






Article

Simulated Annealing Applied to Alternative Assets in Mexican Stock Exchange

Jose Luis Purata Aldaz ¹, Juan Frausto Solís ^{1,*}, Juan J. Gonzalez Barbosa ¹, Guadalupe Castilla-Valdez ¹
and Juan Paulo Sánchez Hernández ²

¹ Graduate Program Division, Tecnológico Nacional de México/Instituto Tecnológico de Ciudad Madero, Ciudad Madero 89440, Mexico; luis.pa@cdmadero.tecnm.mx (J.L.P.A.); juan.gb@cdmadero.tecnm.mx (J.J.G.B.); guadalupe.cv@cdmadero.tecnm.mx (G.C.-V.)

² Informatics, Electronics and Telecommunications Division, Universidad Politécnica del Estado de Morelos, Blvd Cuauhnáhuac 566, Jiutepec 62574, Mexico; juan.paulosh@upemor.edu.mx

* Correspondence: juan.frausto@gmail.com

Abstract

Accurate price forecasting and portfolio optimization in emerging financial markets remain challenging due to the non-stationary dynamics, structural breaks, and heterogeneous behavior of traded instruments. This paper proposes the Time-series Adaptive Forecast Ensemble (TAFE), a method that combines single popular forecasting methods, such as ARIMA, by using an algorithm derived from both the simulated annealing (SA) and Threshold Accepting algorithms. The TAFE is applied to twenty-four weekly price series of Mexican exchange-traded funds (ETFs) and Real Estate Investment Trusts (FIBRAS) over the period 2020–2025. A top-K pre-selection strategy is used, mitigating the adverse cross-model interaction effect of some assets over others, in other words, reducing the propagation of errors from poorly performing base learners. In addition, the sample results show that the TAFE achieves the lowest mean SMAPE across the panel, with statistical superiority over the equal-weight benchmark and a Hybrid Model, confirmed by Diebold–Mariano and Harvey–Leybourne–Newbold tests. Out-of-sample evaluation over a 26-week horizon reveals a regime-shift-driven performance reversal consistent with the bias–variance tradeoff in adaptive combination schemes. Portfolio optimization using SA-generated forecasts yields with an expected return of 35.77%; thus, the model presents a slight overestimation of the return, with a variance of 2.4%. However, it has an acceptable level of risk.

Keywords: investment portfolios; optimization; forecast; heuristics



Academic Editors: Daniel E. Hernández, Marcela Quiroz-Castellanos, Leonardo Trujillo, Pierrick Legrand and Oliver Schütze

Received: 30 December 2025

Revised: 5 May 2026

Accepted: 7 May 2026

Published: 13 May 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](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In a financial environment characterized by high volatility, structural changes in markets, and a growing number of instruments available for investment, the efficient building of investment portfolios has become a technical and strategic challenge. In this context, investors seek to maximize risk-adjusted returns by integrating assets from the Mexican Stock Exchange, such as exchange-traded funds (ETFs) and Real Estate Investment Trusts (FIBRAS) or Real Estate Investment Trusts, REITs in the USA, which offer an attractive combination of performance, diversification, and liquidity.

The present study explores the growing importance of ETFs and FIBRAS in the Mexican financial market, particularly with regard to their role in diversification and the attraction of capital. A study of the exchange-traded funds (ETFs) listed on the Mexican

Stock Exchange reveals a consistent rise in liquidity and participation, thereby solidifying their status as cost-effective instruments for replicating both domestic and international indices [1–3]. Prominent examples include iShares NAFTRAC and Vanguard FTSE BIVA Mexico Equity. Conversely, FIBRAs attained a net asset value in excess of 891 billion pesos in 2025, exhibiting an average occupancy rate approaching 95%, propelled by industrial demand stemming from nearshoring [4]. Furthermore, the Fibras Index demonstrated a notable accumulation of returns, reaching 19.5% in 2025. This figure surpassed the 15.9% return of the S&P/BMV IPC, underscoring its capacity to generate value through both dividend yield and certificate appreciation [5]. These data confirm that the analysis of ETFs and FIBRAs is essential to understanding the current dynamics of the Mexican market and its role in attracting investment in a context of economic transformation.

The classic theory of portfolio optimization, based on the mean–variance model proposed by Markowitz [6], has been largely superseded by more flexible and adaptive approaches, especially in scenarios where assumptions of normality and linearity do not hold. In recent years, heuristic algorithms, such as simulated annealing (SA) [7], have been robust tools for the efficient exploration of complex and non-convex search spaces, which are typical in asset allocation problems with constraints.

The present work proposes an integrated strategy for the selection and weighting of high-return assets in investment portfolios, using a combination of financial time series forecasting methods and metaheuristic optimization algorithms. In the initial phase, a range of forecasting models is employed, including ARIMA, seasonal decomposition, feed-forward neural networks, and recurrent neural networks (RNNs), to estimate the future evolution of the prices of the selected assets. Subsequently, an ensemble strategy is constructed to combine the forecasts, where the weighting is optimized using the Threshold Accepting algorithm, with the aim of minimizing the prediction error [8].

The implementation of a portfolio selection model using simulated annealing is informed by the projections generated, with the objective function being optimized based on a modification of the Sharpe ratio. This incorporates a penalty for asset correlation and dispersion. The proposed approach is empirically evaluated on a set of ETFs and FIBRAs listed in emerging and developed markets, emphasizing its capacity to generate investment combinations that exhibit superior performance and a diversified structure.

This study addresses two fundamental questions in the management of portfolios of alternative assets in emerging markets: First, it asks whether ensemble forecasting methods optimized through metaheuristic algorithms produce more accurate return predictions than individual forecasting models or simple combination strategies. Second, it asks whether the integration of optimized forecast ensembles with metaheuristic portfolio allocation generates superior risk-adjusted returns compared to traditional portfolio construction approaches. In order to answer these questions, we formulated three testable hypotheses.

- (a) **Forecast Ensemble Superiority:** The ensemble forecasting model optimized using Threshold Accepting (Comb_TA) produces significantly lower forecast errors compared to individual forecasting methods (ARIMA, STLF, NNETAR, TBATS, ARFIMA) and naive ensemble approaches (equally weighted combination, quick model) when predicting the weekly returns of Mexican ETFs and FIBRAs. One way to evaluate this idea is through the application of metrics: the Symmetric Mean Absolute Percentage Error (sMAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE).

To provide statistical certainty to the proposed premise, a Diebold–Mariano test for pairwise forecast accuracy comparisons with a significance level of $\alpha = 0.05$ will be applied.

- (b) **Portfolio Performance Enhancement.** The portfolio constructed using simulated annealing optimization with forecasted returns from the optimized ensemble (final portfolio) generates a significantly higher Sharpe ratio compared to the initial portfolio based on historical returns alone (initial portfolio), after accounting for forecasting uncertainty. The metrics to evaluate are the Sharpe ratio, expected return, portfolio risk (standard deviation), correlation structure, and maximum drawdown. For statistical validation, a Ledoit–Wolf test is conducted for Sharpe ratio differences, and bootstrap confidence intervals are used at a 95% confidence level.
- (c) **Risk Administration Through Optimization.** The integration of optimized forecasting and portfolio rebalancing achieves substantial risk reduction (measured by portfolio standard deviation) while maintaining competitive expected returns, demonstrating the effectiveness of the modified Sharpe ratio objective function that incorporates penalties for asset correlation and weight dispersion. For evaluation, it will be necessary to calculate portfolio volatility and correlation. To confirm the above, an out-of-sample performance evaluation will be carried out on the test set, which is 10% of the data. This evaluation will entail comparing risk metrics between the initial and optimized portfolios and conducting statistical significance tests.

These hypotheses are designed to be directly testable using the experimental framework described in Section 4, with the results evaluated against well-defined statistical criteria to ensure robust and reproducible findings.

1.1. Research Gap

Despite extensive research in portfolio optimization and financial forecasting, significant gaps persist in the integration of these methodologies, particularly in emerging market contexts. The existing literature exhibits three primary limitations that this study addresses:

- (a) **Limited Integration of Forecasting Optimization and Portfolio Construction**

While ensemble forecasting methods and metaheuristic portfolio optimization have been independently investigated, the literature lacks comprehensive approaches that optimize forecast combination weights using metaheuristic algorithms before employing these forecasts in portfolio selection. Most studies either use simple averaging for forecast combinations or optimize portfolio weights with historical returns but rarely integrate both optimization stages in a unified framework. This sequential optimization approach (first optimizing forecast ensembles, then using these forecasts for portfolio optimization) represents a little-explored approach to research with potential for substantial performance improvements.

- (b) **Insufficient Focus on Mexican Alternative Assets**

The portfolio optimization literature heavily concentrates on developed markets (primarily U.S., European, and Japanese equities) or broadly defined emerging market indices. Research specifically targeting Mexican ETFs and FIBRAs using advanced optimization techniques is notably absent. Given Mexico's unique position in Latin American markets (characterized by nearshoring trends, evolving monetary policy, and growing financial market sophistication), this asset class merits dedicated research attention. The distinct risk–return characteristics and correlation structures of Mexican alternative assets require tailored optimization strategies that account for market-specific features.

- (c) **Neglect of Forecast Uncertainty Propagation**

Existing portfolio optimization studies that incorporate forecasting typically do not explicitly model how forecast errors propagate into portfolio performance metrics. The implicit assumption that forecasts are point estimates without uncertainty intervals fails to

account for the reliability of return predictions and their impact on portfolio composition. A robust portfolio optimization framework should acknowledge forecast uncertainty and demonstrate performance stability across different forecast accuracy scenarios.

1.2. Contributions of This Work

This study makes three distinct contributions to the portfolio optimization and financial forecasting literature:

We propose a fully integrated methodology that combines Threshold Accepting (TA) for optimizing forecast ensemble weights with simulated annealing for portfolio asset selection and allocation. Unlike previous approaches that treat forecasting and portfolio optimization as independent problems, our framework establishes a direct optimization linkage where ensemble forecast quality directly influences portfolio performance. The analytical parameter tuning method [9] ensures systematic hyperparameter selection for both optimization stages, enhancing reproducibility and practical applicability.

The methodological novelty extends beyond a simple combination of existing techniques. The specific algorithm assignment (TA for forecast ensemble optimization and SA for portfolio construction) is justified by their complementary characteristics: TA's computational efficiency suits the high-frequency forecast weight adjustment problem, while SA's robust global search capabilities address the more complex portfolio allocation problem with multiple constraints and non-convex objective landscapes.

This study provides the first systematic evaluation of integrated forecast–portfolio optimization specifically for Mexican ETFs and FIBRAs. Our empirical analysis spans 335 weekly observations (January 2019–June 2025), encompassing 26 alternative assets that represent diverse sectors of the Mexican economy. The train–validation–test split enables rigorous out-of-sample evaluation, addressing concerns about the in-sample overfitting common in portfolio optimization studies.

The empirical results establish performance benchmarks for this under-researched asset class and demonstrate the practical viability of advanced optimization techniques in emerging market contexts. Specifically, we document substantial improvements in Sharpe ratios and risk reduction, providing evidence that metaheuristic approaches can generate economically meaningful benefits in real-world investment scenarios.

We introduce and empirically validate a modified Sharpe ratio objective function that explicitly penalizes high inter-asset correlation and excessive portfolio concentration. This extension addresses two practical concerns in portfolio management: (1) correlation instability during market stress periods that can undermine diversification benefits and (2) overly concentrated portfolios that fail to achieve adequate risk dispersion despite optimization.

The modified objective function balances return maximization, risk minimization, correlation management, and diversification enforcement within a single optimization framework. Our empirical results demonstrate that this multi-dimensional objective yields portfolios with superior out-of-sample performance compared to standard Markowitz–Sharpe optimization approaches, suggesting that the explicit incorporation of correlation and concentration penalties enhances portfolio robustness.

2. Background

2.1. Portfolio Optimization with Metaheuristic Algorithms

The application of metaheuristic algorithms to portfolio optimization has attracted considerable research attention due to the computational challenges posed by realistic portfolio models. Modern portfolio optimization problems frequently incorporate non-linear objective functions, cardinality constraints, and transaction costs, which render traditional quadratic programming methods computationally inefficient [10,11]. Erwin &

Engelbrecht (2023) conducted a comprehensive survey of over 140 publications spanning three decades (1993–2023), systematically categorizing approaches according to problem type (unconstrained vs. constrained) and algorithmic methodology (single-objective vs. multi-objective) [10]. Their analysis demonstrates that metaheuristic approaches, particularly evolutionary and swarm intelligence algorithms, offer robust alternatives for approximating optimal solutions in complex portfolio models.

Simulated annealing has emerged as a particularly effective metaheuristic for portfolio optimization due to its capacity to escape local optima through the probabilistic acceptance of suboptimal solutions [7]. Recent applications demonstrate its continued relevance: Gunjan & Bhattacharyya (2024) compared classical SA with Quantum-Inspired Simulated Annealing (QiSA) for portfolio optimization, finding that quantum-inspired techniques can match or potentially exceed traditional methods in computational efficiency while maintaining solution quality [11]. Lai et al. (2023) extended SA to multi-objective portfolio selection, demonstrating its capability to simultaneously optimize return, risk, and other relevant objectives [12].

Threshold accepting represents a deterministic variant of SA that eliminates the stochastic component by accepting solutions within a predefined threshold [8]. This deterministic acceptance criterion offers computational advantages in specific optimization contexts, particularly when the objective landscape is well-understood. The analytical tuning methodology provides a systematic approach for parameter selection in both SA and TA, enhancing their practical applicability [9]. Recent metaheuristic innovations continue to expand the algorithmic toolkit for portfolio optimization [13,14].

2.2. Ensemble Forecasting in Financial Markets

Ensemble methods have gained prominence in financial time series forecasting due to their capacity to leverage the complementary strengths of heterogeneous models while mitigating individual model weaknesses. In [15], it was demonstrated that forecast combination and model averaging strategies can substantially reduce variance and improve generalization performance in financial applications. This theoretical foundation supports the widespread adoption of ensemble approaches in portfolio management contexts.

The integration of classical statistical models with machine learning techniques represents a particularly active research direction. In [16], a deep learning ensemble model combining Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Autoregressive Moving Average (ARMA) models showed that hybrid architectures can effectively capture both linear and non-linear patterns in financial time series. The empirical results showed that ensemble-based forecasting models consistently outperform individual components across multiple financial assets. Other works have reported combination forecasting for stock market return prediction, finding that mean combination forecasting models achieve superior out-of-sample performance compared to individual forecasting methods [17]. This finding is consistent with the broader literature on ensemble learning, which emphasizes the reduction in prediction variance through model aggregation. The `hybridModel` package in R is documented in [18], where it is implemented in an ensemble framework that combines multiple base models (ARIMA, ETS, neural networks, TBATS) with flexible weighting schemes. In addition, several advances in ensemble methodologies have explored more sophisticated combination strategies. Adaptive ensemble approaches that dynamically weight individual forecasts based on recent performance have shown promise in capturing regime shifts and non-stationary patterns common in financial markets. The optimization of ensemble weights through metaheuristic algorithms represents a natural extension of this research direction, potentially offering

superior forecast combinations compared to simple averaging or historical error-based weighting schemes.

2.3. *Integrated Approaches: Forecasting and Portfolio Optimization*

The integration of forecasting methods with portfolio optimization represents a critical yet underexplored research area. Traditional portfolio optimization typically assumes that expected returns are known or estimated from historical data. However, this approach neglects forecast uncertainty and its propagation into the portfolio selection process. The explicit incorporation of forecasting models into the portfolio optimization framework offers the potential to improve out-of-sample portfolio performance, provided that forecast errors and their impact on portfolio composition are properly accounted for.

Research combining forecasting ensembles with metaheuristic portfolio optimization remains relatively limited. The existing literature has largely treated forecasting and portfolio allocation as sequential but independent problems. This separation may be suboptimal, as forecast uncertainty directly affects the reliability of portfolio weights and risk estimates. A fully integrated approach that simultaneously optimizes forecast combinations and portfolio allocations could potentially yield superior risk-adjusted returns.

2.4. *Portfolio Optimization in Emerging Markets*

Emerging market portfolio optimization presents unique challenges due to higher volatility, lower liquidity, and distinct market microstructures compared to developed markets. Latin American markets, including Mexico, have received relatively limited attention in the portfolio optimization literature despite their growing importance in global investment portfolios. The MSCI Emerging Markets Latin America Index has demonstrated substantial volatility and divergent performance patterns compared to broader emerging market indices, suggesting that region-specific strategies may offer meaningful advantages.

ETFs and REITs, known as FIBRAs in Mexico, represent an attractive asset class for portfolio optimization due to their combination of diversification benefits, liquidity, and accessibility to retail and institutional investors. However, the academic literature specifically addressing portfolio optimization for Mexican alternative assets using advanced metaheuristic techniques is notably sparse. This gap is particularly evident when considering integrated approaches that combine ensemble forecasting with metaheuristic portfolio allocation.

Recent market developments in Latin America, including nearshoring trends, evolving monetary policies, and structural economic reforms, create a dynamic environment that may reward sophisticated portfolio optimization strategies. The application of metaheuristic algorithms to this specific market context, coupled with advanced forecasting techniques, represents a potentially valuable contribution to both the academic literature and practical portfolio management.

2.5. *Research Gap and Positioning*

The literature review reveals some gaps that this study addresses.

- (a) Limited integration of forecasting and optimization: While ensemble forecasting and metaheuristic portfolio optimization have been studied independently, their full integration—where forecasting weights are themselves optimized using metaheuristics—remains underexplored.
- (b) Underrepresentation of emerging markets: Most portfolio optimization studies focus on developed markets (U.S., Europe, Japan, and recently China) or broadly defined emerging market indices. Research specifically targeting Mexican alternative assets (ETFs and FIBRAs) using advanced optimization techniques is notably absent.

- (c) Insufficient attention to forecast uncertainty: Many portfolio optimization studies that incorporate forecasting do not explicitly address how forecast errors propagate into portfolio performance or implement systematic robustness checks against forecast uncertainty.
- (d) Need for algorithm-specific justification: While numerous metaheuristic algorithms have been proposed, the rationale for selecting specific algorithms (e.g., TA for ensemble weights vs. SA for portfolio allocation) is often underspecified in the existing literature.

This study contributes to addressing these gaps by proposing and empirically evaluating an integrated methodology that combines Threshold Accepting-optimized forecast ensembles with simulated annealing-based portfolio selection, specifically tailored to the Mexican alternative asset market. This research provides both methodological contributions (demonstrating the viability of fully integrated forecast–portfolio optimization) and empirical contributions (establishing performance benchmarks for this asset class using advanced techniques).

3. Methodological Framework

The present study employs a quantitative, exploratory, and applied approach based on forecasting techniques, machine learning, and metaheuristic optimization. The methodology is structured in four main phases:

3.1. Data Collection and Preparation

ETFs and FIBRAs with high liquidity, sector coverage, and historical price availability are selected. The data is obtained from platforms such as Yahoo Finance, in weekly frequencies, and transformed into a series to calculate the fractional change (relative change) between the current and a prior element.

In the analysis of financial time series, the presence of missing data—particularly in recently listed assets such as certain FIBRAs and ETFs—poses significant statistical challenges. The exclusion of these instruments would result in a reduction in the representativeness of the sample, while the utilization of simple methods such as forward-fill or linear interpolation leads to an underestimation of actual volatility. To address this, a proxy imputation technique (market-relative) was implemented, based on preserving the covariance structure and the dynamics of systematic risk.

The method, known in practice as proxy filling, is related to academic approaches to the systematic imputation of financial data [19] and to Expectation–Maximization techniques for missing data [20], which allow for the reconstruction of time series under the assumption of market-relative behavior.

Methodologically, the primary local market benchmark (NAFTRAC) was utilized as a market anchor. For each asset A with an incomplete history, backward extrapolation was performed to estimate its price at a given time $t - 1$ ($\hat{P}_{A,t-1}$) calculated from the known price in t ($\hat{P}_{A,t}$) and the relative performance observed in the proxy index (1):

$$\hat{P}_{A,t-1} = P_{A,t} \times \frac{I_{t-1}}{I_t} \quad (1)$$

This is critical to the robustness of optimization models, as it allows for the calculation of a positive-definite variance–covariance matrix and prevents bias in the Sharpe ratio by not omitting periods of high systemic volatility. By stabilizing the input matrix, the simulated annealing algorithm achieves a more efficient exploration of the frontier, identifying asset combinations that more accurately reflect the risk–return balance for investors in the context of the Mexican real estate and financial markets.

This approach ensures that the variance injected into the imputed data is not arbitrary but reflects market conditions during the corresponding period.

3.2. Asset Selection

The process of constructing an investment portfolio has become increasingly complex in the contemporary financial landscape, due to the numerous factors that can influence the selection, integration, and weighting of assets within an investment portfolio. It is imperative that robust and reliable tools are employed in order to facilitate the analysis of a multitude of candidate assets and the data that integrates them. As outlined in Section 1, the MPT model, proposed by Markowitz, is the seminal work in portfolio management which considers two main objectives: maximizing the expected return and minimizing the risk [21]. The Sharpe model [22], as shown below, constitutes an MPT extension that considers the two objective functions within a single framework. In the present work, asset selection is defined by the following Equations (2)–(6):

$$\max SR = \frac{E(R_{pf}) - Rf}{\sqrt{\sum_{i=1}^n \sigma_i^2 x_i^2 + \sum_{i=1}^n \sum_{j \neq i}^n \sigma_{ij} x_i x_j}} \quad (2)$$

which is subject to

$$\sum_{i=1}^n x_i = 1 \quad (3)$$

$$\sum_{i=1}^n E(R_{pf}) > MARR \quad (4)$$

$$0 < x_i \leq 1 \quad (5)$$

$$\sum_{i=1}^n \sigma_i^2 x_i^2 + \sum_{i=1}^n \sum_{j \neq i}^n \sigma_{ij} x_i x_j \leq \Phi_{max}^2 \quad (6)$$

where $E(R_{pf})$ is the expected return, i and j represent assets, Rf is a risk-free option, x_i, x_j are the contributions of i and j assets, σ_i, σ_j are the standard deviations of the assets, and Φ_{max}^2 is the Maximum Risk Rate Allowable. The parameter MARR represents the Minimum Acceptable Rate of Return that an investor could accept, but the Markowitz model uses a risk-free rate. The Φ_{max}^2 risk parameter defines the assets that integrate the portfolio. This parameter quantifies the risk associated with the asset.

The simulated annealing algorithm [7] is based on the Metropolis algorithm [23], which is used in the heat treatment of metals. SA begins with the selection of an initial solution, followed by the random selection of a neighbor solution. A new solution is accepted if it is shown to be superior to the previous one. If not, the new solution is accepted based on a probability calculated using the Boltzmann distribution. This approach ensures that the acceptance rate of incorrect solutions decreases as the algorithm is executed. This strategy is employed to evade local optima. The parameters of this algorithm include an initial temperature T_i , a final temperature T_f , a cooling rate α , and the number of iterations for each Metropolis cycle. In the present study, the SA algorithm functions as both an optimization heuristic and a variable selection filter, thereby contributing to the methodological depth of this research. Moreover, the objective is to mitigate variance. It has been observed that assets that demonstrate resilience across multiple random seeds of the SA algorithm tend to possess optimal diversification characteristics for FIBRAs and ETFs.

In Algorithm 1, the process of the SA algorithm [24] that seeks to optimize the set of assets is shown. It is important to note that the algorithm is based on decreasing the initial temperature, and looking for the best target value is necessary to adjust the final value of the

SR, which is achieved by multiplying the final result by minus 1. The maximization of the SR function of Equation (2) is performed using the property $\max SR = \min(-SR) = \min \tau$ where $\tau = -SR$. Therefore, Algorithm 1 maximizes SR by minimizing τ .

Algorithm 1. SA algorithm applied

```

1:  Parameters ( $T_i, T_f, \alpha, L_k, Q$ )
2:  Generate random initial solution  $X_c = X_{best}$ 
3:   $s_c = f(X_n) = SR_{neg} = s_{best}$ 
4:   $k = 0; T_i = T_k$ 
5:  while  $T_k \geq T_f$  do
6:    while  $k < L_k$  do
7:       $X_n = Perturbatiok(X_c); s_n = f(X_n)$ 
8:       $\Delta = s_n - s_c$ 
9:      if  $\Delta < 0$  then
10:        $s_c = s_n; X_c = X_n$ 
11:       if  $s_c < s_{best}$  then
12:         $s_{best} = s_c; X_{best} = X_c$ 
13:       end if
14:       else if  $\text{random}(0, 1) < e^{-\Delta/T_k}$  then
15:         $s_c = s_n; X_c = X_n$ 
16:       end if
17:        $k = k + 1$ 
18:     end while
19:      $L_{k+1} = \beta * L_k$ 
20:      $T_{k+1} = \alpha * T_k$ 
21:   end while
22:    $SR = -s_{best}$ 
23:   return  $SR, X_{best}$ 
24: end SA

```

The Threshold Accepting algorithm was first introduced by [8] and is comparable to simulated annealing. The key distinction lies in the acceptance criteria for new solutions. In TA, a suboptimal solution may be acceptable provided the decrease remains within a predefined tolerance or threshold, which is subject to reduction during the execution of the algorithm. This criterion ensures that probabilities are not calculated and no random decisions are made. The parameters of this algorithm are as follows: the number of iterations, n_{iter} ; the number of steps, n_{steps} ; and a threshold sequence.

Algorithm 2 presents a general overview of the pseudocode for the optimization model developed for this study, the Time-series Adaptive Forecast Ensemble (TAFE). The TAFE is responsible for finding the optimal weights of the individual forecasting methods that make up the ensemble and yield the lowest error metric. This algorithm's notable feature is its adaptive capability in both the initial and final temperature, which is based on the initial reference error generated by the equally weighted forecast assembly, multiplied by an arbitrarily assigned τ factor. In this experiment, the τ_1 value was set to 50, and the τ_2 value was set to 0.0001. These values were determined through experimentation and provide a margin in the algorithm's search space. The tolerance value was assigned based on the previously mentioned error.

Algorithm 2. TAFE algorithm

```

1:  Threshold Accepting ( $T_o, T_f, \alpha, \gamma, L_k$ )
2:   $T_o = \text{err\_initial} * \tau_1$ 
3:   $T_f = \text{err\_initial} * \tau_2$ 
4:   $\alpha$ : Threshold reduction factor ( $\alpha \in (0, 1)$ )
5:   $\gamma$ : Tolerance reduction factor.
6:  err_initial: Initial reference error.
7:  pv_obs: Observed time series values.
8:  pv_models: Set of individual forecasts
9:   $s_c$  = Initialize best local and global solutions ( $X_{best}$ ) with uniform weights.
10:  $k = 0; Tol_k = \text{err\_initial}$ .
11: while  $T_k \geq T_f$  do
12:   while  $k < L_k$  do
13:      $X_n = \text{NewSolution}()$  #new weights
14:     FctTA = new_comb( $X_n, \text{pv\_models}$ ) #Calculate combined forecast
15:      $s_n = \text{error\_calculation}(\text{FctTA}, \text{pv\_obs})$ . #Evaluate combination error
16:      $\Delta = s_n - s_c$ 
17:     if  $\Delta < Tol_k$  then
18:        $S_c = S_n; X_c = X_n$ 
19:       if  $s_c < s_{best}$  then
20:          $s_{best} = s_c; X_{best} = X_c$ 
21:       end if
22:     end if
23:      $k = k + 1$ 
24:   end while
25:    $Tol_{k+1} = \gamma * Tol_k$ 
26:    $T_{k+1} = \alpha * T_k$ 
27:    $k = 0$ 
28: end while
29: return  $s_c, X_{best}$  # minimum error and optimal weights.
30: end TAIPO

```

It is important to note that both algorithms require the tuning of their hyperparameters. As illustrated in both [9,24], an analytical method is proposed for the determination of initial and final values (temperatures) for the two algorithms. This approach is expected to yield a range of solutions at the initial temperature of the algorithm, prevent stagnation at local optima, and identify optimal solutions as the algorithm approaches its final temperature.

3.3. Yield Forecasting

Forecasting empowers financial analysts to anticipate and adapt to evolving market conditions, enabling them to make more timely and effective decisions. The selection of forecasting models for this study is guided by the distinctive statistical properties exhibited by financial time series, particularly those derived from emerging market alternative assets such as Mexican ETFs and FIBRAs. Financial return series typically display several stylized facts that standard forecasting approaches may fail to capture adequately [25]: (i) non-stationarity and structural breaks induced by macroeconomic shocks or policy regime shifts; (ii) volatility clustering, where periods of high volatility tend to be followed by high volatility; (iii) potential long-range dependence or “long memory” in which autocorrelations decay slowly; (iv) non-linear dynamics that cannot be adequately represented by

linear models; and (v) in certain contexts, multiple seasonalities or calendar effects that complicate the forecasting task.

To address these challenges, we employ a heterogeneous ensemble of forecasting models that differ fundamentally in their underlying assumptions, structural complexity, and the types of temporal patterns they are designed to capture. This diversity is intentional: by combining models with complementary strengths, the ensemble can achieve superior out-of-sample performance relative to any individual component [15,17]. The selected models span a spectrum of complexity and flexibility, ranging from parsimonious linear specifications to flexible non-linear architectures.

3.3.1. ETS

The ETS method [26] is an advanced version of traditional exponential smoothing techniques. Classical methods only provide point estimates. The ETS approach is based on state space models. This allows for a statistical foundation based on maximum likelihood and the generation of prediction intervals.

The model's name represents three main parts:

- i. E (Error): This can be either additive or multiplicative.
- ii. T (Trend): This describes how things will behave over time. There are three types: none (N), additive (A), and additive damped (A_d).
- iii. S (Seasonality): This captures repeating patterns. These patterns can be none (N), additive (A), or multiplicative (M) [27].

In its simplest form (ANN or Simple Exponential Smoothing model), the model is defined by the interaction of an observation equation and a state equation:

The observation equation is given by Equation (7):

$$y_t = l_{t-1} + \varepsilon_t \quad (7)$$

The state equation at the indicated level is given by (8):

$$l_t = l_{t-1} + \alpha \varepsilon_t \quad (8)$$

y_t is the observed value, l_t is the level of the series, and $0 \leq \alpha \leq 1$ is the smoothing parameter. The most recent values are given more weight, but this weight decreases over time. Some reasons to use it for financial forecasting are as follows:

- Adaptability to Volatility: It adapts more quickly to changes in level than simple moving average models.
- Noise Management: Since financial markets have a high signal-to-noise ratio, ETS's ability to break the series down and treat "error" explicitly helps filter out short-term changes from the underlying trend.
- Automated Model Selection: Information criteria (like AICc) are used to make objective choices about the best structure. For example, it helps decide whether the trend should be smoothed or whether the error is multiplicative. This reduces researcher bias in the fitting phase.

3.3.2. STLF

The `stlf()` method combines seasonal decomposition using STL (Seasonal and Trend decomposition using Loess) with an exponential smoothing (ETS) model applied to the seasonally adjusted series. First, it extracts the seasonal component with STL, then adjusts an ETS to the seasonally adjusted series, and finally reincorporates seasonality to generate the forecast. This technique is effective for series with seasonal patterns that change over time [28].

This method addresses the possibility of evolving seasonal patterns or structural trend components that ARIMA may inadequately model. Although weekly financial returns are not typically seasonal in the classical sense, certain calendar effects (month-end rebalancing, tax loss harvesting, regulatory reporting cycles) can introduce quasi-periodic patterns that STLF can exploit.

3.3.3. TBATS

The TBATS model (Trigonometric seasonality, Box–Cox transformation, ARMA errors, Trend, Seasonal components) extends the exponential smoothing framework of state space to handle multiple complex seasonalities. The Box–Cox transformation is employed to stabilize variance. The models are capable of displaying trends either with or without damping. The model incorporates ARMA errors. The representation of seasonality is achieved through the utilization of trigonometric (Fourier) terms. The algorithm automatically selects components and tunes parameters by minimizing the AIC, facilitating adjustment to series with irregular seasonal behavior [29]. In a general way, it handles multiple seasonal cycles and variance stabilization, useful for detecting latent cyclical patterns in alternative assets.

The three models shown are very similar, but there are some differences in how they are used for financial series:

- STLF stands for “seasonal and trend-following strategy.”
- It is a trading strategy that focuses on recurring patterns. Examples of these patterns include Christmas sales and quarter-end closings.
- TBATS stands for “time-based autocorrelation of seasonal time series.” It is used when the series is very noisy and seasonality changes a bit over time.
- ETS: This is used if the behavior is more like a “random walk,” which is typical of stocks.

3.3.4. ARIMA/AUTO.ARIMA

The `auto.arima()` function implements the Hyndman–Khandakar algorithm, which combines unit root tests (KPSS), likelihood maximization, and stepwise search to automatically select the optimal order (p, d, q) of an ARIMA model. It first determines the number of differences required using repeated KPSS tests, then explores model candidates by varying p and q in a stepwise manner, and chooses the one that minimizes the AICc. This approach automates much of the Box–Jenkins process and allows ARIMA to be applied in a practical way to varied datasets [30].

ARIMA (`auto.arima`) provides the parsimonious linear baseline; its automatic selection algorithm ensures that the model order is data-driven and avoids overfitting, making it an essential benchmark for evaluating whether more complex models offer meaningful gains. Financial analysts widely regard ARIMA as the minimum acceptable standard for return forecasting, and its inclusion ensures methodological rigor [25].

3.3.5. NNETAR

The `nnetar()` function is employed to train feed-forward neural networks (FFNNs) with a single hidden layer, utilizing lags from the series as inputs. By default, it automatically determines the optimal number of lags (p) based on the Akaike information criterion (AIC) in a linear autoregressive (AR) model. It then trains multiple networks with different initial weights, averaging their forecasts to improve stability. In the context of seasonal series, the model encompasses both non-seasonal $(1, \dots, p)$ and seasonal $(m, 2m, \dots, Pm)$ lags, thereby yielding the $\text{NNAR}(p, P, k)_m$ model [25].

This FFNN is characterized by its capacity for non-linear modeling without the necessity of the explicit specification of the functional form of non-linearity. Financial returns have been observed to demonstrate regime-dependent behavior, such as the distinction between bull and bear markets, as well as between high- and low-volatility regimes. Financial returns have also been shown to exhibit threshold effects and asymmetric responses to positive and negative shocks. Neural networks offer a malleable approximation framework for these intricate dynamics. The autoregressive structure with ensemble averaging (training multiple networks and averaging forecasts) mitigates the risk of overfitting [25].

3.3.6. ARFIMA

Autoregressive Fractionally Integrated Moving Average (ARFIMA) models are a generalization of Autoregressive Integrated Moving Average (ARIMA) models, as they allow for a fractional degree of differentiation (real d). This captures the “long memory” property, where the autocorrelation function decays slowly according to a power law. The parameter d governs the level of persistence: (a) The series is stationary if and only if d is less than 0.5. If $0.5 \leq d < 1$, the long-term trend is retained, and (b) in the event of the value of d being less than -0.5 and greater than 0, the property of anti-persistence is exhibited.

The identification and estimation of ARFIMA models can be achieved through the utilization of parametric and semiparametric methodologies in the identification procedure for stationary series. Within the R programming environment, the ARIMA package is employed to fit these models through the direct optimization of the exact likelihood. The package provides functionality for both simulation and forecasting, incorporating ARMA errors and fractional components. ARFIMA models fractional integration to capture persistent autocorrelation in returns or volatility [31].

3.3.7. The hybridModel Function

The `hybridModel()` function, implemented in the `forecastHybrid` package, generates an assembly comprising two to five base models (namely, `auto.arima`, `ETS`, `thetam`, `nnetar`, `stlm`, `tbats` and `snaive`). This ensemble baseline combines multiple forecasts with equal or error-based weights, providing a benchmark for evaluating whether metaheuristic weight optimization (via Threshold Accepting) offers meaningful gains [18].

The model is fitted independently, with either the default or custom hyperparameters. The weights for each forecast are calculated using error criteria (RMSE, MAE, or MASE) on in-sample fits or through cross-validation. The process of combining predictions involves a two-step procedure. Firstly, the predictions are weighted according to the magnitude of their respective fitting errors. Secondly, the predictions are weighed equally. This hybrid approach capitalizes on the strengths inherent in each methodology, thereby enhancing the overall accuracy of forecasts.

3.4. Optimized Assembly Strategy

The proposed forecasting framework, designated as the Time-series Adaptive Forecast Ensemble, integrates eight individual forecasting models (ETS, ARIMA, STL, NNAR, TBATS, ARFIMA, Prophet, and Random Forest) into an optimized weighted combination through a simulated annealing algorithm. For each asset, the available weekly price series was partitioned into a training set and a hold-out test set, the latter comprising approximately 24 weeks. All base models were fitted in a single pass on the training set, and their in-sample residuals were used to construct a pseudo-validation window representing the most recent 17% of the training observations. Prior to weight optimization, the individual forecasting accuracy of each model was assessed on this pseudo-validation window using the Symmetric Mean Absolute Percentage Error (SMAPE) as the loss function. Only the top- $K = 5$ models ranked by the lowest pseudo-validation SMAPE were admitted

into the ensemble. The SA algorithm then sought the optimal weight vector across the selected models through 30 independent parallel restarts, with each restart executing 300 inner iterations. The initial and final temperatures were scaled relative to the pseudo-validation SMAPE to ensure consistent exploration across assets of varying price scales. The selection of optimal weights from the restart yielded the lowest SMAPE globally. To generate accurate out-of-sample forecasts, all models were subsequently re-estimated on the complete available series (train + test) and projected 26 weeks ahead. The resulting ensemble forecast was evaluated against a dedicated hold-out dataset of realized prices. The accuracy of the forecasts was assessed using several statistical metrics, including the SMAPE, RMSE, MAE, and MAPE. Additionally, the statistical significance of the differences in predictive accuracy was examined through the Diebold–Mariano (DM) test, with Harvey–Leybourne–Newbold (HLN) small-sample correction. The process described is illustrated in Figure 1 and includes the parameter values used for the experimentation.

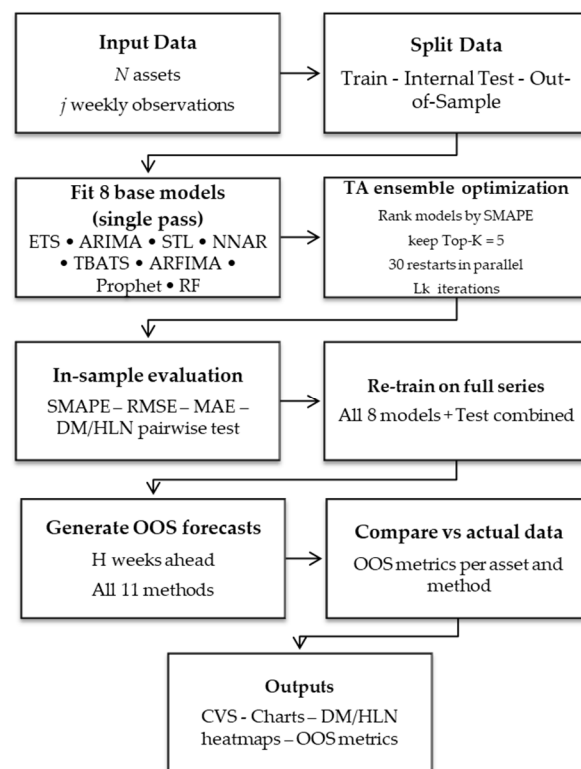


Figure 1. General outline for forecasting and ensemble optimization. Own elaboration.

Statistical Test (DM and HLN)

The differences in how accurate the forecasts are important; we use a test created by Diebold and Mariano (1995) [32]. This nonparametric test lets us compare how well two competing models can predict future outcomes under the assumption that they are equally accurate, i.e., $E[d_t] = 0$, where d_t is the difference in losses. However, the original DM test often overdiagnoses significance in small samples or in forecast horizons greater than one step ($h > 1$). This analysis is supported by the modification of Harvey, Leybourne, and Newbold (1997) [33]. The HLN statistic adjusts the bias of the DM test by correcting sample size and using a Student’s t distribution to calculate critical values. When we combine these two tests, we get a strong system for checking how well one method works compared to another. This makes sure that the results we see are not just random noise or due to problems with the samples.

3.5. Portfolio Optimization

The implementation of a simulated annealing algorithm, predicated on projected expected returns, is undertaken for the purpose of selecting the optimal subset of assets and assigning weights to them. The objective function is designed to maximize a modified Sharpe ratio, incorporating penalties for correlation between assets and weight dispersion. Portfolio performance is evaluated over a test horizon using metrics such as expected return, standard deviation (volatility), and Sharpe ratio.

4. Experimentation and Results

This section reviews the equipment and material resources used for developing this study. For this work, was used an Intel(R) Core (TM) i7-13620H, CPU 2.40 GHz, RAM: 16 GB. The languages used for this experiment are Python 3.7 for initial and optimized portfolios and R 4.3.3 for forecasting.

A description of the time series utilized and an account of the experimentation conducted with this series through the portfolio integration and portfolio forecasting algorithms are also provided. A concise overview of both is included. The results and discussion of these experiments are subsequently presented.

4.1. Dataset

To evaluate the algorithms and the proposed solution methodology, a set of assets and their data series of the Mexican Stock Exchange was used, consisting of 51 ETFs and FIBRAs listed between January 2020 and the end of February 2026, listed in Table 1, giving a total of 322 observations (weekly prices). For the purposes of this study, weekly returns were utilized instead of daily or monthly returns. This decision is based on the need to balance the richness of the information with the robustness of the statistical analysis. In emerging capital markets, the use of daily data often introduces significant “microstructure noise” due to low liquidity and the phenomenon of asynchronous trading. This can distort estimates of correlation and variance [34].

Table 1. List of ETFs and FIBRAs.

XLFMX	IGFMX	VGKMX	EFA.MX	IVWMX	IAU.MX
SOXX.MX	AIQ.MX	SOXS.MX	EPU.MX	EWZ.MX	CORPMX
SDIAN.MX	PSOTRAC15.MX	SMARTRC14.MX	CETETRCISHRS.MX	MEXX.MX	NAFTRACISHRS.MX
IBTAN.MX	EMBXXN.MX	IUAAN.MX	EWWMX	QVGMEX18.MX	VMEX19.MX
MEXTRAC09.MX	IVVPESOISHRS.MX	GENIUS21.MX	FIBRATC14.MX	FLMX.MX	ESGMEXISHRS.MX
DLRTRAC15.MX	DIABLO10.MX	CHNTRAC11.MX	ANGELD10.MX	STORAGE18.MX	EDUCA18.MX
FCFE18.MX	FHIPO14.MX	SOMA21.MX	TERRA13.MX	FUNO11.MX	FSHOP13.MX
FPLUS16.MX	FMTY14.MX	FINN13.MX	FIHO12.MX	FIBRAPL14.MX	FIBRAMQ12.MX
FIBRAHD15.MX	DANHOS13.MX	FNOVA17.MX			

To ensure a rigorous evaluation, the total dataset (2020–2026) was partitioned chronologically into two main distinct windows: a Development Window (2020–2025, representing 92% of the data) and a final evaluation window (2025–2026, representing 8%); see Figure 2.

Within the designated Development Window, standard three-way partitioning is employed for training, hyperparameter tuning (referred to as the “Validation Set”), and the initial unbiased performance estimate (“Internal Test Set”).

The evaluation window is reserved exclusively for a single, final simulation of real-world deployment, which is referred to throughout this manuscript as the “Out-of-Sample Final Test.” This set was not visible to the models during the training or synchronization phases, thereby preventing any temporal data leakage. All references to “hold-out” or generic “test sets” were harmonized to adhere to this defined structure.

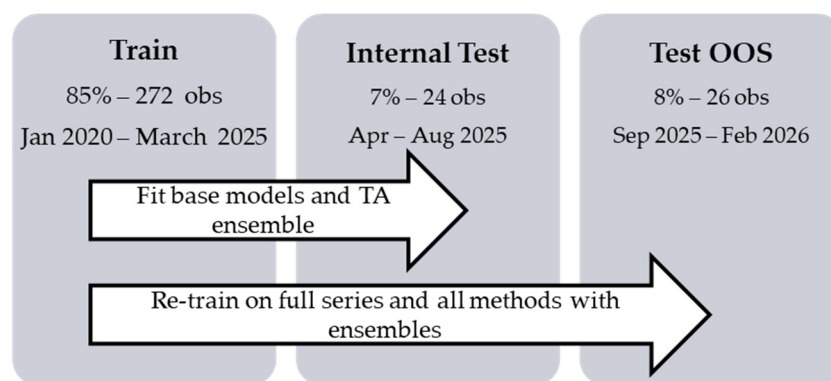


Figure 2. Data partitioning.

The 85% training allocation makes sure that the optimization algorithms and forecasting models have enough historical data to learn complex patterns and volatility patterns. This is particularly important because the financial series analyzed are stochastic.

The 7% internal test segment is the final step in checking whether the model can be used to make predictions. The three-stage partitioning reduces the risk of overfitting during the hyperparameter tuning phase. This ensures that the results in the 26-week period accurately reflect the model's predictive performance. The 26-week out-of-sample period (8% of the data) is chosen to match standard industrial and financial planning cycles. A six-month timeframe is enough time to see how well the models can predict medium-term trends and seasonal shifts. This is important for making practical decisions.

During the data cleaning process, the set of time series was analyzed to verify their quality. After thorough analysis, time series with less than 70% of the available data were excluded, leaving 38 time series for further consideration. For cases where data was missing, the proxy filling method described in Section 3.1 was applied.

4.2. Initial Portfolio

The SAIPO algorithm is employed in the construction of the initial portfolio. This has previously undergone rigorous testing and comparative analysis with other available options, thereby substantiating its superior performance and yielding optimal results.

The hyperparameters of the SAIPO metaheuristic were tuned by applying the analytical tuning method described in [9,24]. The values obtained for the parameters are an initial temperature of 0.944879, final temperature of 0.000074, and cooling ratio of 0.93, while the equilibrium cycle length (L) is dynamically calculated as a function of the number of assets and the temperature scheme; in most cases, the value is around 100. After 30 executions, the initial portfolio is complete, and the SR's behavior can be observed in Figure 3. The results indicate that when a risk-free rate of 0% is employed, the average return is 0.0017, the risk is 0.0169, and the Sharpe ratio is 0.1043. With 21–29 assets selected on a weekly basis, the following observations were made: an annual expected return of 0.0964 and an annual Sharpe ratio of 5.67.

Figure 2 illustrates the convergence dynamics of the SAIPO algorithm over the course of the evaluations. An initial phase of intense stochastic exploration is observed, characterized by high variance in the Sharpe ratio. This indicates the system's ability to escape local optima by accepting suboptimal solutions under high computational temperature. As the cooling process progresses, the algorithm transitions to a neighborhood exploitation stage, stabilizing the portfolio composition until reaching asymptotic convergence at an approximate Sharpe ratio. This outcome confirms the effectiveness of the metaheuristic approach in navigating complex search spaces within the ETF/FIBRA market.

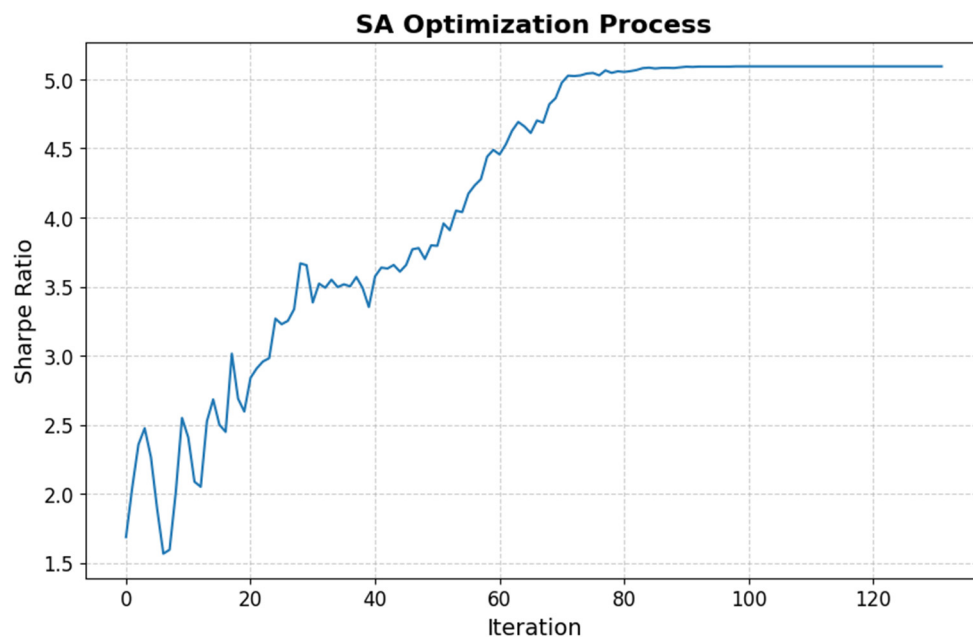


Figure 3. SAIPO algorithm solution for portfolio selection for weekly values.

4.3. Forecast and Ensemble

The hyperparameters of the TA metaheuristic were tuned by applying the analytical tuning method mentioned before. The values obtained for the parameters are an initial temperature of 250, final temperature of 0.00016, cooling ratio of 0.95, and a Markov length of 100. However, during the experimentation process, it was observed that while these hyperparameters worked well and produced satisfactory results for several of the analyzed time series, many of the forecasts were completely off the mark. Therefore, it was decided to use a custom tuning scheme for each time series. The solution entailed the utilization of the error calculated in the validation sub-stage—specifically, SMAPE (*smape_pv*)—as the base value and a predetermined factor to obtain the initial temperature. Similarly, for the final temperature, a small factor close to zero was estimated, such that the tuned and designated hyperparameters were as follows: the initial temperature is equivalent to the product of *smape_pv* and 50, the final temperature is equivalent to the product of *smape_pv* and 0.0001, the cooling ratio is equivalent to 0.985, and the Markov length is set to 300. The acceptance ratio is equivalent to 0.979.

Figure 4 offers a comparison of forecasting methods. The dotted lines represent the various forecasting methods employed. The reference displayed serves as an illustration of how certain methods are more effective in aligning with the data presented in the test section, while others tend to overestimate or underestimate the projected data, thereby expanding the margin of error. Conversely, the assembled methods are represented by solid lines. It is evident that the three methods adhere to the pattern of and trend in the test data.

When these methods are assembled, they generate forecasts for a particular series. In this case, it is evident that the equally weighted method and the quick model do not adequately follow the actual data, while the TA-optimized method demonstrates better performance and test data tracking.

As illustrated in Figure 5, the optimized ensemble forecast aligns closely with the actual test data, indicating the model's capacity to replicate the movement patterns of the original data.

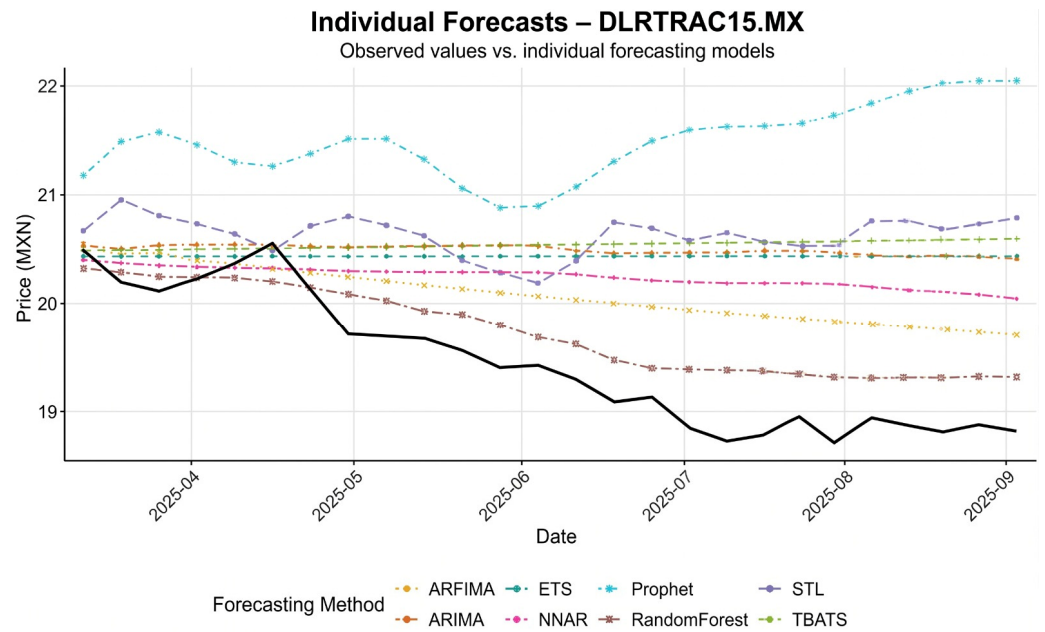


Figure 4. The performance of individual forecasting models for an asset.

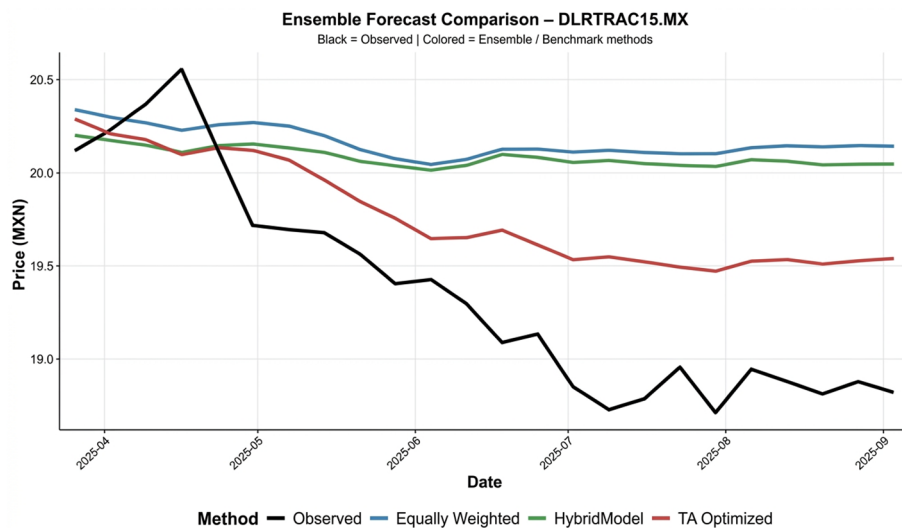


Figure 5. Combined forecast model performance for an asset.

Figure 6 presents the SMAPE distribution for all eleven forecasting methods across the asset series, ordered from the lowest to highest mean error. The diamond symbol represents the cross-sectional mean and the horizontal bar the median, while individual dots correspond to each asset.

Three performance tiers are visually distinguishable. The first tier, to the left of the first dashed line, comprises Comb_TA, TBATS, RF, ETS, ARIMA, and Comb_EW, all with a mean SMAPE below 0.076. Comb_TA leads this group with the lowest mean (0.0703) and notably the most compact interquartile range, indicating not only superior average accuracy but also greater consistency across assets—a desirable property in heterogeneous financial panels. The second tier, between the two dashed lines, includes HybridModel and STL, whose means and medians are slightly elevated but whose distributions remain relatively contained. The third tier, to the right of the second dashed line, groups Prophet, ARFIMA, and NNAR, which exhibit both higher central tendency and substantially wider dispersion. NNAR presents the most extreme outlier in the panel (~0.67 SMAPE), reflecting its instability on certain series with irregular dynamics.

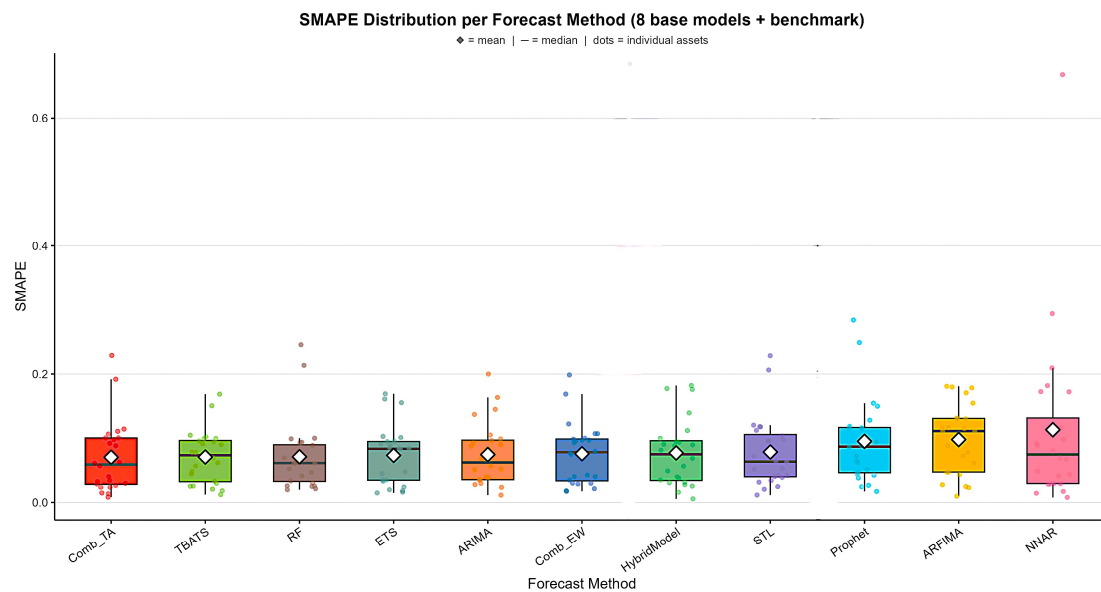


Figure 6. Overall sMAPE distribution for individuals and combined forecast methods in the dataset.

A notable feature shared by all methods is the right-skewed distribution, reflecting that most assets are forecast with moderate error, while a small subset—likely those with structural breaks or high volatility—drives the upper tail. The relatively tight distribution of Comb_TA suggests that the top-K pre-selection and SA weight optimization effectively mitigate the influence of poorly performing base models on individual assets, reducing the occurrence of high-error episodes compared to both the equal-weight benchmark (Comb_EW) and several individual methods.

Statistical Test

Pairwise forecast accuracy was assessed using the Diebold–Mariano (DM) test and its Harvey–Leybourne–Newbold (HLN) small-sample correction over 24 weekly asset series. Both tests yielded consistent conclusions in terms of direction and significance level, confirming the robustness of the results to the limited hold-out window (~24 observations per series).

Figure 7 presents the average DM statistic for each method pair, where blue cells indicate that the row method outperforms the column method, and red cells indicate the opposite; significance stars denote differences at the 1%, 5%, and 10% levels. Three structural patterns emerge from the heatmap. First, ARFIMA and Prophet display predominantly red rows, indicating that both methods are systematically outperformed across the panel—ARFIMA by seven competitors (net score = -7) and Prophet by five (net score = -5)—consistent with their elevated mean SMAPE values of 0.098 and 0.095, respectively. Second, ETS, ARIMA, TBATS, and HybridModel show predominantly blue rows, reflecting competitive accuracy relative to most counterparts, with net scores ranging from $+3$ to $+4$. Third, and most relevant to the proposed framework, the Comb_TA row is predominantly light-blue to neutral, indicating that the TA ensemble does not significantly underperform any major competitor. Its only statistically significant loss was against TBATS (DM = 0.25, $p = 0.027$), a method specifically designed for complex seasonal structures. Critically, Comb_TA significantly outperformed Comb_EW (DM = -1.30 , $p = 0.021$), providing statistical support for the value added by the SA weight optimization over naive equal weighting. These findings, together with Comb_TA achieving the lowest mean SMAPE across the panel (0.0703), suggest that the proposed ensemble constitutes a competitive and statistically defensible forecasting strategy for weekly Mexican ETF and FIBRA price series.

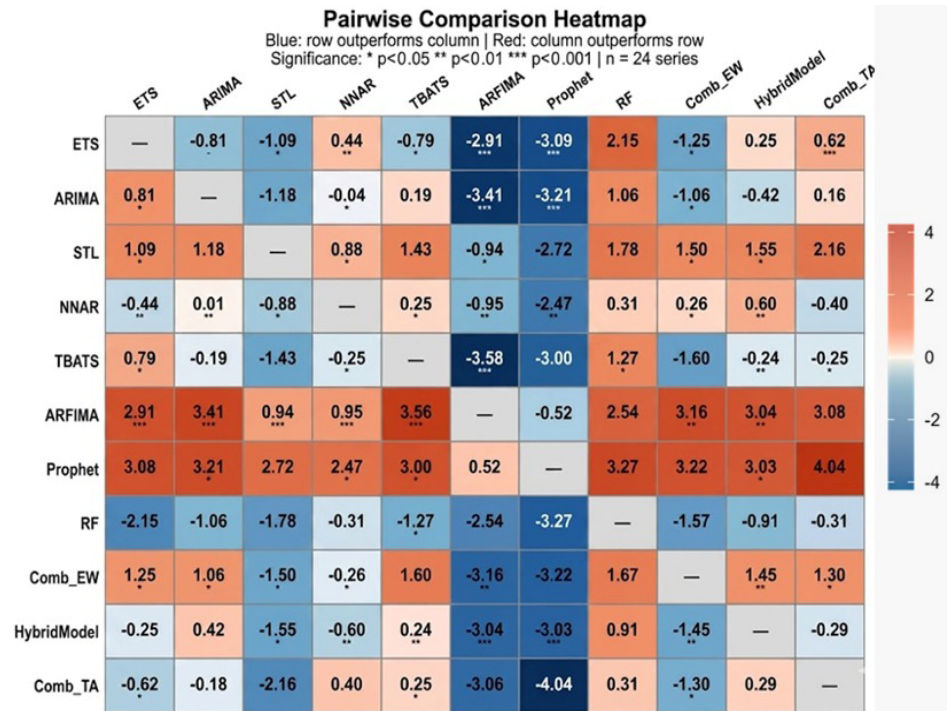


Figure 7. Average DM statistic.

Figure 8 displays the average *p*-values from the Harvey–Leybourne–Newbold (HLN) test, which corrects the standard DM statistic for small-sample bias using a $t(T - 1)$ distribution. Deep red cells indicate statistically significant differences in predictive accuracy, while lighter tones denote non-significant pairs. Given that the effective hold-out window comprises approximately 24 weekly observations per series, the HLN correction is particularly appropriate in this context, as the standard DM statistic is known to over-reject the null hypothesis of equal predictive accuracy in small samples.

The overall pattern of significance is largely consistent with the DM results, confirming that the conclusions are not an artifact of sample size. The most prominent feature of the heatmap is the deep red column associated with ARFIMA, which registers highly significant differences ($p < 0.001$) against ETS, ARIMA, and TBATS and significant differences ($p < 0.05$) against NNAR, STL, Comb_EW, and HybridModel. This pervasive significance confirms that ARFIMA is the systematically weakest method in the panel. Similarly, the ARFIMA row shows that it is significantly outperformed by virtually all competitors, reinforcing the evidence from the DM statistical analysis.

Regarding the proposed ensemble, Comb_TA presents a notably sparse pattern of significant cells. Under the HLN correction, its only statistically significant pairwise results are a superiority over ETS ($p = 0.038$) and Comb_EW ($p = 0.022$) and a significant loss against TBATS ($p = 0.028$). The remaining ten pairwise comparisons yield *p*-values above 0.05, indicating that Comb_TA performs statistically on par with ARIMA, NNAR, RF, STL, Prophet, HybridModel, and ARFIMA under the more conservative small-sample test. The significant advantage over Comb_EW is particularly relevant, as it provides corrected statistical evidence that the TA-optimized weighting scheme adds measurable value beyond naive averaging even under conservative inferential conditions. Taken together, the HLN results corroborate the DM findings and support the conclusion that Comb_TA represents a statistically competitive ensemble strategy for weekly financial series forecasting.

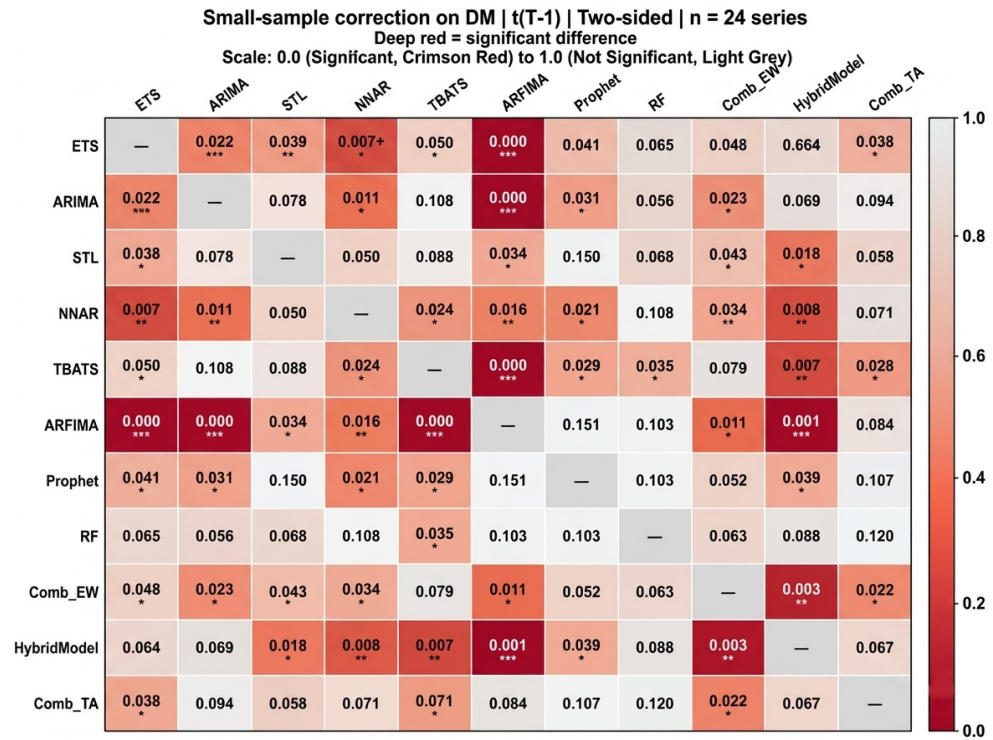


Figure 8. Average *p*-values from the Harvey–Leybourne–Newbold (HLN) test. Note: * *p* < 0.05: Significance at 95% (there is strong evidence), ** *p* < 0.01: 99% significance (very solid evidence), *** *p* < 0.001: Significance at 99.9%. (extremely strong evidence), and without *: The difference is not statistically significant.

In Table 2, the proposed method, Comb_TA, demonstrates notable leadership with an SMAPE of 0.0703 and the lowest standard deviation among the competing methods (0.0559), thereby underscoring its consistency across assets. The sole vulnerability in the DM count is its paucity of significant wins, with only two recorded victories, in contrast to the more substantial tallies of ETS and HybridModel. This observation lends credence to the notion that Comb_TA’s advantage, while genuine, is modest in magnitude.

Table 2. In-sample forecast accuracy and DM pairwise significance.

Method	Rank	SMAPE Mean	SMAPE Median	SMAPE Sd	RMSE	MAE	DM Wins	DM Losses	Net
Comb_TA	1	0.0703	0.0589	0.0559	28.47	24.47	2	1	1
TBATS	2	0.0707	0.0734	0.0412	25.97	22.78	5	2	3
RF	3	0.071	0.0613	0.0553	36.22	30.74	1	0	1
ETS	4	0.0732	0.0834	0.0458	26.74	23.24	6	2	4
ARIMA	5	0.0745	0.0621	0.0488	26.24	22.99	4	1	3
Comb_EW	6	0.0758	0.0781	0.0475	29.29	25.2	3	4	−1
HybridModel	7	0.077	0.0749	0.0518	30.68	26.72	5	1	4
STL	8	0.0784	0.0634	0.0547	21.17	17.62	1	4	−3
Prophet	9	0.0951	0.0868	0.066	37.25	31.69	0	5	−5
ARFIMA	10	0.0978	0.1111	0.0541	41.9	37.05	0	7	−7
NNAR	11	0.1132	0.0746	0.1384	72.12	59.47	4	4	0

Note: 24 assets ~24-week test set ordered by mean SMAPE; DM: significant wins/losses at *p* < 0.05.

4.4. Final Portfolio Optimization

Once the forecasting process is completed, we take the forecasted data from the series and proceed to carry out the final optimization of the portfolio using the algorithm again. At this point, the same hyperparameters of the previous stage of the simulated annealing

are maintained, with the only difference being the change in the equilibrium cycle; for greater opportunity in the solution space, the equilibrium cycle length L is set to 300.

The results show that by establishing a minimum rate of return of 6% (the average annual yield on 28-day government bonds between 2020 and February 2026 ranged from approximately 4% to 7.5%, with an average of approximately 6%), an annual expected return of 30.08% is obtained, an average risk of 8.08%, a correlation of -0.0488 , and a Sharpe ratio of 2.98, leaving on average 9.5 assets in the portfolio.

The portfolio's strategic composition, which includes MXN-listed ETFs, Mexican FIBRAs, thematic ETFs (ESG, innovation, semiconductors), and defensive ETFs, is a key factor in these results. This approach helps to stabilize volatility, but it does result in a reduction in the maximum potential return.

U.S. ETFs (SOXX, XLF, EFA) demonstrated notable success, particularly SOXX, which profited significantly from the semiconductor boom (2020–2025). FIBRAs offered stability and dividends, but their growth potential was comparatively limited when benchmarked against the Mexican Price and Stock Index (IPC). The inclusion of defensive assets (DIA-BLOI, SDIAN, ANGELD) has been shown to reduce returns in bull markets, although they provide protection during downturns.

In terms of volatility, it was lower than that of the IPC, thanks to the mix of defensive and real estate assets.

The IPC exhibited a volatility of 18% during this period; our portfolio is well below this threshold, indicating a lower risk profile compared to the Mexican market as a whole.

Conversely, FIBRAs have consistently generated dividends, which enhance total returns when reinvested. However, this aspect is not addressed in the scope of this study.

4.5. Out-of-Sample Forecasting

To further validate this methodology, out-of-sample (OOS) forecasts were generated based on the training process from the previous stage for the individual methods and on the optimized weighting generated by the TA algorithm. The forecasts demonstrate clear dispersion and illustrate how models such as NNAR or ARIMA can exhibit erratic or flat behavior in the presence of trend breaks. Conversely, in the ensemble forecasts, the visual discernibility of the role of the combination (particularly Comb_TA) as a noise filter is pronounced, resulting in a smoothing of the error and the maintenance of a trajectory that aligns more closely with the actual series.

As shown in Figure 9, the evaluation of OOS forecasting constitutes the definitive test for validating the generalization ability of the proposed models. While in-sample fitting entails the possibility of overfitting historical noise, out-of-sample analysis quantifies the robustness of the ensemble under unobserved market conditions. The present study demonstrates the significance of the OOS analysis in showing that optimization via TA, in Figure 10, not only minimizes historical error but also captures underlying structures that persist over time. As demonstrated in the comparisons of the FIBRA and ETF series, the ensemble reduces the extreme variance in the individual models, providing a more stable and statistically reliable estimate for financial decision-making in scenarios of uncertainty.

In Table 3, the Δ vs. IS column is the most informative. Regarding parsimony-based methods, the ARIMA method demonstrated a 9.2% decrease, the STL method demonstrated a 10.0% decrease, and the HybridModel method demonstrated a 6.1% decrease. Conversely, methods that optimize over the historical period, such as the RF method (37.8%) and the Comb_TA method (25.7%), performed less well. The Spearman correlation coefficient, denoted by ρ , was found to be 0.236, thereby confirming that the rankings underwent substantial changes between the two periods. This finding serves to substantiate the decision to present both tables within the context of this paper.

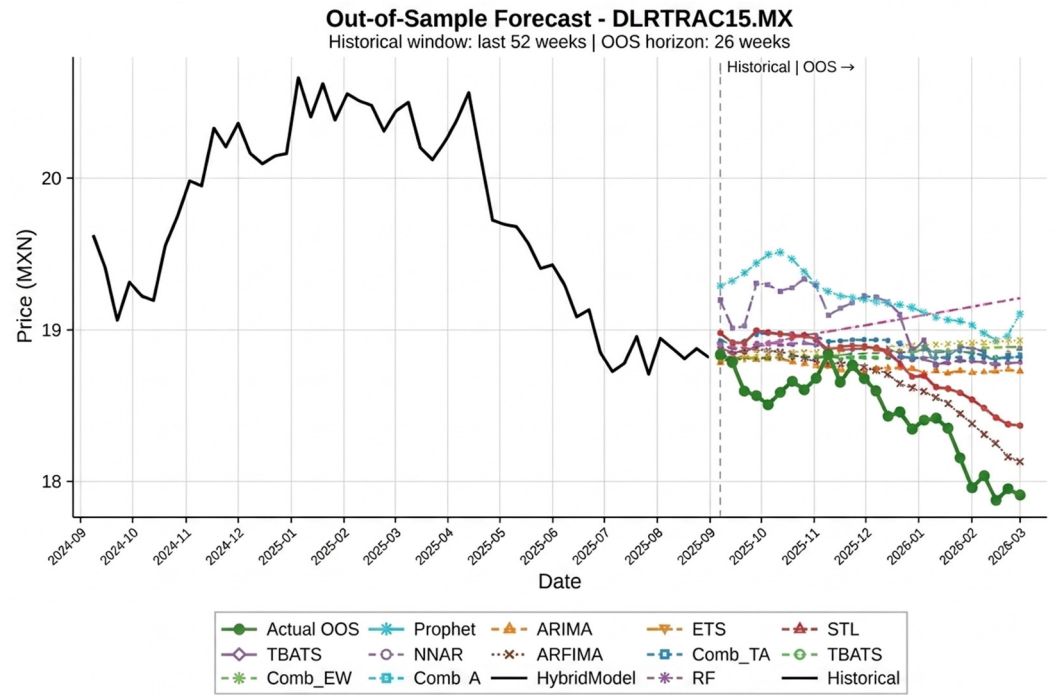


Figure 9. This example illustrates OOS forecasts generated by individual methods.

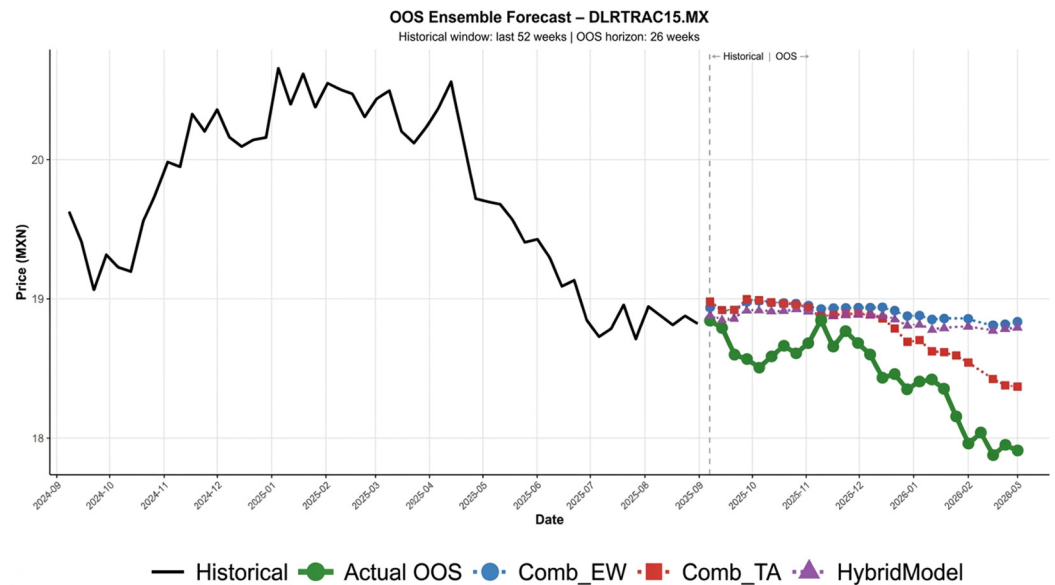


Figure 10. This example illustrates OOS forecasts generated by ensemble methods.

Table 4 presents the out-of-sample (OOS) SMAPE for the 26-week forecast horizon covering September 2025 through March 2026. The OOS ranking differs substantially from its in-sample counterpart, with a Spearman rank correlation of only 0.236 between the two periods, indicating a marked change in the relative performance of methods across the two regimes. ARIMA and STL emerge as the best-performing methods in the OOS period, improving their mean SMAPE by 9.2% and 10.0%, respectively, relative to their in-sample values, consistent with the well-documented advantage of parsimonious linear models in out-of-sample extrapolation under structural change. HybridModel and NNAR also generalize well, registering OOS improvements of 6.1% and 18.0%. By contrast, RF and Comb_TA exhibit the largest OOS degradation, with mean SMAPE increases of 37.8% and 25.7%, respectively, a pattern attributable to the overfitting of model-specific dynamics present in the training window but absent in the forecast horizon.

Table 3. Out-of-sample forecast accuracy—26-week horizon (Sep 2025–Mar 2026).

Method	IS Rank	OOS Rank	SMAPE Mean	SMAPE Median	SMAPE Sd	RMSE	MAE	MAPE	Δ IS \rightarrow OOS
ARIMA	5	1	0.0676	0.0637	0.0533	61.77	52.23	0.0643	−9.2%
STL	8	2	0.0706	0.0615	0.0466	56.08	49.11	0.0677	−10.0%
Hybrid Model	7	3	0.0723	0.0624	0.0537	62.05	53	0.0684	−6.1%
ETS	4	4	0.0744	0.0683	0.0516	61.6	52.12	0.0702	1.60%
TBATS	2	5	0.0747	0.0614	0.0506	59.84	50.56	0.0709	5.70%
Comb_EW	6	6	0.0817	0.0662	0.0594	66.47	58.19	0.0771	7.80%
Comb_TA ★	1	7	0.0884	0.0679	0.0642	71.18	62.63	0.0844	25.70%
Prophet	9	8	0.0921	0.0562	0.0758	62.93	57.4	0.0926	−3.1%
NNAR	11	9	0.0928	0.0846	0.0711	74.84	64.84	0.0861	−18.0%
RF	3	10	0.0979	0.0843	0.0789	83.13	74.26	0.0909	37.80%
ARFIMA	10	11	0.1246	0.105	0.0807	92.3	81.07	0.1128	27.30%

Note: 24 assets ordered by mean SMAPE (OOS); Δ = relative change from in-sample mean SMAPE; highlighted row = proposed ensemble; ★ Comb_TA: IS rank 1^o (SMAPE = 0.0703); OOS rank 7^o (SMAPE = 0.0884); Spearman ρ (IS, OOS) = 0.236.

Table 4. Out-of-sample forecast accuracy (26-week horizon, Sep 2025–Mar 2026).

Method	IS Rank	OOS Rank	SMAPE Mean	SMAPE Median	SMAPE Sd	RMSE	MAE	Δ vs. IS
ARIMA	5	1	0.0676 *	0.0637	0.0533	61.77	52.23	−9.2%
STL	8	2	0.0706	0.0615	0.0466	56.08	49.11	−10.0%
HybridModel	7	3	0.0723	0.0624	0.0537	62.05	53	−6.1%
ETS	4	4	0.0744	0.0683	0.0516	61.6	52.12	1.60%
TBATS	2	5	0.0747	0.0614	0.0506	59.84	50.56	5.70%
Comb_EW	6	6	0.0817	0.0662	0.0594	66.47	58.19	7.80%
Comb_TA	1	7	0.0884	0.0679	0.0642	71.18	62.63	25.70%
Prophet	9	8	0.0921	0.0562	0.0758	62.93	57.4	−3.1%
NNAR	11	9	0.0928	0.0846	0.0711	74.84	64.84	−18.0%
RF	3	10	0.0979	0.0843	0.0789	83.13	74.26	37.80%
ARFIMA	10	11	0.1246	0.105	0.0807	92.3	81.07	27.30%

Note: 24 assets ordered by mean SMAPE; Δ vs. IS = relative change in mean SMAPE from in-sample to OOS period; * = lowest error; Spearman rank correlation IS vs. OOS: ρ = 0.236; Comb_TA highlighted in both tables.

The OOS deterioration of Comb_TA is concentrated in five assets (EPU.MX, XLF.MX, FHIPO14.MX, FIBRAMQ12.MX, and GENIUS21.MX) where IS-to-OOS SMAPE increases exceed 170%, while the remaining 19 series exhibit a median degradation of only +21.7%, comparable to that of Comb_EW (+7.8%) and other competitive methods. This asymmetry suggests that SA weight optimization successfully captured the dominant forecasting dynamics for the majority of the panel but proved sensitive to regime shifts in a subset of assets characterized by abrupt structural breaks during the OOS window, a limitation inherent to any optimization-based ensemble that relies on fixed historical weights.

These findings are consistent with the bias–variance tradeoff literature on forecast combination: while adaptive weighting schemes can reduce in-sample error by concentrating weight on the locally best-performing models, they may increase out-of-sample variance when the relative accuracy of base models shifts across periods [35,36]. The equal-weight benchmark Comb_EW, which is immune to this source of instability, achieves a more stable IS-to-OOS transition (+7.8%), reinforcing the argument for combining optimization-based and equal-weight ensembles as a robustness strategy in financial forecasting applications.

4.6. Portfolio Final Optimization in OOS and Portfolio with Actual Asset Values

Finally, the integrated portfolio was evaluated by forecasting out-of-sample data, specifically for the period from September 2025 to the end of February 2026. The algorithm’s

hyperparameters and the risk-free rate were consistent with those utilized during the evaluation of the test period.

The findings suggest a projected return of 14.59%, an average risk level of 1.56%, a correlation coefficient of -0.121 , and a Sharpe ratio of 5.49, leading to an average asset allocation of 5 within the portfolio. Additionally, the portfolio generated with the real data is presented, that is, with the data from January 2020 to February 2026, which shows the following results in Table 5.

Table 5. Final portfolio scenarios.

Scenario	Return	Risk	Correlation	Sharpe	Avg. Assets
Initial	9.64%	1.69%	-0.1109	5.67	23.0
Test	30.08%	8.08%	-0.0488	2.98	12.7
OOS (forecast)	35.77%	9.53%	0.0218	3.12	10.8
Real (2020–2026)	33.37%	8.9%	0.0052	4.4	17.8

A central finding is that both the test and OOS portfolios yield expected returns that meet or exceed the established minimum threshold of 6% annual return on 28-day government bonds.

4.7. Discussion

- i. **Risk–Return Tradeoff and Evolution.** A substantial shift in the risk–return profile is evident as the initial portfolio is compared to the optimization scenarios. The initial scenario indicates a conservative return of 9.64%, accompanied by a very low risk of 1.69%. In contrast, the test and OOS (forecast) scenarios demonstrate a substantial increase in profitability, reaching 30.08% and 35.77%, respectively. Although the risk increases proportionally to approximately 8.08–9.53%, the transition suggests that the optimization algorithm (Threshold Accepting) successfully identified higher-yield opportunities that were previously untapped in the initial allocation.
- ii. **Efficiency and Sharpe Ratio Analysis.** The analysis of the Sharpe ratio reveals critical insights into the efficiency of the portfolios, especially considering the change in the risk-free rate (R_f).
 Initial vs. Test: While the initial portfolio demonstrates a high Sharpe ratio of 5.67 (evaluated at $R_f = 0$), the test scenario exhibits a robust 2.98.
 OOS and Real Scenarios: Despite the implementation of a more stringent 6% annual R_f hurdle for the OOS and real scenarios, the portfolio’s efficiency persists at a remarkably high level. The real (2020–2026) scenario attained a Sharpe ratio of 4.4, signifying that for each unit of risk assumed, the portfolio yielded substantial excess returns, even within a higher-interest-rate environment.
- iii. **Predictive Accuracy (OOS vs. Real).** A salient finding is the alignment between the out-of-sample (OOS) forecast and the real data performance. The OOS forecast predicted a return of 35.77%, with a 9.53% risk. The real performance yielded a 33.37% return with an 8.9% risk. The proximity of these figures to one another indicates a high degree of predictive accuracy in the underlying forecasting models. The model exhibited a slight overestimation of the return, with a variance of 2.4%. However, it also demonstrated an accurate capture of the risk level, thereby substantiating the robustness of the hybrid forecasting approach for real-world financial applications.
- iv. **Portfolio Concentration and Diversification.** A discernible trend toward asset concentration emerges during the optimization process. The initial portfolio exhibited high diversification, with 23.0 assets, whereas the OOS scenario optimized this down to 10.8 assets. The real portfolio distribution (average of 17.8 assets) suggests that while the forecast leans toward high concentration to maximize returns, the actual market

dynamics over the 2020–2026 period required a slightly broader base to maintain the observed stability and a higher Sharpe ratio (4.4 vs. 3.12).

5. Conclusions

The present study evaluated the forecasting performance of eight individual models and three ensemble strategies (equal-weight combination (Comb_EW), HybridModel, and a simulated annealing-optimized weighted ensemble (Comb_TA)) applied to 24 weekly price series of Mexican ETFs and FIBRAs over the period 2020–2025, with a 26-week out-of-sample horizon extending through March 2026.

The in-sample results demonstrate that the proposed Comb_TA ensemble attained the lowest mean SMAPE across the panel (0.0703), surpassing all individual models and the equal-weight benchmark. The Diebold–Mariano test with Harvey–Leybourne–Newbold small-sample correction provided statistical support for this superiority, rejecting the null hypothesis of equal predictive accuracy against Comb_EW at the 5% significance level ($DM = -1.30, p = 0.021$) while registering no significant losses against the majority of individual competitors. These results underscore two salient advantages of the combination approach. Firstly, the utilization of ensemble methods was demonstrated to systematically reduce the dispersion of forecast errors across assets. For instance, Comb_TA exhibited the most compact interquartile range in the panel, reflecting greater consistency in heterogeneous financial series. Secondly, SA-based weight optimization demonstrated that adaptive weighting adds measurable value over naive averaging, particularly in assets where the dominant forecasting dynamic is concentrated in a subset of models.

However, the out-of-sample results demonstrate a significant limitation inherent to optimization-based ensembles. Comb_TA did not participate in the OOS period (SMAPE = 0.0884, $\Delta = +25.7\%$), while parsimonious models such as ARIMA (rank 1, SMAPE = 0.0676, $\Delta = -9.2\%$) and STL (rank 2, SMAPE = 0.0706, $\Delta = -10.0\%$) exhibited superior generalization capabilities under the structural changes observed in the forecast horizon. The low Spearman rank correlation between in-sample and out-of-sample rankings ($\rho = 0.236$) confirms a substantial regime shift across periods, consistent with the volatility and structural breaks characteristic of Latin American financial markets during the study period. This finding is consistent with the theoretical bias–variance tradeoff in forecast combination, which posits that adaptive weighting reduces in-sample bias but increases out-of-sample variance when model rankings are unstable across regimes.

These findings underscore the importance of the ongoing exploration of ensemble strategies in contexts analogous to ETFs and FIBRAs, where price dynamics are influenced by macroeconomic shocks, liquidity constraints, and cross-market spillovers that generate non-stationary behavior. Future research could pursue at least three promising avenues. Initially, the incorporation of shrinkage mechanisms that integrate SA-optimized weights with equal-weight baselines has the potential to mitigate out-of-sample variability while maintaining the in-sample benefits of adaptive optimization. Secondly, the employment of time-varying weight schemes, such as rolling-window or regime-switching SA optimization, has the potential to enable the ensemble to adapt to structural breaks in real time, thereby obviating the necessity of relying on fixed historical weights. Thirdly, the expansion of the asset universe to encompass other Latin American ETFs, commodity-linked instruments, or currency-hedged products would serve to test the generalizability of the framework. This expansion could potentially reveal asset classes in which adaptive ensembles maintain their advantage in both evaluation windows. The evidence presented here establishes Comb_TA as a statistically competitive and theoretically grounded forecasting strategy. However, it also indicates that weight stability and regime robustness are the central challenges for its practical deployment in dynamic financial environments.

Finally, the implementation of the SA algorithm was instrumental in achieving superior portfolio optimization, raising returns from an initial 9.64% to over 30% in both prospective and real-world scenarios. The primary benefit of employing this metaheuristic lies in its capacity to navigate complex, non-convex search spaces, which are characteristic of financial time series. In contrast to local optimization methods, SA employs a stochastic acceptance mechanism for suboptimal solutions based on the Boltzmann probability $e^{-\Delta/Tk}$ that facilitates the escape of the system from local minima and the exploration of regions of the solution space that are more conducive to success. This approach engenders a noteworthy degree of model robustness, as evidenced by a Sharpe ratio of 4.4 in real-world performance from 2020 to 2026. Furthermore, the close alignment between out-of-sample (OOS) forecasts and actual historical data validates the algorithm's efficiency in identifying optimal and resilient weight configurations, minimizing the risk of overfitting even in highly volatile environments.

Author Contributions: Conceptualization, J.L.P.A., J.F.S. and J.J.G.B.; methodology, J.L.P.A., J.F.S. and J.J.G.B.; software, J.L.P.A.; investigation, J.L.P.A., J.F.S. and J.J.G.B.; validation, G.C.-V., J.P.S.H. and J.J.G.B.; formal analysis, G.C.-V., J.P.S.H., J.F.S. and J.J.G.B.; writing—original draft preparation, J.L.P.A. and J.J.G.B.; writing—review and editing, J.F.S., J.P.S.H. and J.J.G.B.; visualization, J.F.S. and J.P.S.H.; supervision, J.F.S.; project administration, J.F.S. and J.J.G.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: https://github.com/DrJuanFraustoSolis/SA_TAFE.git (accessed on 4 May 2026).

Acknowledgments: The authors would like to acknowledge, with appreciation and gratitude, SECIHTI, TecNM, and Laboratorio Nacional de Tecnologías de la Información (LaNTI) in the Instituto Tecnológico de Ciudad Madero for the access to the cluster. J.L.P.-A. would also like to thank SECIHTI for the Ph.D. scholarship.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Blanco, C. Mejores ETFs de FIBRAS. 2025. Available online: <https://www.rankia.mx/blog/fondos-indexados/5782438-mejores-etfs-fibras> (accessed on 2 February 2025).
2. Joaquin, G. Mejores ETFs Para 2025 en México. 2025. Available online: <https://www.rankia.mx/blog/fondos-indexados/3774550-mejores-etfs-mexico> (accessed on 30 January 2025).
3. Borja, J.; Terras, X. Mejores ETFs de la Bolsa Mexicana de Valores (BMV) en 2025. Available online: <https://finantres.mx/mejores-etfs-bolsa-mexicana-valores-bmv/> (accessed on 1 January 2026).
4. Solano, J.R. Análisis Económico, Cambiario y Bursátil. FIBRÓMETRO. CDMX, 2025. Available online: <https://www.monex.com.mx/portal/download/reportes/Fib%20Agost2%202025.pdf> (accessed on 6 May 2026).
5. González, V. Fibras Inmobiliarias Resilientes, Rentables y Con Buen Panorama Para el 2S25. Available online: <https://realestatemarket.com.mx/noticias/mercado-inmobiliario/48764-fibras-inmobiliarias-resilientes-rentables-y-con-buen-panorama-para-el-2s25> (accessed on 30 December 2025).
6. Markowitz, H. Portfolio Selection. *J. Financ.* **1952**, *7*, 77–91. [[CrossRef](#)]
7. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680. [[CrossRef](#)]
8. Dueck, G.; Scheuer, T. Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing. *J. Comput. Phys.* **1990**, *90*, 161–175. [[CrossRef](#)]
9. Frausto-Solis, J.; Román, E.F.; Romero, D.; Soberon, X.; Liñán-García, E. Analytically Tuned Simulated Annealing Applied to the Protein Folding Problem. In *Computational Science—ICCS 2007*; Shi, Y., van Albada, G.D., Dongarra, J., Sloat, P.M.A., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 370–377.
10. Erwin, K.; Engelbrecht, A. Meta-heuristics for portfolio optimization. *Soft Comput.* **2023**, *27*, 19045–19073. [[CrossRef](#)]

11. Gunjan, A.; Bhattacharyya, S. Chapter 7—Portfolio optimization using simulated annealing and quantum-inspired simulated annealing: A comparative study. In *Hybrid Computational Intelligence for Pattern Analysis and Understanding*; Bhattacharyya, S., Köppen, M., De, D., Panigrahi, B.K., Eds.; Academic Press: Cambridge, MA, USA, 2024; pp. 213–243. [[CrossRef](#)]
12. Lai, G.; Cai, X.; Yin, H. A Simulated Annealing Approach for Multi-objective Portfolio Optimization. In *Proceedings of the International Conference on Internet of Things, Communication and Intelligent Technology*; Dong, J., Zhang, L., Eds.; Springer Nature: Singapore, 2023; pp. 392–401.
13. Silva, Y.L.T.V.; Herthel, A.B.; Subramanian, A. A multi-objective evolutionary algorithm for a class of mean-variance portfolio selection problems. *Expert Syst. Appl.* **2019**, *133*, 225–241. [[CrossRef](#)]
14. Xue, J.; Shen, B. Dung beetle optimizer: A new meta-heuristic algorithm for global optimization. *J. Supercomput.* **2023**, *79*, 7305–7336. [[CrossRef](#)]
15. Winkler, R.L.; Makridakis, S. The Combination of Forecasts. *J. R. Stat. Soc. Ser. A* **1983**, *146*, 150–157. [[CrossRef](#)]
16. He, K.; Yang, Q.; Ji, L.; Pan, J.; Zou, Y. Financial Time Series Forecasting with the Deep Learning Ensemble Model. *Mathematics* **2023**, *11*, 1054. [[CrossRef](#)]
17. Lv, W.; Qi, J. Stock market return predictability: A combination forecast perspective. *Int. Rev. Financ. Anal.* **2022**, *84*, 102376. [[CrossRef](#)]
18. Shaub, D.; Ellis, P. forecastHybrid: Convenient Functions for Ensemble Time Series Forecasts. 2025. Available online: <https://cran.r-project.org/web/packages/forecastHybrid/index.html> (accessed on 30 December 2025).
19. Bryzgalova, S.; Lerner, S.; Lettau, M.; Pelger, M. Missing Financial Data. *Rev. Financ. Stud.* **2025**, *38*, 803–882. [[CrossRef](#)]
20. Meng, X.; Rubin, D.B. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **1993**, *80*, 267–278. [[CrossRef](#)]
21. Markowitz, H.M. *Portfolio Selection*; Yale University Press: New Haven, CT, USA, 1959. Available online: <http://www.jstor.org/stable/j.ctt1bh4c8h> (accessed on 30 December 2025).
22. Sharpe, W.F. A Simplified Model for Portfolio Analysis. *Manag. Sci.* **1963**, *9*, 277–293. [[CrossRef](#)]
23. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [[CrossRef](#)]
24. Frausto Solis, J.; Purata Aldaz, J.L.; González Del Angel, M.; González Barbosa, J.; Castilla Valdez, G. SAIPO-TAIPO and Genetic Algorithms for Investment Portfolios. *Axioms* **2022**, *11*, 42. [[CrossRef](#)]
25. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2018; p. 504. Available online: <https://otexts.com/fpp2/> (accessed on 30 December 2025).
26. Hyndman, R.; Koehler, A.; Ord, K.; Snyder, R.D. *Forecasting with Exponential Smoothing*; Springer: Berlin/Heidelberg, Germany, 2008. [[CrossRef](#)]
27. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd ed.; OTexts: Melbourne, Australia, 2021. Available online: <https://otexts.com/fpp3/> (accessed on 30 December 2025).
28. Cleveland, R.B.; Cleveland, W.S.; McRae, J.E.; Terpenning, I. STL: A Seasonal-Trend decomposition Procedure Based on Loess. *J. Off. Stat.* **1990**, *6*, 3–73.
29. De Livera, A.M.; Hyndman, R.J.; Snyder, R.D. Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.* **2011**, *106*, 1513–1527. [[CrossRef](#)]
30. Hyndman, R.J.; Khandakar, Y. Automatic Time Series Forecasting: The forecast Package for R. *J. Stat. Softw.* **2008**, *27*, 1–22.
31. Hosking, J.R.M. Fractional Differencing. *Biometrika* **1981**, *68*, 165–176. [[CrossRef](#)]
32. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [[CrossRef](#)]
33. Harvey, D.; Stephen, L.; Newbold, P. Testing the equality of prediction mean squared errors. *Int. J. Forecast.* **1997**, *13*, 281–291. [[CrossRef](#)]
34. Brooks, C. *Introductory Econometrics for Finance*, 4th ed.; Cambridge University Press: Cambridge, UK, 2019. [[CrossRef](#)]
35. Timmermann, A. Forecast Combinations. In *Handbook of Economic Forecasting*; Elliott, G., Granger, C.W.J., Timmermann, A., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 1, pp. 135–196. [[CrossRef](#)]
36. Genre, V.; Kenny, G.; Meyler, A.; Timmermann, A. Combining expert forecasts: Can anything beat the simple average? *Int. J. Forecast.* **2013**, *29*, 108–121. [[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.