Time series forecasting with exogenous variables: a literature review to identify promising gaps in computational research

Rafael Diniz Toscano de Lima[1]*, Sergio Murilo Maciel Fernandes[2] , Sidney Marlon Lopes de Lima[3]

[1] rdtl@ecomp.poli.br, [2] smurilo@ecomp.poli.br. Escola Politécnica de Pernambuco, Universidade de Pernambuco (UPE), Recife, Pernambuco, Brazil
[3] sidney.lima@ufpe.br. Departamento de Eletrônica, Universidade Federal de Pernambuco (UFPE), Recife, Pernambuco, Brazil
*Corresponding author

Abstract
This study presents a comprehensive literature review on integrating exogenous variables in time series forecasting, with a particular focus on financial markets. Both traditional statistical methods, such as ARIMA and SARIMA, and machine learning techniques, like neural networks and support vector regression, are systematically examined, emphasizing their impact on prediction accuracy and the practical challenges involved. Significant research gaps are identified through an in-depth analysis of the existing literature, especially regarding the application of hybrid models and advanced computational techniques. Unlike previous reviews, this paper highlights the potential of combining traditional and machine learning approaches to handle multidimensional data from various external sources better. Addressing these gaps could enhance the robustness and precision of financial forecasting models, offering valuable insights to academic researchers and industry practitioners navigating volatile market environments. The paper consolidates current knowledge and underscores promising computational opportunities that could transform time series forecasting.
Keywords: computational techniques; exogenous variables; financial markets; time series forecasting.

Previsão de séries temporais com variáveis exógenas: uma revisão da literatura para identificar lacunas promissoras na pesquisa computacional

Resumo
Este estudo apresenta uma revisão abrangente da literatura sobre a integração de variáveis exógenas na previsão de séries temporais, com foco particular nos mercados financeiros. Tanto métodos estatísticos tradicionais, como ARIMA e SARIMA, quanto técnicas de aprendizado de máquina, como redes neurais e regressão de vetores de suporte, são examinados sistematicamente, enfatizando seu impacto na precisão da previsão e os desafios práticos envolvidos. Lacunas significativas de pesquisa são identificadas por meio de uma análise aprofundada da literatura existente, especialmente em relação à aplicação de modelos híbridos e técnicas computacionais avançadas. Ao contrário de revisões anteriores, este artigo destaca o potencial de combinar abordagens tradicionais e de aprendizado de máquina para lidar melhor com dados multidimensionais de várias fontes externas. Abordar essas lacunas pode aumentar a robustez e a precisão dos modelos de previsão financeira, oferecendo insights valiosos para pesquisadores acadêmicos e profissionais da indústria que navegam em ambientes de mercado voláteis. O artigo consolida o conhecimento atual e destaca oportunidades computacionais promissoras que podem transformar a previsão de séries temporais.
Palavras-chave: *mercados financeiros; previsão de séries temporais; técnicas computacionais; variáveis exógenas.*

1 Introduction

In the dynamic field of financial markets, precision in predicting market behaviors is essential for both academic researchers and industry practitioners. Time series forecasting plays a critical role by providing key insights into future market trends based on historical data (Sezer; Gudelek; Ozbayoglu, 2020).

Traditional time series models primarily focus on the historical values of the target variable, often overlooking the impact of exogenous variables – factors external to the time series that can

significantly influence forecasting outcomes (Beeram; Kuchibhotla, 2021). These variables may include economic indicators, policy shifts, geopolitical events, and environmental factors, all of which can profoundly affect market dynamics.

Incorporating exogenous variables into time series models represents a major advancement in financial forecasting, enhancing both the accuracy and practical relevance of predictive analytics. However, the systematic inclusion of these variables presents notable challenges. These challenges encompass selecting and preprocessing relevant variables, a task complicated by the vast amount of available data (Ren *et al*., 2023).

Furthermore, the integration of exogenous variables into sophisticated predictive models demands advanced computational techniques, which can increase model complexity. Recent advancements in computational methods, along with the growing availability of large datasets, have opened new avenues for modeling and forecasting financial time series (Rundo *et al*., 2019).

This paper presents a comprehensive and systematic literature review aimed at consolidating existing knowledge on the application of time series models that incorporate exogenous variables in financial markets. Through the examination of a broad spectrum of literature, this review seeks to elucidate methodological advancements, identify persistent gaps, and outline emerging opportunities in computational research. This scholarly effort not only deepens the understanding of the current research landscape but also sets the foundation for future investigations that may enhance and refine financial forecasting models.

By synthesizing findings from a variety of studies, this review constructs an integrated perspective on this complex subject, employing a meticulous and reproducible methodological approach. It is anticipated that this research will inspire and equip scholars with the insights needed to explore new and promising directions in applied computational research across financial markets, economics, and econometrics.

This article is organized as follows. Section 2 provides a comprehensive theoretical foundation, discussing various time series forecasting models and previous works, emphasizing models incorporating exogenous variables. Section 3 outlines the methodology, focusing on the systematic literature review process, including the selection and exclusion criteria, database sources, and quality assessment methods. In Section 4, the selected papers are analyzed to extract key insights regarding integrating exogenous variables into forecasting models, followed by a discussion of their impact on accuracy and performance. Section 5 identifies research gaps and opportunities for future work, proposing several promising directions to enhance the role of exogenous variables in financial time series forecasting. Lastly, Section 6 concludes the study by summarizing the findings, emphasizing the importance of exogenous variables in improving forecasting accuracy, and recommending areas for further research.

2 Theoretical foundation and previous works

Modeling and forecasting in time series analysis are inherently complex tasks. A variety of methods are available to construct models that effectively capture historical data and project future trends. These models support informed decision-making, thereby mitigating risks and optimizing returns.

The development of robust and accurate forecasting models is crucial for precisely predicting market trends. Each modeling technique, from conventional statistical approaches to contemporary computational methods, presents distinct advantages and limitations.

This section provides a summary of previous works in this field (Lima; Fernandes; Melo, 2019), designed to enhance the reader's understanding of the diverse methodologies available for time series analysis. The summary, displayed in Table 1 at the end of this section, spans a spectrum of techniques from straightforward to more complex approaches, offering preliminary insights that contribute to the advancement of computational research opportunities.

2.1 Regression models

Regression analysis has long been a fundamental method for examining relationships among variables across various research fields. This statistical technique primarily aims to quantify the association between a dependent variable and one or more independent variables (Montgomery; Peck;

Vining, 2021). However, when dealing with non-linear data, traditional linear regression models may fail to capture the complexity of these relationships. To address this limitation, advanced techniques such as polynomial regression, spline regression, and non-parametric methods like kernel regression are commonly employed. Moreover, hybrid approaches that combine regression models with machine learning techniques, such as support vector machines or artificial neural networks, provide greater flexibility in modeling non-linear patterns, enabling more accurate predictions across a wide range of datasets.

2.1.1 Exponential smoothing models

The application of Holt-Winters methods for time series forecasting can be particularly challenging when datasets contain sparse values, as noted by Barakat *et al*. (1990). To address the limitations posed by data sparsity, El-Keib, Ma e Ma (1995) developed a hybrid method that combines exponential smoothing with autoregressive and spectral analysis techniques. Despite its origins in the mid-20th century, exponential smoothing remains a valuable tool for addressing various forecasting challenges. This is demonstrated by the adaptations of the Holt-Winters model described in Goodwin's research (2010), as well as tailored models for financial and economic forecasting (Sasongko; Prasetyo; Purbaningtyas, 2017).

2.1.2 Harmonic regression models

Harmonic regression offers an advanced statistical approach distinct from traditional linear models, particularly effective for time series data with cyclical patterns. This technique excels in modeling periodic behaviors by using trigonometric functions such as sine and cosine to capture seasonality.

This model has been effectively applied in environmental science, particularly in predicting daily and monthly averages of particulate matter (PM10) in urban areas. As demonstrated in a study, harmonic regression provided superior accuracy for forecasting PM10 concentrations in London compared to traditional time series methods due to its ability to capture multiple seasonal cycles spanning from 7 days to 15 months (Okkaoglu; Akdi; Ünlü, 2020). Its capability to finely tune seasonal fluctuations makes it a better choice than ARIMA models in this context.

2.1.3 ARIMA models

Autoregressive Integrated Moving Average (ARIMA) models are advanced forecasting tools, particularly effective when time plays a critical role. These models enhance predictive accuracy by leveraging current data, historical observations, and adjustments based on the residual errors of previous forecasts (Taskaya-Temizel; Casey, 2005). One notable advantage of ARIMA is its capacity to correct biases and inconsistencies that often arise in regression analyses, especially when lagged dependent variables are used as predictors. Moreover, ARIMA models frequently outperform traditional multivariate regression and other time series techniques due to their ability to integrate and adjust residual errors, thereby improving overall forecasting accuracy (Peng *et al*., 2003).

However, ARIMA has its limitations, particularly when dealing with data that exhibits strong seasonality or non-linear patterns. While ARIMA can be extended to handle seasonal data through the Seasonal ARIMA (SARIMA) variant, its capacity to capture complex non-linear relationships remains constrained. In such cases, hybrid approaches that combine ARIMA with machine learning techniques, or the transition to models such as exponential smoothing or neural networks, may yield better results. These limitations highlight the importance of selecting the appropriate model based on the data's characteristics, especially in applications where non-linearity or seasonality plays a significant role.

2.2 Stochastics models

Financial market data are often noisy and multi-dimensional. To address these challenges, Vapnik (1998) developed a time series forecasting model based on the Support Vector Machine (SVM) algorithm, a widely recognized machine learning technique for linear solutions. An, Liu, and Venkatesh (2007) later extended this work by combining SVM with genetic algorithms to enhance financial forecasting.

Huang and Wang (2006) demonstrated the efficiency of SVM in classification through concurrent optimization processes, using financial datasets from the University of California Irvine (UCI). Kabran and Ünlü (2021) proposed a dual-step machine learning strategy to predict market bubbles in the S&P 500, initially identifying potential bubbles with a right-tailed unit root test and then employing SVM with macroeconomic indicators for prediction. Their model achieved a prediction accuracy exceeding 96%.

Support Vector Regression (SVR) is closely related to SVM but is designed for regression tasks. It employs various kernel functions to construct hyperplanes, ensuring precise and reliable predictions (Hani'ah *et al*., 2023). SVR is highly resilient to changes in epsilon values, using lossless functions to keep the variance in predicted results within acceptable limits. Through kernel functions, SVR transforms complex, non-linear data into a higher-dimensional feature space (Zaman *et al*., 2021).

Zhang and Qi (2005) explored different types of Artificial Neural Networks (ANNs) for modeling trends and seasonal patterns in time series data, demonstrating their effectiveness in quarterly forecasts. Hamzacebi (2008) enhanced this by developing a novel ANN model specifically for improving seasonal time series predictions.

Taskaya-Temizel and Casey (2005) also evaluated hybrid ANN techniques for time series forecasting, using diverse random and gradient search algorithms to optimize ANN models based on specific data characteristics. Kim and Han (2000) introduced a method combining Genetic Algorithms with RNA models to forecast stock market indices.

Batres-Estrada (2015) reviewed prevalent methodologies for time series forecasting, including stacked autoencoders, convolutional neural networks, and deep belief networks.

2.3 Markov models

A significant development in time series forecasting research is the application of Markov chains. Markov chains provide a robust framework for modeling sequential data, where predictions are based solely on the relationships between observations. These models operate under the assumption that future states depend only on the current state, without influence from prior states. This approach is particularly useful for time series data exhibiting long short-term memory characteristics, such as financial market data (Al-Anzi; Zeina, 2017).

In financial contexts, for instance, Markov chains have been successfully applied to model stock price movements, where the price at any given time is treated as dependent only on the most recent price, rather than the entire price history. Liu (2010) emphasizes that a key technique within this framework is the storage of transition probabilities in matrix format, enabling future state predictions through matrix multiplication. The next state is determined by selecting the one with the highest probability.

This method is particularly valuable when dealing with sparse or highly volatile data, as the transition matrix can be dynamically adjusted to account for the probability of rare events, making it more adaptable to fluctuations. In highly volatile markets, where prices can change drastically, the Markov chain's ability to focus on immediate transitions rather than long-term historical data allows for more responsive and agile forecasting.

2.4 Chapter insights

Based on the observations it is evident that addressing the challenges presented by dynamic financial markets necessitates further exploration of the potential for incorporating exogenous variables into computational research. External factors such as economic indicators and geopolitical events significantly impact market volatility and dynamics.

Recognizing the profound influence of these variables underscores the need to explore their integration within time series models more comprehensively. This approach can bridge existing research gaps and enhance the effectiveness of financial forecasting tools, providing decision-makers with more robust methodologies for navigating the complexities of global financial environments. By thoroughly understanding the intricacies of this issue, the aim is to methodically identify gaps and opportunities that could refine financial time series forecasting.

Incorporating exogenous variables into time series models presents significant challenges, particularly in selecting relevant variables and determining the optimal lag between the exogenous

inputs and the target variable. The vast volume and variety of external data, ranging from macroeconomic indicators to social media sentiment, often require advanced techniques to filter and preprocess the data before integration. Failure to select the most relevant variables or account for appropriate time lags can lead to overfitting or underfitting, thereby reducing the model's predictive power. This underscores the importance of employing robust feature selection methods and cross-validation techniques to ensure that the chosen exogenous variables meaningfully contribute to the forecasting process.

Additionally, fine-tuning model parameters becomes increasingly complex when integrating exogenous variables. Standard time series models like ARIMA may require modifications or extensions, such as ARIMAX models, which incorporate exogenous data. However, determining the optimal configuration of parameters, including lag length or integration order, often necessitates iterative processes like grid search or more advanced optimization algorithms. In practice, this requires balancing model complexity with computational efficiency, particularly when handling large datasets or high-dimensional data spaces. Hybrid models that combine statistical methods with machine learning approaches, such as ARIMA integrated with neural networks or support vector machines, have shown promise in managing these complexities effectively while improving prediction accuracy.

Moreover, the dynamic nature of exogenous variables often necessitates adaptive techniques capable of recalibrating model parameters as new data become available. Methods such as Kalman filtering or reinforcement learning can assist models in adjusting to shifts in external conditions without requiring complete retraining. This adaptability is crucial in highly volatile environments, such as financial markets, where real-time data continuously influences the underlying time series. By incorporating these adaptive mechanisms, forecasters can develop more resilient models that are better suited to handle the unpredictable nature of exogenous influences, ultimately enhancing the robustness and accuracy of their predictions.

3 Methodology

This section presents the methodologies and fundamental statistics employed in conducting the Systematic Literature Review (SLR) and the bibliometric analysis of the articles selected for this study.

3.1 Inclusion/exclusion criteria

Following the protocol established by Thomé, Scavarda, and Scavarda (2016), the steps were: formulating a research question, conducting a literature search, collecting data, establishing quality criteria, analyzing and synthesizing the data, and interpreting gaps and opportunities for further research.

As specified by this SLR protocol, the following Research Question (RQ) was formulated: Can incorporating exogenous variables affect the performance of time series forecasting models? To address this question, a set of strict inclusion and exclusion criteria were applied during the search, as outlined below.

Inclusion criteria

- Works published in English;
- Works published in peer-reviewed journals;
- Works published within the last five years;
- Search string containing: "time series" AND ("exogenous variable" OR "explained variable" OR "input variables" OR "forecast variable" OR "explanatory variable" OR "dependent variable").

Exclusion criteria

- Works that are not fully available;
- Unfinished works;
- Works that do not include mathematical, statistical, or computational analysis;
- Works lacking clear specifications.

Following the establishment of these criteria, exploratory searches were conducted across various article databases to identify the most relevant studies for achieving the research objectives.

3.2 Paper databases

Initially, the search was performed across the entire scope of the topic. A search string using broad keywords was developed, ensuring that the search would not artificially restrict the number of studies retrieved, yet remain specific enough to include only studies pertinent to the subject. This approach aligns with Petticrew and Roberts' (2006) recommendation to maintain sensitivity while ensuring specificity to exclude irrelevant studies. Consequently, the search was conducted in four of the most influential global databases: IEEE, ACM, Science Direct, and Scopus.

This search yielded 3,801 articles published between 2019 and 2023. Table 2 details these results by publication databases and year.

Table 2 – Publications by database and year of publication

Database	Year of publication				
	2019	2020	2021	2022	2023
Science Direct	532	592	712	724	687
ACM	12	19	21	39	33
Scopus & Web of Science	46	45	71	64	44
IEEE	15	40	42	31	32

Source: research data (2024)

Subsequently, a bibliometric analysis was conducted on 160 papers, excluding those without citations. This exclusion criterion ensures that the analysis is focused on works that have demonstrated some impact or relevance within the scientific community. Articles without citations may indicate a lack of interest or relevance in the field of study and thus would not meaningfully contribute to the objectives of the bibliometric analysis. The purpose of this approach is to comprehend the publication landscape and the influence of these works within the context of the research.

3.2.1 Paper stratification

The Pareto principle, or the 80-20 rule, was applied to the analysis. This principle posits that approximately 80% of the results are produced by 20% of the causes. In the context of bibliometric research, the principle helps identify the most significant contributions, allowing for a focus on key trends in the field, the topics that attract the most attention, and the journals with the highest impact. Such insights are crucial for researchers, academic institutions, and decision-makers seeking to understand the dynamics of research in specific fields and allocate resources more efficiently.

It is important to acknowledge, however, that this approach does not entirely exclude journals with fewer articles or papers with low citation counts. Such works may hold considerable value in specific research niches. While it may exclude some emerging research that has not yet garnered significant citations, these relevant studies can still be integrated into future research as they gain recognition over time, ensuring they are not permanently overlooked.

Nevertheless, applying the Pareto principle allows for an emphasis on analyzing the most influential portion of the dataset. Despite the prominence of generalist journals, the relevance and impact of works published in specialized journals within fields such as computing, data science, pattern recognition, machine learning, and artificial intelligence remain notable. These areas have driven significant technological advancements, particularly in electrical and electronic engineering research.

The predominance of generalist journals is natural in many disciplines, often due to the interdisciplinary nature of the research, which attracts a broader audience. However, it is crucial to recognize that beneath this surface generalism, significant specialized contributions play a critical role in advancing knowledge and fostering innovation.

3.3 Quality selection criteria

A systematic literature review is a critical, structured approach to analyzing and synthesizing the available evidence within a specific research area. To ensure that the articles included in the review are of high quality, it is essential to define solid evaluation criteria and assign a quality score to each article. In this context, seven key questions, each with an associated weight, were established to assess the quality of the articles selected for an in-depth systematic review.

- (Q1) Is the proposed approach clearly described in the abstract? (Weight: 3)

Justification: the abstract is the reader's first point of contact and provides a concise summary of the research. A clear and well-articulated abstract suggests that the authors understand their approach and can effectively communicate their research goals and findings. This criterion is assigned a weight of 3 to reflect its importance in ensuring that the essence of the research is immediately understandable, though it is considered less critical than other aspects of the study.
Justification: The abstract is the reader's first point of contact and provides a concise summary of the research. A clear and well-articulated abstract suggests that the authors understand their approach and can effectively communicate their research goals and findings. This criterion is assigned a weight of 3 to reflect its importance in ensuring that the essence of the research is immediately understandable, though it is considered less critical than other aspects of the study.

- (Q2) Was the research context adequately described (experiments, configurations, techniques)? (Weight: 5)

Justification: the research context, including the description of experiments, configurations, and techniques, is crucial for evaluating the study's validity and reproducibility. A well-described research context allows for replication and verification, which are fundamental to scientific rigor. This criterion carries the highest weight (5) because a thorough understanding of the research context is essential for assessing the study's credibility and reliability.

- (Q3) Is the discussion of the study results sufficiently comprehensive? (Weight: 2)

Justification: A comprehensive discussion of the results demonstrates that the authors have critically engaged with their findings, considering the implications, limitations, and potential areas for future research. This criterion is assigned a weight of 2, highlighting its significance in ensuring that the study offers valuable insights. However, it is considered slightly less critical than the clarity of the research context or the practical applicability of the findings in real-time scenarios.

- (Q4) Does the study involve real-time forecasting or analysis? (Weight: 5)

Justification: real-time forecasting or analysis is particularly relevant in fields such as finance, where timely predictions can have significant consequences. This criterion is assigned the highest weight (5) because real-time forecasting demonstrates the study's practical applicability and relevance to decision-making in dynamic environments.

- (Q5) Does the approach support multiple domain? (e.g.: identification, mining, quantification, metrics) (Weight: 3)

Justification: an approach that can be applied across multiple domains demonstrates robustness and broad applicability—valuable traits in a forecasting model. This criterion is weighted at 3, reflecting its importance in assessing the generalizability of the approach while recognizing that domain-specific studies may still offer significant contributions.

- (Q6) Does the approach incorporate the advantages of classical forecasting models? (Weight: 4)

Justification: Integrating classical forecasting models can enhance the performance of new approaches by leveraging well-established methodologies. This question is weighted at 4, highlighting the importance of building upon existing knowledge and techniques to develop more effective forecasting models.

- (Q7) Does the approach focus on financial time series forecasting? (Weight: 5)

Justification: forecasting financial time series is complex and high-stakes, where accuracy is critical. This criterion is weighted at 5, reflecting the importance of financial forecasting within the review, given the dynamic and impactful nature of financial markets.

These seven questions were empirically developed to assess the quality and relevance of the articles within the context of time series forecasting. Each question's weight reflects its relative importance in ensuring that the selected studies contribute meaningfully to the field, particularly in critical areas of practical application and theoretical rigor.

The most relevant papers were selected based on the quality scores assigned using these criteria. This approach narrowed the selection to less than 10% of the initially cataloged 160 papers, as shown in Table 3, which outlines the final research funnel process.

Table 3 – Most relevant paper based on quality selection criteria

Paper	Score
The role of weather predictions in electricity price forecasting beyond the day-ahead horizon (Sgarlato; Ziel, 2023)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(×); Q6(×); Q7() Score: 22
A hybrid swarm-based system for commodity price forecasting during the covid-19 pandemic (Xavier; Fernandes; Oliveira, 2023)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(×); Q6(×); Q7() Score: 22
An architecture to improve energy-related time-series model validity based on the novel rMAPE performance metric (Mares *et al*, 2023)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(×); Q6(×); Q7() Score: 22
Climate indices impact in monthly streamflow series forecasting (Toledo *et al*, 2023)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(×); Q6(×); Q7() Score: 22
Spatio-temporal characterization of stochastic dynamic transportation networks (Filipovska; Mahmassani, 2023)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(×); Q6(×); Q7() Score: 22
Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting (Bukhari *et al*, 2020)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(); Q6(); Q7(x) Score: 20
Short-term photovoltaic power forecasting based on long short-term memory neural network and attention mechanism (Zhou *et al*, 2019)	Q1(×); Q2(×); Q3(×); Q4(×);Q5(); Q6(); Q7() Score: 15
A new hybrid model for short-term electricity load forecasting (Haq; Ni, 2019)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(); Q6(); Q7() Score: 15
The use of mutual information to improve value-at-risk forecasts for exchange rates (Antwi; Kyei; Gill, 2020)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(); Q6(); Q7() Score: 15

Deep learning-based multistep solar forecasting for PV ramp-rate control using sky images (Wen *et al*, 2021)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(); Q6(); Q7() Score: 15
LSTM-MSNet: leveraging forecasts on sets of related time series with multiple seasonal patterns (Bandara; Bergmeir; Hewamalage, 2021)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(); Q6(); Q7() Score: 15
State-space models for online post-covid electricity load forecasting competition (Vilmarest; Goude, 2022)	Q1(×); Q2(×); Q3(×); Q4(×); Q5(); Q6(); Q7() Score: 15
A causal framework for distribution generalization (Christiansen *et al*, 2022)	Q1(×); Q2(×); Q3(×); Q4(); Q5(×); Q6(); Q7() Score: 13
A quality-related fault detection method based on the dynamic data-driven algorithm for industrial systems (Sun *et al*, 2022)	Q1(×); Q2(×); Q3(×); Q4(); Q5(×); Q6(); Q7() Score: 13
Modeling fuel consumption and NOX emission of a medium-duty truck diesel engine with comparative time-series methods (Ozmen *et al*, 2021)	Q1(×); Q2(×); Q3(×); Q4(); Q5(); Q6(); Q7() Score: 10

Source: research data (2024)

4 Analysis of selected papers

The analysis of the reviewed papers provides a comprehensive understanding of the advanced techniques and methods applied in time series forecasting across various domains, including weather forecasting, electrical and solar energy generation, asset forecasting, financial prediction, flow forecasting, and transportation flows. These studies represent significant advancements in both theoretical knowledge and practical applications, addressing specific challenges and illustrating how time series forecasting serves as a powerful tool in an increasingly data-driven world.

A recurring theme in these papers is the increasing importance of integrating machine learning and statistical modeling techniques into time series forecasting. These approaches have proven highly effective in diverse fields, delivering notable results in terms of accuracy and adaptability across different data types.

One of the key highlights is the application of deep neural networks, such as Long Short-Term Memory (LSTM) networks. For instance, Zhou *et al*. (2019) propose a framework utilizing two LSTM networks, one for temperature forecasting and another for photovoltaic power generation prediction. These networks are capable of learning complex features from time series data and modeling long-term dependencies, making them particularly effective for longer-term forecasts. Similarly, Bukhari *et al*. (2020) present a hybrid model that combines the fractional order derived from ARFIMA with the dynamic characteristics of LSTM networks, applied to high-frequency financial market forecasts, demonstrating superior accuracy compared to traditional models.

Wen *et al*. (2021) present a solar forecasting method based on deep convolutional networks (ConvNets), which integrate additional exogenous variables. This technique stands out for its simplicity and effectiveness in predicting solar energy generation, thereby contributing to the efficient integration of renewable energy into power grids. Bandara, Bergmeir, and Hewamalage (2021) proposed a novel forecasting framework, LSTM-MSNet, for time series with multiple seasonal patterns. The authors combined a globally trained LSTM network with state-of-the-art multiseasonal decomposition techniques to improve forecasting accuracy. These decomposition techniques were employed to extract various seasonal components from time series data, thereby enhancing the LSTM learning process.

Another prominent trend involves hybrid models that combine different techniques to improve forecasting precision. For instance, Haq and Ni (2019) introduced a model utilizing Empirical Mode Decomposition (IEMD) to decompose electrical charge time series, an innovative approach that addresses the limitations of standard empirical mode decomposition (EMD). Moreover, it incorporates T-Copula correlation analysis and binary variables indicative of peak load, resulting in improved

forecasting accuracy. This hybrid model illustrates how the integration of various techniques can yield substantial improvements in predicting electrical loads.

Xavier, Fernandes, and Oliveira (2023) developed a hybrid system that integrates the Particle Swarm Optimization (PSO) algorithm with ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables) and Support Vector Regression (SVR) models to predict commodity prices during the COVID-19 pandemic. This system demonstrates how parameter optimization and feature selection can significantly enhance forecast accuracy, particularly in periods of heightened market volatility. In this context, hybrid models combining Particle Swarm Optimization (PSO) with ARIMAX and Support Vector Regression (SVR) are particularly effective. PSO facilitates the search for optimal parameters by simulating a population of potential solutions that iteratively improve based on a fitness function. Compared to traditional methods, PSO navigates complex, non-linear search spaces more efficiently, resulting in better-tuned models and improved predictive performance, especially in volatile and high-dimensional datasets.

Ozmen *et al*. (2021) applied advanced time series modeling techniques to forecast fuel consumption and NOx emissions of a medium-duty truck diesel engine. Their approach included various techniques, such as Nonlinear Autoregressive Network with Exogenous Inputs (NARX), ARIMAX, Multiple Linear Regression (MLR), and Regression Error with Autoregressive Moving Average (RegARMA). These methods were used to estimate NOx emissions and fuel consumption based on input variables like exhaust gas recirculation temperature, engine coolant temperature, engine speed, and exhaust gas pressure. Comparative analyses of these modeling methods highlighted the critical role of intelligent modeling techniques in achieving accurate predictions and enhancing internal combustion engine performance analysis.

In Sgarlato and Ziel (2023), exogenous variables were incorporated using an autoregressive multivariate linear model and LASSO to forecast wholesale electricity prices in Germany. The inclusion of weather forecasts as external factors proved crucial in improving forecast accuracy, particularly for shorter prediction horizons. This highlights the importance of accounting for external factors when modeling time series, especially in sectors highly sensitive to external disturbances.

Mares *et al*. (2023) combined inputs from four models – ANN, SVR, LSTM, and RTE – to forecast energy demand and electricity market time series. The model's performance was validated using a rolling window technique, with a step size of k depending on the number of models in each subset. The results showed that LSTM and RTE outperformed SVR and ANN, confirming the superiority of these techniques under specific conditions.

Similarly, Toledo *et al*. (2023) explored the use of feature selection techniques and machine learning models such as Bayesian Neural Networks, Support Vector Regression, and Gaussian Processes to predict monthly flow time series. The inclusion of climate indices significantly enhanced prediction accuracy, showcasing the value of incorporating external information to improve model performance.

Antwi, Kyei, and Gill (2020) utilized several techniques to enhance value-at-risk forecasts for exchange rates, such as incorporating mutually dependent covariate returns to create exogenous break variables and augment GARCH models. They also applied mutual information and Pearson's method to evaluate the strength of linear dependencies.

Sun *et al*. (2022) proposed a dynamic kernel entropy component regression (DKECR) framework to address the instability of quality-related fault detection due to the dynamic characteristics of the process. In comparison to the typical kernel entropy component analysis method, this approach establishes a direct relationship between process states and quality states, thereby elucidating the impact of faults on product quality.

Vilmarest and Goude (2022) employed statistical correction techniques for meteorological variables, standard statistical and machine learning models, state-space models (e.g., Kalman filter and Viking), and expert aggregation methods. Statistical correction of meteorological variables was used during data preprocessing to account for factors such as temperature and cloud cover, which influence electricity load forecasting.

Christiansen *et al*. (2022) addressed the challenge of predicting a response variable Y from a set of covariates X when the training and test distributions differ. Recognizing that such differences may

have causal explanations, they focused on minimizing the worst-case risk, accounting for test distributions arising from interventions within a structural causal model.

Change detection and community analysis have also emerged as valuable tools. Filipovska and Mahmassani (2023) concentrated on modeling transport networks with randomly varying travel times, employing change and community detection techniques to characterize the spatial and temporal features of the network. Their two-part approach first applies a modified community detection technique to assess the network structure and travel time data. Then, a change point detection method is used to identify the moments when shifts in travel time distributions occur, revealing robust spatial and temporal network patterns under changing exogenous and endogenous conditions.

In conclusion, these studies collectively advance the field of time series forecasting by demonstrating innovative methodologies and cutting-edge techniques across diverse domains. They not only improve forecast accuracy but also provide valuable insights for decision-makers, researchers, and practitioners reliant on time series forecasting.

4.1 Insights for enhanced time series forecasting

Table 4 summarizes critical insights into time series forecasting across various domains based on the analysis of multiple studies. These insights underscore the importance of carefully selecting, preprocessing, and integrating exogenous variables, which are vital for improving the accuracy and reliability of forecasting models.

Table 4 – Insights into the selection, preprocessing, and integration of exogenous variables

Process	Insights
Selection of exogenous variables	The importance of identifying relevant exogenous variables, such as weather forecasts, climate indices, and market-specific externalities, is highlighted in numerous studies. Techniques like feature selection (e.g., Bayesian Neural Networks, Support Vector Regression) are employed to determine which external factors significantly improve model accuracy (Antwi; Kyei; Gill, 2020; Toledo *et al.*, 2023)
Preprocessing of exogenous variables	Statistical correction of meteorological variables is applied to adjust for the effects of temperature and cloud cover on energy load forecasting (Vilmarest; Goude, 2022) Change detection and community analysis are employed to characterize the spatial and temporal features of transport networks and identify when changes in exogenous conditions occur (Filipovska; Mahmassani, 2023) Mutual information and Pearson's method are used to measure the strength of linear dependencies between exogenous variables and time series data (Antwi; Kyei; Gill, 2020)
Integration of exogenous variables	Deep neural networks, such as LSTM and convolutional networks, are frequently utilized to integrate exogenous variables into time series forecasting models. These networks are effective at capturing complex dependencies and improving long-term forecast accuracy (Wen *et al.*, 2021; Zhou *et al.*, 2019) Hybrid models, which combine traditional methods (e.g., ARIMAX, ARFIMA) with machine learning techniques (e.g., LSTM, SVR), show enhanced forecasting performance in complex and volatile domains such as financial markets and energy demand (Bukhari *et al.*, 2020; Mares *et al.*, 2023; Xavier; Fernandes; Oliveira, 2023) Models that include exogenous variables like weather forecasts (e.g., LASSO for electricity price forecasting) significantly improve forecast accuracy, particularly for shorter horizons (Sgarlato; Ziel, 2023)

Source: research data (2024)

4.1.1 Selection of exogenous variables

The selection of appropriate exogenous variables is a critical step that directly affects the performance of forecasting models. In time series forecasting, exogenous variables refer to external factors that can significantly influence the target variable. The insights provided in Table 4 emphasize the importance of identifying relevant exogenous variables, such as weather forecasts, climate indices, and specific market externalities like economic indicators. Advanced techniques, such as feature selection through Bayesian Neural Networks and Support Vector Regression, enable researchers to identify which external factors are most impactful, resulting in more precise and reliable predictions. This step is crucial, as it ensures the model is not overwhelmed by irrelevant data, which could otherwise degrade its performance and reduce accuracy.

Building on the identification of these relevant factors, the next step involves preparing the data for optimal use in the model.

4.1.2 Preprocessing of exogenous variables

Once the relevant exogenous variables have been identified, the next essential step is preprocessing. Preprocessing involves preparing the data to maximize its usefulness for the forecasting model. The insights from Table 4 highlight various preprocessing techniques, including the statistical correction of meteorological data and change detection in transport networks. For instance, adjusting for the effects of temperature and cloud cover in energy load forecasting ensures that the model accurately reflects real-world conditions influencing energy consumption. Similarly, change detection and community analysis help identify temporal shifts in transport networks, enabling the model to adapt to changing conditions. Preprocessing is vital, as it transforms raw data into a refined input that the model can effectively utilize, ultimately enhancing its forecasting accuracy. With the data properly preprocessed, the final step is to integrate these variables into the forecasting model in a way that maximizes predictive performance.

4.1.3 Integration of exogenous variables

The final step in leveraging exogenous variables is their integration into the forecasting model. As highlighted in Table 4, there is a growing trend towards using deep learning models, such as Long Short-Term Memory (LSTM) networks and convolutional networks, to incorporate these variables, capturing complex dependencies and improving long-term forecast accuracy. Furthermore, hybrid models that combine traditional statistical methods with machine learning techniques have demonstrated enhanced forecasting capabilities, especially in complex and volatile environments like financial markets and energy demand forecasting. The integration process is crucial, as it allows the model to account for external influences, making the forecasts more robust and applicable to real-world scenarios.

4.2 Objective enhancing metrics

The analysis of enhancement metrics from the highest-rated research papers reveals substantial improvements in forecasting accuracy when exogenous variables are integrated into predictive models, as summarized in Table 5. For example, studies focused on electricity price forecasting using weather predictions beyond the day-ahead horizon demonstrated a 10-20% reduction in Root Mean Square Error (RMSE) when exogenous variables were included. Similarly, the hybrid swarm-based system for commodity price forecasting during the COVID-19 pandemic showed superior performance, achieving the best Mean Squared Error (MSE) results in 63% of the datasets analyzed. These findings emphasize the potential of exogenous variables in improving model precision and adapting to external influences.

Table 5 – Objective enhancements metrics

Paper	Enhancement
The role of weather predictions in electricity price forecasting beyond the day-ahead horizon (Sgarlato; Ziel, 2023)	Improvement of the RMSE by 10-20%
A hybrid swarm-based system for commodity price	Best MSE results in 63% of datasets

forecasting during the covid-19 pandemic. (Xavier; Fernandes; Oliveira, 2023)	
An architecture to improve energy-related time-series model validity based on the novel rMAPE performance metric. (Mares *et al*, 2023)	Forecasting results showed an improvement in MAPE of up to 23%
Climate indices impact in monthly streamflow series forecasting.(Toledo *et al*, 2023)	RMSE improvements ranged from 0.45% to 25.52% when models with climate indices as exogenous variables were employed
Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. (Bukhari *et al*, 2020)	The proposed model achieved the lowest MAPE; the ARFIMA-LSTM model improved financial series prediction accuracy by 80% compared to traditional models
Short-term photovoltaic power forecasting based on long short-term memory neural network and attention mechanism. (Zhou *et al*, 2019)	The ALSTM model significantly reduced MAPE, RMSE, and MAE values compared to LSTM
A new hybrid model for short-term electricity load forecasting. (Haq; Ni, 2019)	MAPE decreased by 15.27%, and RMSE decreased by 13.86% compared to other techniques
Deep learning-based multistep solar forecasting for PV ramp-rate control using sky images. (Wen *et al*, 2021)	Forecasting accuracy improved by 5.6% to 17.7% in MAE and RMSE
State-space models for online post-covid electricity load forecasting competition. (Vilmarest; Goude, 2022)	The proposed method showed a 3% lower MAE compared to other aggregation strategies

Source: research data (2024)

Additionally, the impact of climate indices on monthly streamflow series forecasting is particularly noteworthy. Models that incorporated these indices as exogenous variables achieved significant gains, with RMSE improvements ranging from 0.45% to 25.52%. Another compelling example is the application of the Fractional Neuro-Sequential ARFIMA-LSTM model for financial market forecasting, which outperformed traditional models, improving prediction accuracy by an impressive 80%. These results underscore the crucial role of exogenous variables in capturing external factors that significantly influence the target series, leading to more accurate and reliable predictions.

The effectiveness of integrating exogenous variables is further demonstrated in energy-related forecasting studies. For short-term photovoltaic power forecasting, the ALSTM model, which combines LSTM with an attention mechanism, substantially reduced MAPE, RMSE, and MAE values compared to standard LSTM models. Similarly, a new hybrid model for short-term electricity load forecasting reported a 15.27% decrease in MAPE and a 13.86% reduction in RMSE, outperforming other techniques. In the context of post-COVID electricity load forecasting, state-space models exhibited a 3% lower MAE compared to other aggregation strategies. These improvements in accuracy metrics highlight the significant impact of exogenous variable integration on enhancing the robustness and predictive power of time series forecasting models across various domains.

5 Discussion on research opportunities and gaps

The analysis of the studies in Table 3 and the techniques in Table 6 reveals several relevant factors concerning research opportunities and gaps. This analysis proposes multiple pathways for enhancing the integration of exogenous variables in time series forecasting models. Although underexplored by the authors, several factors present notable research opportunities or gaps, underscoring areas for future inquiry and suggesting directions for advancing the field. Some of the identified research gaps and opportunities are discussed in the subsections that follow.

Table 6 – Techniques applied in each paper

Paper	Techniques
Zhou *et al*. (2019)	LSTM, Fully Connected Layers, Attention Mechanism
Haq; Ni (2019)	IEMD, T-Copula, Deep Belief Network
Bukhari *et al*. (2020)	ARFIMA, LSTM
Antwi, Kyei, and Gill (2020)	Covariate Returns, Mutual Information, GARCH
Wen *et al*. (2021)	Deep ConvNets, Stacked Sky Images
Bandara; Bergmeir; Hewamalage (2021)	LSTM, Multiseasonal Decomposition Techniques
Ozmen *et al*. (2021)	NARX, ARIMAX, MLR, RegARMA
Christiansen *et al*. (2022)	Causal Regression Models
Sun *et al*. (2022)	Kernel Entropy Component Regression (DKECR)
Vilmarest; Goude (2022)	Statistical Correction, Machine Learning, Expert Aggregation
Sgarlato; Ziel (2023)	Autoregressive Multivariate Linear Model, LASSO
Xavier; Fernandes; Oliveira (2023)	Particle Swarm Optimization, ARIMAX, SVR
Mares *et al*. (2023)	ANN, SVR, LSTM, RTE
Toledo *et al*. (2023)	Bayesian Neural Networks, SVR, Gaussian Process
Filipovska; Mahmassani (2023)	Community Detection, Change Point Detection

Source: research data (2024)

5.1 Theoretical relevance of variables

While some studies acknowledge the theoretical importance of exogenous variables, there remains an opportunity to investigate more sophisticated methodologies, such as complex network theories. These approaches could prove instrumental in identifying causal links and synergies between exogenous variables and the target variable. Exploring these relationships could deepen the understanding of how exogenous variables affect time series, potentially leading to enhanced forecasting accuracy.

5.2 Causality and correlation

An underexplored area involves applying advanced causal inference methods to examine interactions between exogenous variables and the target variable. While numerous studies touch upon causality and correlation, further research is needed to employ machine learning algorithms to uncover causal connections within complex time series. Additionally, the possibility of reverse causation between exogenous variables and the target variable warrants closer investigation, particularly when the causality is unclear.

5.3 Data quality

While some studies recognize the importance of data quality for exogenous variables, there remains significant potential to explore advanced data cleaning and processing methods. Machine learning-based data imputation techniques could address issues of missing or sparse data. Utilizing more sophisticated methods to ensure data reliability presents a substantial research opportunity to improve model accuracy.

5.4 Temporal lag

Though some studies discuss the temporal lag between exogenous variables and the target variable, there is an opportunity to investigate more advanced lag selection methods. Techniques based on neural networks, which can detect complex temporal patterns among variables, offer promising directions for further research. These advanced lag selection methods could yield a more precise understanding of the relationship between exogenous variables and the target variable, leading to improved forecasting models that account for appropriate temporal delays.

5.5 Interaction with endogenous variables

Although the interaction between exogenous and endogenous variables is noted in the literature, there is considerable scope for employing more sophisticated modeling techniques to capture these dynamics better and improve model efficacy.

5.6 Effect of sample size

The impact of sample size on the integration of exogenous variables into forecasting models is acknowledged but remains insufficiently explored. Further research could examine how variations in sample size influence model accuracy across different scenarios and investigate more robust strategies for managing limited data. Additionally, deeper analysis of the role of exogenous variables in predictions, particularly through advanced variable selection methods and improved model interpretability techniques, could help identify the most critical variables.

5.7 Model adequacy

While some researchers highlight the importance of model adequacy, future work could benefit from comparing various modeling approaches under specific conditions. The application of more sophisticated techniques, such as deep learning, may also help boost predictive performance.

5.8 Data update

Some studies discuss the need for regular updates to data on exogenous variables. Future research should investigate optimal frequencies for data updates and how these adjustments impact the long-term accuracy and stability of forecasting models.

5.9 Generalization of results

Although generalization is occasionally addressed, advancing research could focus on employing sophisticated cross-validation methods and transfer learning strategies to improve model generalization. Additionally, the exploration of advanced prediction techniques, such as reinforcement learning or Bayesian models, could assess their efficacy across different contexts and integrate domain-specific knowledge to enhance the selection and interpretation of exogenous variables.

5.10 Multi-scale forecasting

The concept of multi-scale forecasting remains underexplored. While some authors discuss forecasting at various scales, future investigations could develop models that incorporate exogenous variables to forecast multiple time series across different scales more effectively.

5.11 Use of unconventional exogenous variables

The potential for incorporating non-traditional exogenous variables, such as data from social networks or media, which may have indirect relevance to the prediction domain, has yet to be fully explored. Future research could investigate these unconventional data sources, focusing on robust forecasting models capable of withstanding disturbances and accounting for uncertainty, thereby enhancing reliability in rapidly changing environments.

5.12 Use of exogenous variables in real-time forecast

The implementation of real-time forecasting using exogenous variables remains an underexplored area, presenting significant opportunities for future research. One of the key challenges in this domain is the timely availability of exogenous data. In real-time settings, data delays, inconsistencies, and missing information can severely affect the accuracy and reliability of forecasts.

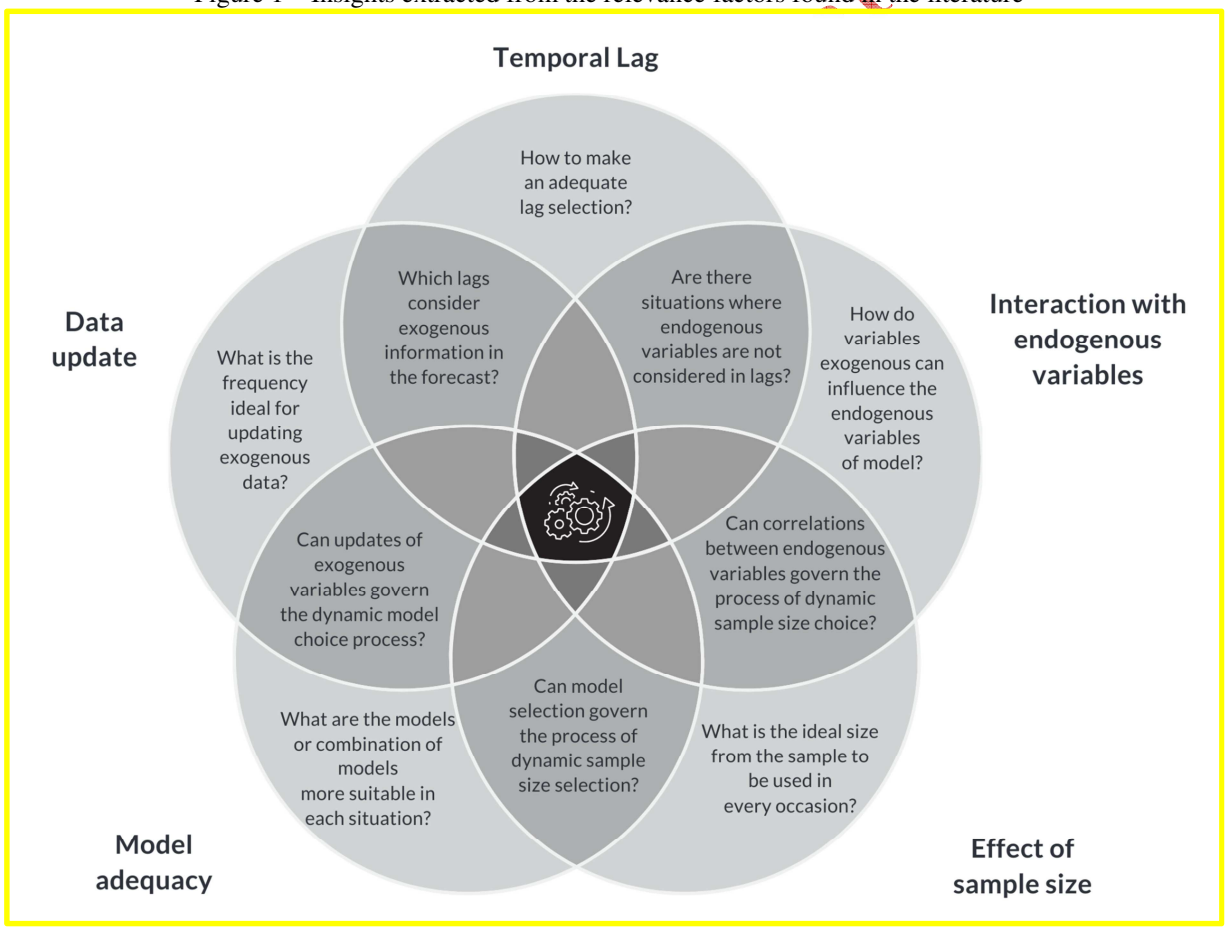
To address this, future research could focus on developing more robust data collection and processing pipelines that ensure the consistent flow of external data in near real-time, minimizing latency. Additionally, adaptive models that can account for sudden changes in the availability or quality of exogenous data would improve the resilience and flexibility of real-time forecasting systems.

Another critical issue is the dynamic nature of exogenous variables, which often require continuous monitoring and updating to reflect real-time changes. For example, variables such as economic indicators, social media sentiment, or weather conditions can fluctuate rapidly, making it challenging to incorporate them effectively into forecasting models. Future studies could explore the development of adaptive machine learning techniques, such as online learning algorithms, that can evolve with incoming data and automatically adjust model parameters in response to changes in exogenous variables. This would not only enhance the accuracy of real-time forecasts but also improve the models' ability to handle volatile and unpredictable data sources.

5.13 Chapter insights

Figure 1 contributes to identifying new overlapping areas between relevant factors, providing a foundation for further research insights that the scientific community can explore. This conclusion underscores the importance of these intersections in advancing the understanding of complex time series forecasting models and highlights promising directions for future investigations.

Figure 1 – Insights extracted from the relevance factors found in the literature



Source: elaborated by authors (2024)

6 Conclusion

The studies presented in Table 3 provide a solid foundation for analyzing critical factors that reveal research opportunities and gaps related to the integration of exogenous variables into time series forecasting models. Progress in these areas could significantly improve the accuracy and reliability of forecasting models across various sectors, including economics, energy, transportation, healthcare, and

climate science. The results presented in the selected studies validate the central hypothesis that incorporating exogenous variables into time series models enhances their accuracy. As further research advances, this area could yield breakthroughs in the performance and practicality of forecasting models.

The scope of this work represents an original contribution to the significant topic of exogenous variables in financial time series, offering a model that addresses key gaps using computational techniques. This model holds the potential to revolutionize approaches to financial time series forecasting. A critical gap identified in the literature is the real-time factor in forecasting financial series, which is essential for effective decision-making. Given the volatile nature of financial markets, the ability to forecast in real time is crucial for ensuring success in algorithm-driven decision-making processes.

Real-time forecasting is essential to maintain the innovative character of scientific advancement. As illustrated in Figure 1, the insights extracted from the relevance factors found in the literature highlight synergies among different problems, framing them as optimization challenges. This interdisciplinary approach emphasizes the need for solutions that span multiple domains, particularly in computer science and engineering, to address complex issues in financial time series forecasting.

Future research should delve into more sophisticated theoretical frameworks for analyzing the significance of exogenous variables, helping to clarify why certain factors have a stronger impact on time series predictions. By employing advanced causal inference techniques, such as those developed by Angrist, Imbens, and Rubin (1996), researchers could uncover deeper insights into the interactions between exogenous variables and the target variable. These methods would facilitate a better understanding of how different variables jointly influence forecasting outcomes.

There is also a pressing need to improve the preprocessing of exogenous data. Enhancing data cleaning and processing methods could significantly improve model accuracy by ensuring higher-quality input data. Techniques capable of addressing noisy, incomplete, or biased data, and dynamically adjusting to the nature of exogenous variables, would be valuable for improving forecasting performance.

Additionally, integrating neural network-based methodologies for lag selection could be a fruitful avenue for further research. These techniques could help detect complex temporal patterns in time series data, leading to more accurate models. Investigating the effects of sample size on model accuracy, especially in contexts with limited data, is another area requiring attention. Strategies such as data augmentation or transfer learning could help maintain model accuracy even when data availability is scarce.

Finally, the integration of domain-specific knowledge into forecasting models is a promising direction for future research. While exogenous variables provide important insights, combining them with expert knowledge from specific sectors could further enhance performance. Multi-scale forecasting, hybrid modeling, and the inclusion of unconventional exogenous variables should be explored to address the unique challenges posed by different sectors, including finance, healthcare, and climate science.

Acknowledgment

The authors express deep gratitude to the Polytechnique University of Pernambuco for its invaluable support and infrastructure, which were instrumental to the success of this research project. The university's commitment to fostering an academic environment that promotes excellence has significantly enhanced the quality and scope of this work. The authors also extend their appreciation to the faculty, researchers, staff, and administrative teams, whose mentorship and encouragement were vital to this project. Their professionalism and dedication ensured a seamless and enriching research experience at the university.

Funding

This research did not receive any financial support, nor did it involve human or material resources from external sources.

Competing interest

The authors declare no conflict of interest.

Ethics approval

This study did not include human or animal subjects, and no ethic committee approval was required.

Author contributions

LIMA, R. D. T.: conception and design of the study. **FERNANDES, S. M. M.:** data analysis and interpretation. **LIMA, S. M. L.:** final revision with critical intellectual input. All authors contributed to the writing, discussion, review, and approval of the final version of the manuscript.

References

AN, S.; LIU, W.; VENKATESH, S. Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. **Pattern Recognition**, v. 40, n. 8, p. 2154-2162, 2007. DOI: https://doi.org/10.1016/j.patcog.2006.12.015.

ANGRIST, J. D.; IMBENS, G. W.; RUBIN, D. B. Identification of causal effects using instrumental variables. **Journal of the American Statistical Association**, v. 91, n. 434, p. 444-455, 1996. DOI: https://doi.org/10.2307/2291629.

ANTWI, A.; KYEI, K. A.; GILL, R. S. The use of mutual information to improve value-at-risk forecasts for exchange rates. **IEEE Access**, v. 8, p. 179881-179900, 2020. DOI: https://doi.org/10.1109/ACCESS.2020.3027631.

AL-ANZI, F. S.; ZEINA, D. A. Statistical Markovian data modeling for natural language processing. **International Journal of Data Mining & Knowledge Management Process (IJDKP)**, v. 7, n. 1, p. 25-35, 2017. DOI: https://doi.org/10.5121/ijdkp.2017.7103.

BANDARA, K.; BERGMEIR, C.; HEWAMALAGE, H. LSTM-MSNet: leveraging forecasts on sets of related time series with multiple seasonal patterns. **IEEE Transactions on Neural Networks and Learning Systems**, v. 32, n. 4, p. 1586-1599, 2021. DOI: https://doi.org/10.1109/TNNLS.2020.2985720.

BARAKAT, E. H.; QAYYUM, M. A.; HAMED, M. N.; AL-RASHED, S. A. Short term peak demand forecasting in fast developing utility with inherent dynamic load characteristics. **IEEE Transactions on Power**, v. 5, n. 3, p. 813-824, 1990. DOI: https://doi.org/10.1109/59.65910.

BATRES-ESTRADA, G. Deep learning for multivariate financial time series. 2015. Degree Project (Mathematical Statistics) – **KTH Royal Institute of Technology**, Stockholm, 2015. Available at: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A820891&dswid=-2616. Accessed on: 06 sep. 2024.

BEERAM, S. R.; KUCHIBHOTLA, S. A survey on state-of-the-art financial time series prediction models. *In*: 2021 International Conference on Computing Methodologies and Communication (ICCMC), 5., 2021, Erode. **Proceedings [...].** Erode: IEEE, p. 596-604, 2021. DOI: https://doi.org/10.1109/ICCMC51019.2021.9418313.

BUKHARI, A. H.; RAJA, M. A. Z.; SULAIMAN, M.; ISLAM, S.; SHOAIB, M.; KUMAM, P. Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. **IEEE Access**, v. 8, p. 71326-71338, 2020. DOI: https://doi.org/10.1109/ACCESS.2020.2985763.

CHRISTIANSEN, R.; PFISTER, N.; JAKOBSEN, M. E.; GNECCO, N.; PETERS, J. A causal framework for distribution generalization. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 44, n. 10, p. 6614-6630, 2022. DOI: https://doi.org/10.1109/TPAMI.2021.3094760.

EL-KEIB, A. A.; MA, X. M.; MA, H. Advancement of statistical-based modeling for short-term load forecasting. **Electric Power Systems Research**, v. 35, n. 1, p. 51-58, 1995. DOI: https://doi.org/10.1016/0378-7796(95)00987-6.

FILIPOVSKA, M.; MAHMASSANI, H. S. Spatio-temporal characterization of stochastic dynamic transportation networks. **IEEE Transactions on Intelligent Transportation Systems**, v. 24, n. 9, p. 9929-9939, 2023. DOI: https://doi.org/10.1109/TITS.2023.3276190.

GOODWIN, P. The Holt-Winters approach to exponential smoothing: 50 years old and going strong. **Foresight**, v. 19, p. 30-33, 2010.

HAQ, M. R.; NI, Z. A new hybrid model for short-term electricity load forecasting. **IEEE Access**, v. 7, p. 125413-125423, 2019. DOI: https://doi.org/10.1109/ACCESS.2019.2937222.

HANI'AH, M.; ABDULLAH, M. Z.; SABILLA, W. I.; AKBAR, S.; SHAFARA, D. R. Google trends and technical indicator based machine learning for stock market prediction. MATRIK: **Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer**, v. 22, n. 2, p. 271-284, 2023. DOI: https://doi.org/10.30812/matrik.v22i2.2287.

HUANG, C.-L.; WANG, C.-J. A GA-based feature selection and parameters optimization for support vector machines. **Expert Systems with Applications**, v. 31, n. 2, p. 231-240, 2006. DOI: https://doi.org/10.1016/j.eswa.2005.09.024.

KABRAN, F. B.; ÜNLÜ, K. D. A two-step machine learning approach to predict S&P 500 bubbles. **Journal of Applied Statistics**, v. 48, n. 13-15, p. 2776-2794, 2021. DOI: https://doi.org/10.1080/02664763.2020.1823947.

KIM, K.-J.; HAN, I. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. **Expert Systems with Applications**, v. 19, n. 2, p. 125-132, 2000. DOI: https://doi.org/10.1016/S0957-4174(00)00027-0.

LIMA, R. D. T.; FERNANDES, S. M. M.; MELO, I. P. L. Fourier genetic series: an evolutionary time series modeling technique. *In*: 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI), 2019, Guayaquil. **Proceedings [...]**. Guayaquil: IEEE, p. 1-6, 2019. DOI: https://doi.org/10.1109/LA-CCI47412.2019.9037031.

LIU, T. Application of Markov chains to analyze and predict the time series. **Modern Applied Science**, v. 4, n. 5, 2010. DOI: http://dx.doi.org/10.5539/mas.v4n5p162.

MARES, J. J.; CHARRIS, D.; PARDO, M.; QUINTERO M, C. G. An architecture to improve energy-related time-series model validity based on the novel rMAPE performance metric. **IEEE Access**, v. 11, p. 36004-36014, 2023. DOI: https://doi.org/10.1109/ACCESS.2023.3264713.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. 6. ed. Wiley, 2021.

OKKAOGLU, Y.; AKDI, Y.; ÜNLÜ, K. D. Daily PM10, periodicity, and harmonic regression model: the case of London. **Atmospheric Environment**, v. 238, 117755, 2020. DOI: https://doi.org/10.1016/j.atmosenv.2020.117755.

OZMEN, M. I.; YILMAZ, A.; BAYKARA, C.; OZSOYSAL, O. A. Modelling fuel consumption and NOx emission of a medium-duty truck diesel engine with comparative time-series methods. **IEEE Access**, v. 9, p. 81202-81209, 2021. DOI: https://doi.org/10.1109/ACCESS.2021.3082030.

PENG, H.; OZAKI, T.; HAGGAN-OZAKI, V.; TOYODA, Y. A parameter optimization method for radial bias function type models. **IEEE Transactions on Neural Networks**, v. 14, n. 2, p. 432-438, 2003. DOI: https://doi.org/10.1109/TNN.2003.809395.

PETTICREW, M.; ROBERTS, H. **Systematic reviews in the social sciences: a practical guide**. Wiley, 2006. DOI: https://doi.org/10.1002/9780470754887.

REN, Z. J.; LOWRY, G. V.; BOEHM, A. B.; BROOKS, B. W.; GAGO-FERRERO, P.; JIANG, G.; JONES, G. D.; LIU, Q.; WANG, S.; ZIMMERMAN, J. B. Data science for advancing environmental science, engineering, and technology. **Environmental Science & Technology**, v. 57, n. 46, p. 17661-17662, 2023. DOI: https://doi.org/10.1021/acs.est.3c08700.

RUNDO, F.; TRENTA, F.; STALLO, A. L.; BATTIATO, S. Machine learning for quantitative finance applications: a survey. **Applied Sciences**, v. 9, n. 24, 5574, 2019. DOI: https://doi.org/10.3390/app9245574.

SASONGKO, R. S; PRASETYO, E.; PURBANINGTYAS, R. System prediction production PT. VICO Indonesia using method Holt-Winters. **Journal of Electrical Engineering and Computer Sciences (JEECS)**, v. 2, n. 1, p. 221-230, 2017. DOI: https://doi.org/10.54732/jeecs.v2i1.166.

SEZER, O. B.; GUDELEK, M. U.; OZBAYOGLU, A. M. Financial time series forecasting with deep learning: a systematic literature review: 2005-2019. **Applied Soft Computing**, v. 90, 106181, 2020. DOI: https://doi.org/10.1016/j.asoc.2020.106181.

SGARLATO, R.; ZIEL, F. The role of weather predictions in electricity price forecasting beyond the day-ahead horizon. **IEEE Transactions on Power Systems**, v. 38, n. 3, p. 2500-2511, 2023. DOI: https://doi.org/10.1109/TPWRS.2022.3180119.

SUN, C.-Y.; YIN, Y.-Z.; KANG, H.-B.; MA, H.-J. A quality-related fault detection method based on the dynamic data-driven algorithm for industrial systems. **IEEE Transactions on Automation Science and Engineering**, v. 19, n. 4, p. 3942-3952, 2022. DOI: https://doi.org/10.1109/TASE.2021.3139766.

TASKAYA-TEMIZEL, T.; CASEY, M. C. A comparative study of autoregressive neural network hybrids. **Journal of Neural Networks**, v. 18, n. 5-6, p. 781-789, 2005. DOI: https://doi.org/10.1016/j.neunet.2005.06.003.

THOMÉ, A. M. T.; SCAVARDA, L. F. S.; SCAVARDA, A. J. Conducting systematic literature review in operations management. **Production Planning Control**, v. 27, n. 5, p. 408-420, 2016. DOI: https://doi.org/10.1080/09537287.2015.1129464.

TOLEDO, J. F.; SIQUEIRA, H. V.; BIUK, L. H.; SACCHI, R.; AZAMBUJA, R. R.; ASANO JUNIOR, R.; ASANO, P. T. L. Climate indices impact in monthly streamflow series forecasting. **IEEE Access**, v. 11, p. 21451-21464, 2023. DOI: https://doi.org/10.1109/ACCESS.2023.3237982.

VAPNIK, V. N. The nature of statistical learning theory. 2. ed. **Springer-Verlag**, 2000. DOI: https://doi.org/10.1007/978-1-4757-3264-1.

VILMAREST, J.; GOUDE, Y. State-space models for online post-covid electricity load forecasting competition. **IEEE Open Access Journal of Power and Energy**, v. 9, p. 192-201, 2022. DOI: https://doi.org/10.1109/OAJPE.2022.3141883.

WEN, H.; DU, Y.; CHEN, X.; LIM, E.; WEN, H.; JIANG, L.; XIANG, W. Deep learning based multistep solar forecasting for PV ramp-rate control using sky images. **IEEE Transactions on Industrial Informatics**, v. 17, n. 2, p. 1397-1406, 2021. DOI: https://doi.org/10.1109/TII.2020.2987916.

XAVIER, A. L. S.; FERNANDES, B. J. T.; OLIVEIRA, J. F. L. A hybrid swarm-based system for commodity price forecasting during the COVID-19 pandemic. **IEEE Access**, v. 11, p. 74379-74387, 2023. DOI: https://doi.org/10.1109/ACCESS.2023.3293738.

ZHANG, G. P.; QI, M. Neural network forecasting for seasonal and trend time series. **European Journal of Operational Research**, v. 160, n. 2, p. 501-514, 2005. DOI: https://doi.org/10.1016/j.ejor.2003.08.037.

ZHOU, H.; ZHANG, Y.; YANG, L.; LIU, Q.; YAN, K.; DU, Y. Short-term photovoltaic power forecasting based on long short-term memory neural network and attention mechanism. **IEEE Access**, v. 7, p. 78063-78074, 2019. DOI: https://doi.org/10.1109/ACCESS.2019.2923006.