the validation-period minimum of $\widehat{G}_p^{\mathrm{val}}$, where $\widehat{G}_p^{\mathrm{val}}$ is the cross-instrument geometric mean of validation-set MAE at look-back p. To avoid selecting longer horizons that produce only negligible validation gains, we report the smallest period whose $\widehat{G}_p^{\mathrm{val}}$ is within 0.25% of this minimum; the selected period is then applied to every instrument and accuracy is measured on the test set. Table 2 reports the resulting cross-instrument geometric mean of test-set MAE ratios relative to CMA at its own p^*. The 95% confidence intervals are constructed on the log ratios across the 55 instruments and exponentiated back to the ratio scale. The win rate is the fraction of instruments for which a method has lower test MAE than CMA at CMA's p^*. Regret is computed within each method as the cross-instrument geometric-mean test MAE under the global p^* divided by the same method's per-instrument geometric-mean test MAE under the per-instrument rule of Table 3; values close to 1 indicate that one global period transfers cleanly across the cross-section.

Under cross-instrument selection, CMA at $p^* = 36$ attains the lowest test-set geometric-mean MAE among the ten methods. The closest non-CMA configuration is EMA at $p^* = 7$, with a ratio of 1.0087 (i.e., 0.87% higher than CMA on the geometric-mean MAE scale); its confidence interval lies strictly above 1, so although the gap is small it is statistically distinguishable from parity across the 55 instruments. All other methods are at least 1.9% above CMA. The regret column is uniformly within 0.6% of unity for every method, indicating that a single global period transfers well across the cross-section under validation-based selection—no method recovers a meaningful additional improvement by tuning p per instrument. CMA's validation-MAE curve is essentially flat for all $p \gtrsim 30$; the strict argmin happens to lie at the grid boundary ($p = 300$), but we report $p^* = 36$ under the 0.25% parsimony tie-break. Across the range $p \in [30, 300]$, CMA's test MAE deviates from its value at $p^* = 36$ by less than about 0.7%. We return to this behavior in Section 3.3;

in practice, CMA can be deployed at any sufficiently large look-back without re-tuning, because the optimizer adapts the effective horizon to the task by assigning negligible effective weight to long-lag positions.

Table 2. Cross-instrument selection: p^* is the smallest period whose cross-instrument geometric mean of validation MAE is within 0.25% of its minimum. This is a parsimony tie-break in the spirit of the one-standard-error rule for model selection (Hastie et al. 2009): when the validation-MAE curve is broadly flat over a range of periods, the smallest such period is preferred for interpretability and computational economy. The rule is binding only for CMA (whose curve is flat from $p \approx 30$ to the search-grid boundary because the fitted weights assign negligible effective weight to most additional lags, Section 3.3); for every other method the smallest in-tolerance period coincides with the strict argmin or differs by a few grid points. Reported quantities are computed on the test segment. Ratios are taken against CMA at its own $p^* = 36$.

Method	p^*	Geom. Mean Ratio	95% CI	Win Rate	Regret	Gen. Ratio
SMA	6	1.0332	[1.0301; 1.0364]	0.00%	0.9995	1.2248
EMA	7	1.0087	[1.0073; 1.0101]	7.27%	1.0015	1.2190
WMA	8	1.0190	[1.0168; 1.0213]	1.82%	1.0012	1.2209
SWMA	5	1.0477	[1.0436; 1.0518]	0.00%	0.9967	1.2208
DEMA	14	1.0280	[1.0257; 1.0303]	0.00%	1.0019	1.2227
TEMA	22	1.0396	[1.0365; 1.0428]	0.00%	1.0014	1.2330
T3	4	1.0445	[1.0407; 1.0482]	0.00%	0.9977	1.2196
ZMA	9	1.0313	[1.0270; 1.0356]	1.82%	1.0012	1.2338
KAMA	2	1.0632	[1.0579; 1.0685]	0.00%	0.9944	1.2121
CMA	36	1.0000	—	—	0.9981	1.2263

Table 3. Per-instrument validation-based selection: for each method and each instrument, p^* is the strict arg min of validation MAE for that instrument. Reported ratios are then computed on the test segment, normalized to CMA's per-instrument test MAE under the same selection rule. Values below 1 would indicate lower test MAE than per-instrument-tuned CMA.

Method	Geom. Mean Ratio	95% CI	Win Rate	Gen. Ratio (OOS/IS)
SMA	1.0318	[1.0272; 1.0365]	1.82%	1.2339
EMA	1.0053	[1.0023; 1.0083]	27.27%	1.2264
WMA	1.0159	[1.0126; 1.0193]	5.45%	1.2292
SWMA	1.0492	[1.0430; 1.0553]	0.00%	1.2341
DEMA	1.0241	[1.0202; 1.0280]	3.64%	1.2265
TEMA	1.0362	[1.0310; 1.0414]	3.64%	1.2215
T3	1.0449	[1.0392; 1.0507]	1.82%	1.2233
ZMA	1.0282	[1.0230; 1.0333]	1.82%	1.2407
KAMA	1.0672	[1.0574; 1.0771]	1.82%	1.2145
CMA	1.0000	—	—	1.2283

Per-instrument selection. For each (method, instrument) pair we choose $p^* = \arg\min_p \mathrm{MAE}_p^{\mathrm{val}}$, i.e., the period that minimizes validation MAE for that instrument. For each instrument we then compute the ratio of the method's test MAE to CMA's test MAE under the same per-instrument validation rule. Table 3 reports the cross-instrument geometric mean of these ratios with log-ratio confidence intervals. The win rate is the fraction of instruments for which the ratio is below 1. The generalization ratio is the cross-instrument geometric mean of each method's test MAE divided by its in-sample MAE under the same validation-selected periods.

Per-instrument selection does not overturn the cross-instrument ranking—a key takeaway of this study, and a sharper conclusion than what test-set-based selection would produce. CMA remains the lowest-MAE method on average; every alternative produces a geometric-mean ratio strictly above 1, and every confidence interval lies above unity. In

other words, under honest validation-based selection, CMA is best under both selection granularities. EMA is again the closest competitor at 1.0053, and it wins on 15 of 55 instruments (27.3% win rate)—substantially more than any other benchmark, consistent with the validation-best-by-period results discussed below. The higher-order exponential constructions (T3, TEMA) sit between 1.04 and 1.05; selecting p on the test set itself (an oracle setup) would make T3 appear competitive at the per-instrument level—a comparison that does not survive an honest validation-only selection. Generalization ratios (test/in-sample) cluster near 1.22–1.24 across all methods including CMA, so no method is meaningfully more overfit than the others under this selection rule.

3.2. Robustness to Alternative Error Metrics

Table 2 reports the primary MAE ratios. As a metric-sensitivity check, Table 4 reports RMSE and MAPE for the same trained filters and the same validation-MAE-selected periods; no weights are re-estimated and no periods are reselected for the alternative metrics. The entries are geometric means of the 55 per-instrument test-error ratios relative to CMA at its validation-MAE-selected $p^* = 36$.

Table 4. Alternative-metric robustness using the validation-MAE-selected periods from Table 2. Values are geometric-mean test-error ratios relative to CMA; values above 1 favor CMA.

Method	RMSE Ratio	MAPE Ratio
SMA	1.0354	1.0222
EMA	1.0075	1.0028
WMA	1.0185	1.0100
SWMA	1.0469	1.0353
DEMA	1.0264	1.0154
TEMA	1.0380	1.0250
T3	1.0423	1.0307
ZMA	1.0354	1.0247
KAMA	1.0482	1.0705
CMA	1.0000	1.0000

CMA retains the lowest geometric-mean test error under both additional metrics. EMA remains the closest competitor: relative to CMA, its error is 0.87% higher under MAE, 0.75% higher under RMSE, and 0.28% higher under MAPE. The RMSE ratio has a 95% confidence interval of [1.0062, 1.0088], while the MAPE interval is [1.0008, 1.0049]; all non-CMA confidence intervals under both metrics remain above 1. Thus the aggregate ranking is robust, although the MAPE advantage over EMA is narrow.

3.3. Characteristics of Optimized CMA Weights

The projected subgradient optimization process yielded a unit-sum weight vector for each of the 62 look-back periods p. The optimization permitted signed weights. Negative coefficients are present in the final returned vectors for 18 of the 62 periods, but all lie between -0.000431 and zero, well below the 0.005 threshold in absolute magnitude. Within each affected period, these negative entries share the same value, consistent with the common correction applied by the final re-projection. They are therefore treated as numerically negligible rather than as economically meaningful negative exposures. Figures 1–3 illustrate the optimized weight distributions for representative short, medium, and long look-backs of $p = 6$, $p = 36$, and $p = 120$. The horizontal axis denotes the lag, where lag 0 corresponds to the most recent observation in the window; the vertical axis shows the weight magnitude.

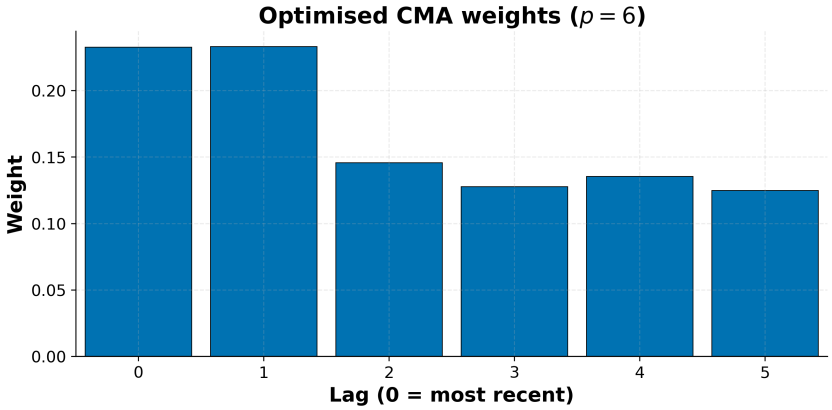

Figure 1. Optimized CMA weights for $p = 6$. All six lags receive non-negligible weight, with mass concentrated on the most recent two.

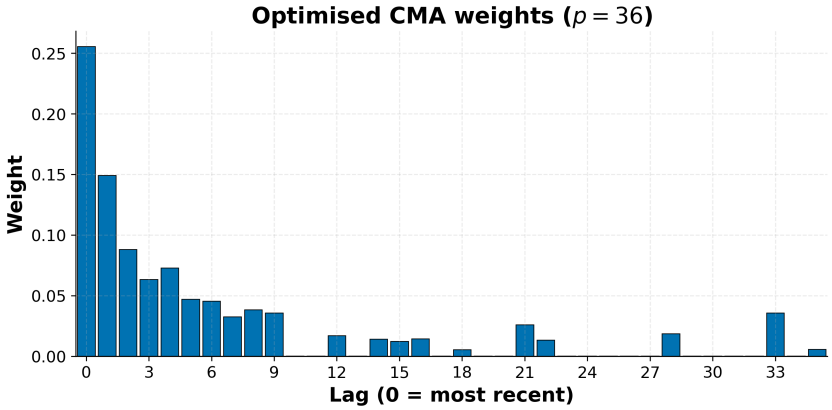

Figure 2. Optimized CMA weights for $p = 36$. 86% of the weight mass falls on lags 0–14; intermediate and far lags receive small but non-zero impulses.

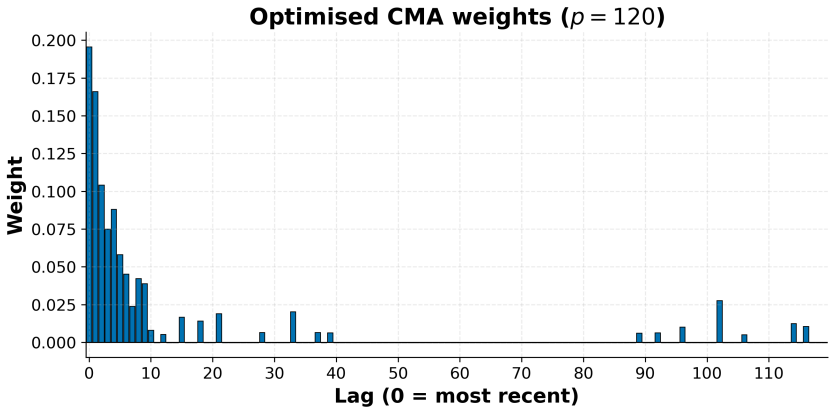

Figure 3. Optimized CMA weights for $p = 120$. 85% of the weight mass again falls on lags 0–14, while the optimizer assigns negligible effective weight to the central span of the window (lags 40–85).

Three regularities are evident:

- **Recency concentration.** In every period, the largest mass lies on the most recent lags, consistent with volatility clustering. For $p = 6$ the solution forms a short plateau over lags 0–2 and then decays; for $p \in \{36, 120\}$ roughly the last 10–15 observations dominate, with about 85–86% of the total mass falling within lags 0–14 in both cases.
- **Concentration on a short effective horizon.** As the nominal window grows from $p = 36$ to $p = 120$, the shape of the weight head barely changes: the optimizer assigns

negligible effective weight to the middle of the window (lags 40–85 for $p = 120$) while preserving a thin tail of small impulses near the boundary. The model thus discovers an effective horizon of roughly 15–20 observations regardless of the nominal p, which explains the flat validation-MAE curve for $p \gtrsim 30$ reported in Table 2.

- **Learned, non-monotonic decay.** Unlike EMA (geometric) or WMA (linear) schemes, CMA learns a mildly non-monotonic head with occasional small "impulses" on intermediate lags. These coefficients are learned solely for predictive accuracy; without separate statistical or structural tests, we do not assign them a calendar, mean-reversion, or other economic interpretation.

3.4. Performance Patterns Across Look-Back Periods

We count, for each period and across the 55 instruments, which method attains the lowest *validation* MAE—that is, which method a per-instrument validation-based rule would select at that look-back. This is the same criterion used to construct Table 3, viewed period by period. Table 5 reports three representative periods (matching the weight figures in Section 3.3), and Figure 4 shows the full short-to-medium range.

Table 5. Number of instruments (out of 55) for which each MA attains the lowest validation MAE at representative periods. For readability, only methods with at least one win at these representative periods are shown.

Period	CMA	EMA	WMA	DEMA	TEMA	SMA	T3
6	0	52	0	1	0	1	1
36	48	4	1	2	0	0	0
120	53	0	1	0	1	0	0
Total (3 periods)	101	56	2	3	1	1	1

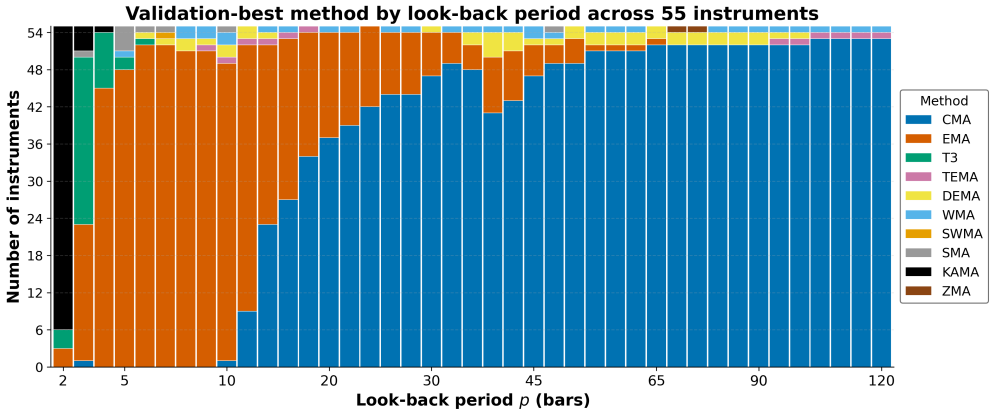

Figure 4. Number of instruments (out of 55) for which each MA attains the lowest validation MAE at each look-back period in the displayed range ($p \leq 120$). Longer periods were included in the quantitative evaluation but omitted here for readability; the CMA-dominated pattern continues monotonically beyond $p = 120$.

Three patterns emerge from Table 5 and Figure 4. First, at very short look-backs ($p \leq 4$), responsiveness dominates and the validation-best method is typically KAMA or T3 (at $p = 2$–3) or EMA (from $p = 3$ onwards). Second, from $p \approx 16$ onwards CMA overtakes EMA as the modal validation-best method; by $p = 30$, it wins on 47 of 55 instruments, by $p = 65$ on 52, and by $p = 120$ on 53. Third, once $p \geq 14$, the higher-order exponential constructions (T3, TEMA, DEMA) account for, at most, 4 of 55 instrument-level wins in any period and never become the modal validation-best method, which is consistent

with the per-instrument Table 3 in which none of these methods improves on CMA on geometric-mean test MAE under honest validation-based selection.

3.5. Single-Instrument Forecast Illustration

To illustrate these patterns on a single asset, Figure 5 overlays one-step-ahead forecasts for the Euro-Bund futures contract (@FGBL) at $p = 36$ over the final nine months of the held-out test window (July 2024–April 2025); the full 5-year test sample is too compressed to be legible at print width and is omitted.

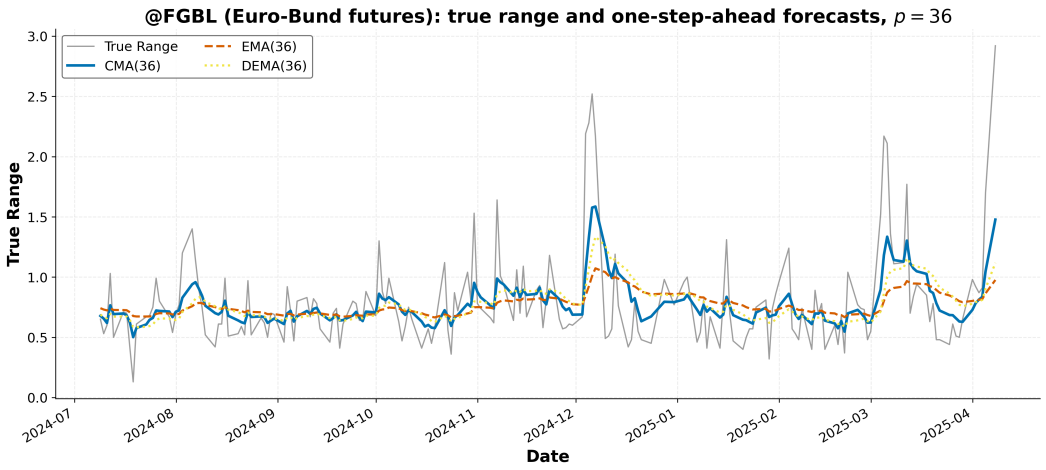

Figure 5. @FGBL (Euro-Bund futures): True Range and one-step-ahead forecasts from CMA(36), EMA(36), and DEMA(36) over the final nine months of the held-out test window (July 2024–April 2025).

CMA tracks large moves more tightly and reverts faster after volatility spikes (most visibly the early-December 2024 episode and the early-March/early-April 2025 spikes), reducing both overshoot and undershoot relative to EMA/DEMA. This overlay illustrates the forecast dynamics at $p = 36$ on a single instrument and is not a substitute for the cross-instrument results reported above.

4. Discussion

This study evaluates whether learning task-specific weighting profiles can improve one-step-ahead forecasting of daily volatility as measured by True Range (TR). The results show a clear period-dependent structure: across the full period grid, the identity of the best-performing smoother depends strongly on the look-back horizon, and the learned CMA weights exhibit systematic regularities that align with well-known properties of volatility such as clustering and short memory.

4.1. Interpretation of Learned CMA Weights

The optimized CMA weights in Section 3.3 place most mass on the most recent observations and decay rapidly, while retaining a very thin long-lag tail and occasional small intermediate impulses. This structure is consistent with the empirical observation that volatility is clustered, meaning the most recent days carry disproportionately high predictive content, while very old observations contribute little incremental information. In this sense, the learned CMA can be viewed as a data-driven analogue of classical exponentially weighted schemes (e.g., EWMA-style volatility filters), but with flexibility to deviate from a strictly geometric decay when intermediate lags contain marginal predictive signal.

The learned weight profiles should be interpreted as average cross-instrument filters over the training sample, not as evidence that the same effective horizon is invariant across all market regimes. Because the current CMA specification uses static weights for each look-

back period, it can respond to crisis periods through larger recent TR observations, but the weighting profile itself does not change unless the model is re-estimated. Regime-specific or walk-forward CMA weights are therefore left for future work.

The non-monotonic intermediate impulses show that the predictive optimum is not restricted to a monotone parametric weight family. However, the coefficients are estimated exclusively by minimizing forecast MAE, without an economic sign restriction or a structural model. We therefore do not interpret the very small signed coefficients as evidence of mean reversion or any other specific market mechanism. Their magnitudes, together with the threshold-and-reprojection procedure, indicate that they should be treated as numerical features of the fitted forecasting filter rather than as economically meaningful positions.

4.2. Cross-Instrument Performance Patterns Across Periods

The validation-best counts in Table 5 and Figure 4 indicate a transition from EMA dominance at very short horizons to CMA dominance from $p \approx 16$ onwards. This pattern is intuitive: when the available history is only a few days, responsiveness is at a premium and a simple exponential weighting can track fast changes with minimal estimation burden. As the window expands, fixed-form parametric schemes face a trade-off between smoothing and lag. CMA partially resolves this trade-off by learning an allocation of weight mass that reduces lag without excessively amplifying noise, which is why it both becomes the modal validation-best method at $p \geq 16$ and attains the lowest geometric-mean test MAE under both selection granularities in Tables 2 and 3.

The single-instrument illustration on @FGBL (Figure 5) is consistent with this cross-instrument picture at medium horizons: CMA(36) tends to react quickly to large volatility shocks while stabilizing faster after spikes, which can reduce both overshoot and undershoot relative to fixed-form alternatives. While such overlays are not a substitute for cross-sectional evaluation, they provide a qualitative view of how the learned weighting profile manifests in practice.

4.3. Selection Granularity Under Validation-Based Selection

Tables 2 and 3 differ only in the granularity at which p^* is chosen: a single global period per method for Table 2, and a per-instrument period for Table 3. In both, p^* is selected on the validation segment of the 60/20/20 split and test MAE is reported on the held-out segment, so both regimes correspond to operational selection rules rather than oracle upper bounds. The central finding is that CMA attains the lowest cross-instrument geometric-mean test MAE under both rules: the result is robust to whether the look-back is allowed to vary per instrument or constrained to a single global value, with EMA as the closest competitor in both cases (0.87% and 0.53% above CMA respectively).

The two tables tell consistent stories. Under cross-instrument selection (Table 2), CMA at $p^* = 36$ attains the lowest test-set geometric-mean MAE; EMA at $p^* = 7$ is the closest competitor at a 0.87% gap, with the confidence interval just above unity. Under per-instrument selection (Table 3), the ordering is essentially unchanged: CMA still has the lowest test MAE on average, and EMA remains the closest competitor at a 0.53% gap. The win rates differ in detail—EMA wins on 27.3% of instruments under per-instrument selection versus 7.3% under cross-instrument selection—but the geometric-mean ranking on test MAE does not flip, and the regret column in Table 2 is uniformly within 0.6% of unity, indicating that one global period transfers cleanly across the cross-section.

The alternative-metric results in Table 4 reinforce, but also qualify, this conclusion. With weights and periods held fixed at their validation-MAE-selected values, CMA remains first under RMSE and MAPE, and the confidence intervals for all non-CMA ratios exclude unity. However, EMA is only 0.28% above CMA under MAPE, compared with 0.75% under

RMSE and 0.87% under MAE. The MAPE difference is therefore statistically distinguishable in this sample but economically small as a forecast-error advantage, and the present analysis does not establish that such a margin improves downstream risk-management outcomes. Because MAPE scales each error by realized TR, the attenuation is consistent with CMA's advantage being stronger in absolute-error terms than in proportional-error terms; we treat this as an interpretation rather than a separately tested mechanism.

The higher-order exponential constructions (T3, TEMA, DEMA) do not improve on CMA under either selection rule. This is a sharper conclusion than what test-set-based selection would have produced: under selection on test MAE, T3 in particular can be made to look more competitive on a per-instrument basis, but that apparent advantage does not survive an honest validation-only selection. The practical interpretation is that the per-instrument periods at which T3 happens to do well on the test set are not the periods that an out-of-sample procedure would identify, and the cross-instrument winner under any honest rule is CMA.

A second feature of CMA visible in Section 3.3 reinforces this conclusion. The optimized weight vector concentrates on a short effective horizon: for $p \in \{36, 120\}$ roughly 85% of the mass falls within lags 0–14, while the fitted weights assign negligible effective weight to the middle of the window. This means the validation-MAE curve for CMA is essentially flat for all $p \gtrsim 30$: the strict argmin happens to lie at the grid boundary, but across $p \in [30, 300]$ the test MAE deviates from its value at $p^* = 36$ by less than about 0.7%. We therefore report $p^* = 36$ under a parsimony tie-break that selects the smallest period within 0.25% of the validation minimum. By contrast, fixed-form smoothers have a sharply peaked validation curve with an optimum at a relatively short period (e.g., EMA at $p^* = 7$, TEMA at $p^* = 22$), reflecting the fact that a fixed-form filter has no mechanism to ignore the added window length.

4.4. Relation to Prior Work and Implications for Volatility Modeling

The findings are consistent with the broader volatility literature in which simple filters (including exponentially weighted constructions) can be competitive for short-horizon volatility tracking, while more flexible structures can improve performance when the effective memory horizon is longer. CMA can be interpreted as a flexible, data-driven linear filter whose coefficients are optimized directly for the forecasting objective, making it conceptually related to distributed-lag models and to the idea of learning the volatility filter rather than selecting it from a fixed family. The empirical weight shapes suggest that most useful information is concentrated in the recent past, while small deviations from monotone decay can capture weak but repeatable structure that parametric averages ignore.

From an applied perspective, the results suggest that practitioners should treat the best moving average as horizon-dependent: short windows reward responsiveness (EMA), whereas medium-to-long windows benefit from learned, task-specific weighting (CMA). Equally important, because the fitted CMA weights concentrate on a short effective horizon ($\approx$20 bars in our data), increasing p beyond that horizon does not degrade out-of-sample accuracy. This is a practically convenient property in a deployment setting where the look-back parameter cannot be retuned frequently, and it stands in contrast to fixed-form smoothers whose accuracy curve is sharply peaked in p.

For practitioners considering live deployment, the main computational difference between EMA and CMA is the weight-optimization step rather than the daily forecast calculation after weights are fixed. EMA uses a fixed analytic weighting rule and requires no numerical training. CMA, by contrast, learns its weights by batch projected subgradient descent, so fitting or re-fitting CMA requires iterative optimization over the training sample. In the present study, this optimization is performed offline over a grid of candidate

look-back periods, and the period-level optimizations are independent. If p is fixed after validation, only the selected CMA specification needs to be re-optimized during periodic recalibration. If the recalibration procedure also re-selects p, the optimization cost scales with the number of candidate periods considered and the number of recalibration dates. Once weights and p are fixed, EMA updates recursively from the previous EMA value, whereas CMA computes a weighted sum over the latest p observations; for daily data and moderate p, this forecast-calculation cost is small relative to offline optimization.

At the same time, the numerical advantage should not be overstated. The main MAE gain over EMA is below 1%, and the MAPE gain is smaller still. These results establish a statistically distinguishable forecasting improvement for TR, but they do not by themselves prove economic value in VaR/ES estimation, position sizing, margin forecasting, hedging, or trading strategies. Those downstream evaluations require separate decision-focused experiments.

4.5. Limitations and Future Research

Several extensions would strengthen the practical and scientific conclusions:

- **Walk-forward and rolling re-estimation.** The current chronological 60/20/20 split is held fixed across all reported numbers. A walk-forward evaluation in which CMA weights and the look-back p^* are refit periodically (e.g., yearly) on an expanding or rolling window would more closely mirror operational deployment and would also surface any non-stationarity in the optimal weighting profile.
- **Broader baseline set.** Moving-average filters are simple benchmarks; future work should compare against standard volatility models (e.g., GARCH-family, stochastic volatility proxies, or HAR-style baselines when realized measures are available) to contextualize gains.
- **Constraint design and regularization.** The current constraint enforces unit-sum weights. Enforcing non-negativity (simplex constraint) and/or adding smoothness or sparsity penalties could improve interpretability and stability, especially at long periods, and may make the concentration of weight on a short effective horizon observed in Section 3.3 explicit rather than emergent.
- **Per-instrument versus global weights.** This study learns global weights per period and transfers them across instruments. Allowing limited per-instrument adaptation (e.g., hierarchical or shrinkage-based deviations from a global template) may yield further gains while controlling overfitting.
- **Economic relevance.** Although MAE on TR is a clean statistical objective, downstream value should be assessed via risk-management metrics (e.g., VaR/ES backtests, margin forecasting, hedging error) and robustness to transaction costs where relevant.
- **Hybrid instrument-specific filter selection.** The horizon-dependent ranking observed here suggests that an ensemble or routing scheme that selects (or blends) different moving-average families for different instruments and look-back regimes could outperform any single global choice, particularly in the short-period regime where EMA is competitive with CMA.

Overall, the evidence supports the working hypothesis that learning weights for the specific forecasting task can improve volatility smoothing beyond fixed-form heuristics, especially at medium-to-long horizons, while also clarifying when traditional exponential filters remain competitive.

5. Conclusions

This paper provided a broad out-of-sample evaluation of CMA and nine benchmark moving averages for forecasting next-day True Range across 55 instruments and 62 look-back periods, under an honest validation-based selection rule applied to a chronological

60/20/20 train/validation/test split. The results show that the relative ranking of moving-average filters is strongly horizon-dependent: EMA is the validation-best method at very short windows, while CMA becomes the modal validation-best method from $p \approx 16$ onwards and dominates by wide margins at medium-to-long horizons, consistent with learned weight profiles that concentrate on recent lags with a thin long-lag tail.

Under both a single-global-period rule and a per-instrument rule, CMA attains the lowest test-set geometric-mean MAE, with EMA as the closest competitor at well under one percent. This is a sharper conclusion than a test-set-based analysis would produce: with the look-back honestly selected on validation rather than on test data, an apparent per-instrument advantage of higher-order smoothers (T3, TEMA) that would emerge under test-set selection does not survive. A practically useful structural finding is that the optimized CMA weights concentrate on a short effective horizon: roughly 85% of the mass falls within the last 15 lags regardless of the nominal p, so the model effectively operates on a horizon of $\approx$20 bars and is robust to the choice of p once it is at least as large as that horizon. Future work should prioritize walk-forward re-estimation, comparison against standard volatility models (GARCH-family, HAR), and evaluation in downstream risk-management settings.

Author Contributions: Conceptualization, A.R. and P.Z.; methodology, A.R., P.Z. and R.V.; software, P.Z., L.L. and L.Ž.; validation, P.Z., L.Ž. and T.P.; formal analysis, P.Z. and A.R.; investigation, P.Z., L.L., L.Ž., T.P., V.V. and J.A.; data curation, P.Z. and L.Ž.; writing—original draft preparation, P.Z. and A.R.; writing—review and editing, L.Ž., A.R., R.V., T.P., V.V. and J.A.; visualization, P.Z. and L.Ž.; supervision, A.R. and R.V.; project administration, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The daily OHLCV series for the 55 instruments analyzed in this study were obtained from TradeStation under a license that does not permit redistribution. The full list of instrument identifiers and the analysis code are available from the corresponding author upon reasonable request, allowing readers with their own access to the same data source to replicate all reported results.

Conflicts of Interest: A.R. is the founder and principal of a Commodity Trading Advisor firm that trades futures contracts of the type analyzed in this study. The research was conducted independently, and the methods, results, and conclusions reflect academic findings. The remaining authors declare no conflicts of interest.

References

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]

Brière, Marie, Alexandre Burgues, and Ombretta Signori. 2010. Volatility exposure for strategic asset allocation. *The Journal of Portfolio Management* 36: 105–16. [CrossRef]

Brooks, Chris, and Gita Persand. 2003. Volatility forecasting for risk management. *Journal of Forecasting* 22: 1–22. [CrossRef]

Chua, Chew Lian, and Sarantis Tsiaplias. 2019. Information flows and stock market volatility. *Journal of Applied Econometrics* 34: 129–48. [CrossRef]

Cont, Rama. 2007. Volatility clustering in financial markets: Empirical facts and agent-based models. In *Long Memory in Economics*. Berlin and Heidelberg: Springer, pp. 289–309. [CrossRef]

Corsi, Fulvio. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7: 174–96. [CrossRef]

Ehlers, John F. 2001. *Rocket Science for Traders: Digital Signal Processing Applications*. Hoboken: John Wiley & Sons.

Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* 50: 987–1007. [CrossRef]

Griffioen, Gerwin A. W. 2003. Technical Analysis in Financial Markets. Doctoral dissertation, University of Amsterdam, Amsterdam, The Netherlands.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer.

Hu, Nan, Xuebao Yin, and Yuhang Yao. 2025. A novel HAR-type realized volatility forecasting model using graph neural network. *International Review of Financial Analysis* 98: 103881. [CrossRef]

Hyndman, Rob J., and Anne B. Koehler. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22: 679–88. [CrossRef]

Joiner, Dakota, Amy Vezeau, Albert Wong, Gaétan Hains, and Youry Khmelevsky. 2022. Algorithmic trading and short-term forecast for financial time series with machine learning models: State of the art and perspectives. In *Proceedings of the 2022 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*. Piscataway: IEEE, pp. 1–9.

Kaufman, Perry J. 2013. *Trading Systems and Methods*, 5th ed. Hoboken: John Wiley & Sons.

Leushuis, Radmir Mishelevich, and Nicolai Petkov. 2026. Advances in forecasting realized volatility: A review of methodologies. *Financial Innovation* 12: 14. [CrossRef]

Luo, Jiawen, Tony Klein, Thomas Walther, and Qiang Ji. 2024. Forecasting realized volatility of crude oil futures prices based on machine learning. *Journal of Forecasting* 43: 1422–46. [CrossRef]

Marcucci, Juri. 2005. Forecasting stock market volatility with regime-switching garch models. *Studies in Nonlinear Dynamics & Econometrics* 9: 6. [CrossRef]

Mulloy, Patrick G. 1994. Smoothing data with faster moving averages. *Technical Analysis of Stocks & Commodities* 12: 11–19.

Patton, Andrew J. 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160: 246–56. [CrossRef]

Poon, Ser-Huang, and Clive W. J. Granger. 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature* 41: 478–539. [CrossRef]

Raudys, Aistis, and Edvinas Goldstein. 2022. Forecasting detrended volatility risk and financial price series using LSTM neural networks and XGBoost regressor. *Journal of Risk and Financial Management* 15: 602. [CrossRef]

Raudys, Aistis, and Židrina Pabarškaitė. 2018. Optimising the smoothness and accuracy of moving average for stock price data. *Technological and Economic Development of Economy* 24: 984–1003. [CrossRef]

Tillson, Tim. 1998. Better moving averages. *Technical Analysis of Stocks & Commodities* 16: 33–37.

Wilder, J. Welles. 1978. *New Concepts in Technical Trading Systems*. Greensboro: Trend Research.

Yamai, Yasuhiro, and Toshinao Yoshiba. 2005. Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance* 29: 997–1015. [CrossRef]

Yang, Ke, Nan Hu, and Fengping Tian. 2024. Forecasting crude oil volatility using the deep learning-based hybrid models with common factors. *Journal of Futures Markets* 44: 1429–46. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.