

Do AI-Based Forecasting Systems Improve Predictability in Foreign Exchange Markets Compared to Classical Econometric Models?

Constantin – Laurențiu STAMA*

National University of Science and Technology POLITEHNICA Bucharest, Bucharest, Romania

**Corresponding author, constantin.stama@upb.ro*

Cătălin – George ALEXE

National University of Science and Technology POLITEHNICA Bucharest, Bucharest, Romania

catalin.alexe@upb.ro

Gheorghe MILITARU

National University of Science and Technology POLITEHNICA Bucharest, Bucharest, Romania

gheorghe.militaru@upb.ro

Cătălina – Monica ALEXE

National University of Science and Technology POLITEHNICA Bucharest, Bucharest, Romania

catalina.alexe@upb.ro

Abstract. *This paper presents a comparative review of the literature on foreign exchange rate forecasting, with focus on whether artificial intelligence methods can improve prediction results when compared with classical econometric models. Recent research has moved from theory-based statistical models toward data-based methods that are able to identify nonlinear patterns and use both numerical market data and information from financial texts. However, the results of these studies are difficult to compare directly because they use different currency pairs, time periods, forecasting horizons, input variables, and evaluation methods. This makes it harder to draw a clear conclusion about how large and stable the reported improvements are.*

The review examines fifteen peer-reviewed studies indexed in Web of Science, most published between 2019 and 2025. The studies are grouped into four categories: classical econometric models, machine learning models, deep learning models, and approaches connected with large language models. Based on document review and qualitative comparison, the paper analyzes whether artificial intelligence methods perform better than econometric benchmarks in out-of-sample forecasting. It also looks at the main design choices that may explain better forecasting results.

The evidence reviewed in this paper suggests that artificial intelligence models often achieve better results than classical econometric models on common forecasting measures such as root mean squared error, mean absolute error, and directional accuracy. Deep learning models, especially long short-term memory variants and attention-based models, tend to show stronger performance in several studies. The review also shows that models using financial text may add useful information that is difficult to include in standard econometric models. The paper brings together findings from different strands of the literature and shows where the improvements are more visible. At the same time, it points out several remaining problems, especially model interpretation, stability across different market conditions, and practical use in real forecasting settings.

Keywords: foreign exchange forecasting, artificial intelligence, machine learning, deep learning, large language models.

DOI: 10.2478/picbe-2026-0051

© 2026 C.– L. Stama; C. – G. Alexe; G. Militaru; C. – M. Alexe, published by De Gruyter.

This work is licensed under the Creative Commons Attribution 4.0 License.

Introduction

The foreign exchange market is the largest and most liquid financial market in the world. Exchange rate movements matter for international trade, capital flows, risk management, investment decisions, and monetary policy. For this reason, the ability to forecast exchange rates has been an important topic in both academic research and financial practice. Even so, reliable forecasting remains difficult, especially when models are tested on data outside the sample used for estimation.

Exchange rates are influenced by many factors at the same time. These include macroeconomic conditions, interest rate expectations, inflation, market sentiment, political events, and the way investors react to new information. These factors often interact in complex ways and their influence may change over time. Because of this, the main difficulty is not simply to explain past exchange rate movements. A more practical question is whether available information can be used to produce forecasts that are stable and useful in real conditions.

Traditional exchange rate forecasting has usually relied on macroeconomic theory. Models based on purchasing power parity, interest rate parity, and monetary fundamentals explain exchange rate movements through differences in prices, interest rates, money supply, and other economic variables. These models remain important because they provide clear economic meaning and help connect forecasts with theory. However, many empirical studies have shown that these models often have weak out-of-sample forecasting performance. One of the best-known studies is that of Meese and Rogoff (1983), who found that structural econometric models often failed to perform better than a simple random walk. Later research reached similar conclusions, showing that exchange rate predictability changes across currencies, time periods, and forecasting horizons (Rossi, 2013).

This difficulty has led researchers to use artificial intelligence methods in foreign exchange forecasting. Machine learning and deep learning models are attractive because they can work with nonlinear relationships and large sets of variables. They can also learn patterns from past data without requiring the same strict assumptions used in many econometric models. Deep learning methods are especially useful for time series because they can model sequences and may capture changes in market behavior more flexibly.

More recent studies have also started to include textual information in forecasting models. Financial news, central bank communication, analyst reports, and market commentary may contain signals about expectations and sentiment. These signals are hard to measure through standard numerical variables. For this reason, some studies combine numerical data with financial text, while others use methods related to large language models to process written information. These developments suggest that better forecasting may come from two sources: more flexible models and richer information.

At the same time, the literature is still hard to compare. Studies often use different datasets, currency pairs, sample periods, forecast horizons, and performance measures. Some papers focus on short-term forecasts, while others examine longer horizons. Some compare artificial intelligence models with random walks, while others use more advanced econometric benchmarks. Because of these differences, it is not always clear whether better results come from the method itself or from the specific research setting.

The main gap addressed in this paper is the lack of a clear comparative view of artificial intelligence methods and classical econometric models in foreign exchange forecasting. Many recent studies focus on a single model, dataset, or type of information. Fewer studies bring together the main approaches and compare them across broader methodological groups. This leaves an

important question open: do artificial intelligence-based forecasting systems actually improve exchange rate predictability compared with classical econometric approaches, and in which conditions do these improvements appear more likely?

This paper answers that question through a comparative literature review of fifteen peer-reviewed studies indexed in Web of Science. The reviewed studies are organized into four categories: classical econometrics, machine learning, deep learning, and approaches related to large language models. The paper does not treat these studies as identical experiments, since their designs differ. Instead, it compares them through a qualitative reading of their forecasting performance, model structure, information inputs, and evaluation choices.

The paper has three main aims. First, it compares the main types of methods used in recent foreign exchange forecasting research. Second, it explains why some artificial intelligence models may produce better forecasts, especially through their ability to capture nonlinear patterns, follow time-based sequences, use attention mechanisms, and include information from text. Third, it identifies the main directions in the current literature, including multimodal forecasting, transformer-based models, and the use of large language models for financial text analysis. The paper also discusses the limits that still affect this area, especially comparability between studies, model transparency, stability across market conditions, and practical implementation.

The review is guided by one central research question: do artificial intelligence-based forecasting systems improve foreign exchange predictability compared with classical econometric approaches? Three hypotheses are used to structure the analysis. The first hypothesis is that, within the reviewed studies, artificial intelligence models tend to perform better than classical econometric baselines on standard forecasting measures and directional accuracy. The second hypothesis is that deep learning models tend to show more stable improvements than traditional machine learning methods, because they are better suited to time series patterns and changing market conditions. The third hypothesis is that models using textual information through multimodal structures or large language model-related components provide additional predictive value compared with models based only on numerical data.

Literature review

Research on exchange rate forecasting has developed around a constant problem: economic theory can explain many exchange rate movements, but it does not always help produce accurate forecasts. In the classical view, exchange rates are connected to macroeconomic factors such as price levels, interest rate differences and monetary conditions. Forecasting is usually done through econometric models, either structural models or simpler reduced-form models. In this area, the random walk has become an important benchmark because it is very hard to beat in out-of-sample tests. The study by Meese and Rogoff (1983) is often used as a reference because it showed that exchange rate models based on economic theory often performed worse than a simple random walk. Later studies reached similar conclusions and showed that predictability depends on the currency, the forecasting horizon and the selected time period (Rossi, 2013).

Classical econometric research has developed over time, but several limits are still present. Some newer models try to account for changes in market conditions. For example, regime-switching models and density forecasting methods are used because exchange rates may behave differently from one period to another. Jaworski (2018) shows that these models can offer a more realistic view of exchange rate movements by allowing changes between regimes and by using simulation-based forecasts. Even so, these models do not remove the main forecasting difficulty.

Econometric models still need a clear mathematical form, a fixed set of numerical variables and assumptions that may become weak when market relations change or become nonlinear. For this reason, classical models remain useful mainly as a reference point. They offer a basis for comparison when newer forecasting methods are tested.

The use of machine learning marks a clear change in this field. Instead of starting mainly from economic theory, these methods learn patterns from data. Compared with traditional econometric models, machine learning methods can work better with nonlinear relations and with interactions between variables. They do not require the researcher to define all relations in advance. This is useful in foreign exchange markets, where the behavior of exchange rates can differ depending on the currency pair, the period analyzed and the forecast horizon. Reviews of machine learning in finance show that these methods are helpful when the data are complex and difficult to describe through fixed statistical models (Nazareth & Reddy, 2023). Studies focused on exchange rate forecasting also report better results than some statistical benchmarks, although the size of these improvements differs from one case to another (Maneejuk & Srichaikul, 2021; Bormpotsis et al., 2023). In this sense, machine learning should be seen as a useful forecasting tool, not as a full replacement for economic reasoning.

Deep learning goes further by using neural networks with several layers that can learn patterns from sequential data. This matters in exchange rate forecasting because currency movements are influenced by past values, but the importance of those past values can change over time. Recurrent neural networks and long short-term memory models are often used for this reason. They are built to process ordered data and to capture links between observations across time (Dautel et al., 2020). Recent studies combine these models with more input variables, hybrid structures and attention mechanisms. Attention mechanisms help the model give more weight to the most relevant past information when making a forecast (Liu et al., 2024; Chi et al., 2025; Ghahremani & Nguyen, 2025). The main idea in this part of the literature is that deep learning models may perform better because they are less dependent on stable linear relations. They can learn changing patterns directly from data.

Another important direction is the use of more types of information in forecasting. Classical econometric models usually rely on numerical indicators. However, exchange rates are also influenced by news, central bank messages, political events, analyst opinions and general market expectations. Because of this, some researchers have started to use models that combine numerical time-series data with text. Windsor and Cao (2022) show that the use of both numerical and textual information in a deep learning model can improve forecasting results. Mabrouk et al. (2025) also argue that financial news can add useful information when it is included in semi-supervised deep learning models. The logic behind these studies is simple: text may contain signals about sentiment, expectations and market interpretation that are hard to express through standard numerical variables. For this reason, multimodal forecasting has become more relevant in the literature.

A newer direction is linked to large language models and transformer-based models. In finance, these models are increasingly used to extract information from large amounts of text. They can help identify sentiment, topics and other signals from financial news, reports or public statements. Studies such as Fatouros et al. (2023) and Yang et al. (2025) show that ChatGPT-type models are becoming more relevant for financial text analysis and exchange rate research. Transformer models are also used directly in time-series forecasting. Their attention mechanism can capture both short-term and long-term relations in the data, without the same limits found in

recurrent models (Zhao & Yan, 2024). This area is still developing, but it is important because it connects text analysis, financial forecasting and sequence models.

Overall, the literature shows that artificial intelligence methods are useful mainly because they are flexible. Machine learning can handle nonlinear relations better than many standard econometric models. Deep learning can work well with time-based data and changing market patterns. Multimodal and language-model-based methods can add information from text, which is difficult to include in a classical numerical model. These features help explain why several recent studies report better forecasting results than traditional benchmarks (Nazareth & Reddy, 2023; Liu et al., 2024; Windsor & Cao, 2022; Yang et al., 2025). Still, the literature is divided. Many studies analyze only one model, one dataset or one type of input. Direct comparisons are difficult because papers use different currencies, time periods, forecast horizons, benchmarks and evaluation measures. Because of this, it is hard to know whether the better results come from the method itself or from the way the study was designed.

The research gap becomes clear at this point. There are many studies on machine learning, deep learning, multimodal forecasting and large language model-related methods, but fewer works bring these findings together in a clear comparative form. This paper addresses that need by treating the literature as a set of different forecasting approaches, each with its own assumptions, inputs and limits. The purpose is to show where artificial intelligence methods seem to bring more stable improvements, what model features are usually linked to better results and what problems remain. The main issues are still model interpretation, stability across different market conditions and the difficulty of comparing studies that use different evaluation designs.

Methodology

This study uses a systematic literature review, followed by a comparative qualitative analysis. The purpose is to examine whether artificial intelligence methods lead to better results in foreign exchange forecasting when they are compared with classical econometric models. The study also looks at the methodological choices that are most often linked to better forecasting performance. The review follows a structured approach, like earlier works on exchange rate forecasting and machine learning in finance, but the focus is placed on comparison between different types of methods, not on bibliometric statistics (Rossi, 2013; Nazareth & Reddy, 2023).

A qualitative synthesis was chosen instead of a statistical meta-analysis because the reviewed studies are very different from one another. They use different currency pairs, periods, forecast horizons, model settings and evaluation methods. Because of these differences, it would be difficult to combine their results in a single numerical estimate without losing important details. For this reason, the analysis compares the studies by looking at repeated patterns, model types, data inputs, benchmarks and reported results.

Data source and search strategy

The main database used for the review was Clarivate Web of Science Core Collection. The search focused on studies published between 2019 and 2025. Three older benchmark studies were also included because they are important for understanding the long-standing forecasting problem in this field. These older studies were used only for theoretical and contextual support. They were not treated as part of the main recent evidence.

The search was carried out in four stages. In the first stage, the aim was to identify studies from the classical econometric area. The search query was:

TS=("foreign exchange" OR forex) AND TS=(forecast* OR prediction) AND TS=(econometric* OR ARIMA OR GARCH OR VAR).

In the second stage, the search was expanded toward machine learning and deep learning studies. The query used was:

TS=("foreign exchange" OR forex) AND TS=(forecast* OR prediction) AND TS=("machine learning" OR "deep learning" OR neural network*).

In the third stage, the search focused on studies related to large language models and generative methods, especially for the 2023 to 2025 period. The query was:

TS=("large language model*" OR LLM OR GPT) AND TS=(finance OR "foreign exchange").

In the fourth stage, an additional search was used to identify studies that directly compared different forecasting approaches. The query was:

TS=("exchange rate forecasting") AND TS=("machine learning" OR "deep learning" OR econometric*).

This search strategy was used to cover the main research directions included in the paper. It also helped avoid a sample focused too much on only one type of method.

Inclusion and exclusion criteria

The review is guided by one main research question: do artificial intelligence forecasting systems improve foreign exchange predictability compared with classical econometric approaches?

Three hypotheses are used to organize the analysis. The first hypothesis is that artificial intelligence models tend to perform better than classical econometric baselines on the standard forecasting measures reported in the literature. The second hypothesis is that deep learning models tend to show more consistent improvements than conventional machine learning models, because they are designed to work with sequential data and changing relations over time. The third hypothesis is that models using textual information, either through multimodal systems or large language model components, tend to add predictive value beyond models based only on numerical data.

Research questions and hypotheses

The review is guided by one main research question: do artificial intelligence forecasting systems improve foreign exchange predictability compared with classical econometric approaches?

Three hypotheses are used to organize the analysis. The first hypothesis is that artificial intelligence models tend to perform better than classical econometric baselines on the standard forecasting measures reported in the literature. The second hypothesis is that deep learning models tend to show more consistent improvements than conventional machine learning models, because they are designed to work with sequential data and changing relations over time. The third hypothesis is that models using textual information, either through multimodal systems or large language model components, tend to add predictive value beyond models based only on numerical data.

Data extraction and comparative synthesis procedure

Each study included in the review was analyzed using the same extraction template. The information collected included the forecasting setting, such as currency pairs and forecast horizon when available, the type of input data, the model class, the model architecture, the benchmark used for comparison, the evaluation measures and the main direction of the results. The studies were

then grouped into four categories: classical econometrics, machine learning, deep learning and methods related to large language models.

For each category, the review summarizes the main model families, the benchmarks used, the evaluation measures reported and the limits mentioned by the original authors. Since the studies use different datasets, time periods, forecast horizons and reporting styles, the findings are compared through qualitative analysis. The focus is placed on results that appear in more than one study, while the limits of comparison are clearly noted where the research designs differ too much.

This approach keeps the comparison clear without treating all studies as identical experiments. It allows the review to show where artificial intelligence methods seem to improve forecasting results, while also keeping attention on the differences between studies.

Results and discussions

Comparative performance across paradigms

The fifteen studies reviewed in this paper show a repeated tendency: models based on artificial intelligence often report better forecasting results than classical econometric models. This result must be read with care, because the studies do not use the same datasets, currency pairs, forecast horizons or evaluation rules. For this reason, the discussion is presented as a comparative reading of the literature, not as a statistical meta-analysis. The aim is to identify the areas where the reported improvements appear more often and to explain which model features may be connected with these results.

Classical econometric models are used in this paper mainly as the reference point for comparison. The study by Meese and Rogoff (1983) showed that structural exchange rate models often failed to perform better than a random walk in out-of-sample tests. Rossi (2013) also showed that exchange rate predictability remains unstable and depends on the selected currency, time period and forecast horizon. Later econometric models, including regime-switching models and density forecasting methods, try to reflect changes in market behavior more clearly. Jaworski (2018), for example, shows that models with regime changes and simulation-based forecasts can describe exchange rate movements in a more realistic way. Even so, these methods still face the basic problem of producing stable forecasts across different market settings.

This point is important for the interpretation of all later results. In exchange rate forecasting, even a small improvement over a simple benchmark can matter, but it has to be checked carefully. A result is more convincing when the model is tested outside the estimation sample, when the benchmark is clearly presented and when the model is not adjusted only for one dataset or one market episode. The reviewed literature therefore shows that forecasting design matters as much as the model name. The key issue is whether the chosen model can deal with nonlinear relations, changing patterns and a large amount of market information.

In the machine learning group, the evidence discussed by Nazareth and Reddy (2023) supports the idea that data-based models can improve prediction in financial applications. In foreign exchange forecasting, Maneejuk and Srichaikul (2021) and Bormpotsis et al. (2023) also report better results than statistical benchmarks in their tested settings. These findings suggest that machine learning can be useful when the relation between variables is nonlinear or when several predictors interact with each other. Still, the reported gains are usually moderate. Machine learning adds flexibility, but it does not remove all forecasting problems. Its performance depends on the selected model, the construction of input variables and the evaluation method.

Deep learning studies report stronger improvements more often, especially when the models are designed for time-series data. Dautel et al. (2020) use deep recurrent neural networks for exchange rate forecasting and show that these models can improve predictive performance. Later studies build on this direction by using long short-term memory models, bidirectional structures, hybrid models and attention mechanisms. Liu et al. (2024) combine LASSO-based feature selection with a BiLSTM ensemble model. Chi et al. (2025) test feature-augmented multivariate LSTM models in a real-time data setting. Ghahremani and Nguyen (2025) use an attention-based LSTM model, which allows the model to give more importance to relevant past information. In these studies, the advantage of deep learning seems to come from its ability to capture temporal patterns, nonlinear links and changes in the importance of past observations.

Another important result is linked to the use of text. Windsor and Cao (2022) show that combining numerical time-series data with textual information can improve forecasting accuracy. Mabrouk et al. (2025) also argue that financial news can add value when included in a semi-supervised deep learning model. This matters because exchange rates are influenced by more than numerical indicators. Markets also react to news, central bank messages, investor expectations and financial commentary. Text-based information can reflect sentiment and market interpretation, which are hard to include in a standard numerical model. In this sense, multimodal forecasting can improve results by using a wider set of information.

The most recent group includes large language models and transformer-based methods. Fatouros et al. (2023) discuss the role of ChatGPT-type models in financial sentiment analysis, while Yang et al. (2025) examine the use of ChatGPT-based text processing in exchange rate analysis and prediction. At the same time, Zhao and Yan (2024) study transformer-based models for currency forecasting. These models are relevant because they can process large amounts of text or time-series information through attention mechanisms. This is one of the newest directions in the field. The results are promising, but they should be treated with caution because the evidence is still recent and the number of directly comparable studies is limited.

Assessment of hypotheses

The first hypothesis states that artificial intelligence models tend to perform better than classical econometric baselines. The reviewed studies generally support this hypothesis. The strongest support appears in deep learning and multimodal studies, while machine learning studies usually report smaller but positive improvements (Nazareth & Reddy, 2023; Liu et al., 2024; Windsor & Cao, 2022). This does not mean that artificial intelligence models are better in every possible case. The evidence shows that they are more likely to bring gains when the forecasting task includes nonlinear relations, changing time patterns or a wider information set.

The second hypothesis states that deep learning models tend to produce more stable gains than conventional machine learning models. This hypothesis is also supported by the reviewed sample. Several deep learning studies report better out-of-sample forecasting results, especially when the model is built to process sequential data and to adjust the weight given to past information (Dautel et al., 2020; Chi et al., 2025; Ghahremani & Nguyen, 2025). The main reason is that deep learning models fit the structure of exchange rate data better than many simpler methods. Currency movements develop over time, and past values may matter differently from one period to another.

The third hypothesis states that textual information and large language model-related components can add predictive value beyond numerical data. The evidence supports this hypothesis, but with more caution. Studies using financial news, text signals or language-model-

based processing report additional value in forecasting systems (Windsor & Cao, 2022; Mabrouk et al., 2025; Yang et al., 2025). However, this part of the literature is newer and less standardized than the literature on deep learning with numerical time series. For this reason, the conclusion is useful, but still more tentative than the conclusions for the first two hypotheses.

Mechanisms that plausibly explain performance differences

The reviewed studies suggest several reasons why artificial intelligence models may obtain better forecasting results than classical econometric models. First, many econometric models depend on fixed mathematical forms. This can be limiting when the relation between variables changes over time or becomes nonlinear. Machine learning and deep learning models can learn more flexible relations from data (Nazareth & Reddy, 2023; Liu et al., 2024). This is important in foreign exchange markets, where the role of predictors can change during periods of monetary policy shifts, crises or market uncertainty.

Second, deep learning models built for sequences can represent time dependence more directly. Models such as LSTM and attention-based architectures can give different importance to past observations. This may explain why these models often report stronger results than simpler methods. In exchange rate forecasting, the effect of past data is rarely constant. Some past values may matter more than others, depending on the market context (Ghahremani & Nguyen, 2025; Zhao & Yan, 2024).

Third, multimodal models can use information that is difficult to express through standard numerical indicators. Exchange rates respond to facts, but they also respond to expectations, interpretations and market narratives. Financial news and public statements can contain signals that numerical data capture only later or incompletely. For this reason, models that combine text and time-series data can gain an advantage (Windsor & Cao, 2022; Mabrouk et al., 2025).

Large language models may add another useful element because they can extract structured signals from large volumes of unstructured text with less manual work. They can help identify sentiment, topics and relevant financial information. Still, this area is at an early stage in exchange rate forecasting. The results should be seen as a promising research direction, not as a settled conclusion (Fatouros et al., 2023; Yang et al., 2025).

Comparative overview

Table 1 presents the main differences between the four forecasting paradigms reviewed in this paper. The table does not rank the methods as universally better or worse. It shows the main strengths and limits that appear in the literature. Classical econometric models remain useful because they are easier to interpret and can work with smaller datasets. Their main weakness is the difficulty of producing stable out-of-sample forecasts. Machine learning models offer more flexibility, especially for nonlinear relations, but their gains are often moderate. Deep learning models are stronger in sequence modeling and learned feature extraction, which helps explain why improvements appear more often in this group. Large language models and transformer-based approaches are promising because they can use richer textual and contextual information, but the evidence is still recent.

Table 1. Comparative summary of forecasting paradigms

Characteristic	Classical econometrics	Machine learning	Deep learning	Large language models and transformers
Typical model families	ARIMA, VAR, GARCH, structural models	tree-based models, support vector machines, shallow neural nets	recurrent networks (LSTM, BiLSTM), attention-based models	transformer-based models for text or time series, GPT-like pipelines
Main strength	interpretability, small-data feasibility	flexible nonlinear mapping	sequence representation and learned features	scalable text processing and attention-based representations
Typical inputs	numerical time series, macro variables	numerical features, engineered signals	numerical time series, optionally engineered features	text and/or numerical sequences
Handling of regime change	explicit modeling, re-estimation	limited unless designed explicitly	can represent changing patterns via learned dynamics	can adapt contextually, depends on design
Interpretability	relatively high	moderate	often limited	often limited
Evidence status in reviewed sample	strong benchmarks, persistent difficulty	improvements reported, often incremental	improvements reported more frequently	emerging and recent evidence

Source: Authors' synthesis based on the systematic review of 15 peer-reviewed publications.

Note: LSTM = Long Short-Term Memory; BiLSTM = Bidirectional Long Short-Term Memory; ARIMA = Autoregressive Integrated Moving Average; GARCH = Generalized Autoregressive Conditional Heteroskedasticity; VAR = Vector Autoregression.

As Table 1 shows, the advantage of artificial intelligence methods differs from one category to another. Improvements appear more often when models move from fixed numerical structures toward methods that can capture time dependence and use richer information. At the same time, interpretation becomes harder as model complexity increases. This creates an important trade-off. In finance, a model must be accurate, but it also has to be understandable enough for practical use, risk control and model supervision.

Evidence extraction and comparability notes

Table 2 presents the main evidence extracted from the fifteen studies reviewed in this paper. It includes the research stream, the main method, the comparison model, the evaluation measures and the main reported outcome. The purpose of the table is to show how the evidence is distributed across different methodological groups. It also helps explain why the results should be interpreted in relation to each study design.

Some papers test more than one model or more than one setting. In such cases, the study may contribute several comparison points to the overall discussion. The results still have to be read within the limits of each paper, because the studies differ in data construction, baseline choice, forecast horizon and reporting style. For this reason, the reported results are not treated as directly interchangeable numerical effects. The table supports a pattern-based comparison by showing where positive results appear more often and where the evidence is still limited.

Table 2. Study-by-study evidence map

Study	Year	Stream	Main method	Baseline/ comparator (as reported)	Metrics reported (examples)	Summary of reported outcome
Meese & Rogoff	1983	Econometrics	structural/econometric vs random walk	random walk	forecast error measures	econometric models often do not beat random walk out of sample
Rossi	2013	Econometrics (review)	synthesis of predictability evidence	random walk and other benchmarks	review-level synthesis	predictability is limited and unstable across settings
Jaworski	2018	Econometrics	Monte Carlo density forecasts with regime switching	econometric baselines	density/foreca st evaluation	richer econometrics improves realism, gains remain limited
Nazareth & Reddy	2023	Machine learning (review)	survey of ML in finance	econometric/s tatistical baselines	error and classification metrics	ML often improves performance in financial prediction tasks
Maneejuk & Srichaikul	2021	Machine learning	ML models for FX forecasting	statistical baselines	error metrics	ML improvements reported in tested settings
Bormpotsis et al.	2023	Machine learning / neural	modular bio-inspired neural network	statistical methods	error metrics	neural approach reports improved forecasting accuracy
Dautel et al.	2020	Deep learning	deep recurrent neural networks	statistical/econometric baselines	error metrics	deep recurrent models report forecasting improvements
Liu et al.	2024	Deep learning hybrid	LASSO + BiLSTM ensemble	econometric and ML baselines	error metrics	hybrid design reports improved performance in tested setup
Chi et al.	2025	Deep learning	feature-augmented multivariate LSTM	econometric baselines	error metrics, directional metrics	improvements reported under real- time data conditions
Ghahremani & Nguyen	2025	Deep learning	attention-based LSTM	baselines in study	error metrics	attention-based model reports gains in tested setting
Windsor & Cao	2022	Multimodal deep learning	multimodal fusion (numerical + text)	numerical- only models	error metrics	multimodal integration reports improved forecasting accuracy
Mabrouk et al.	2025	Text- informed deep learning	semi-supervised deep learning using news	baselines in study	error metrics	news-informed design reports added predictive value
Zhao & Yan	2024	Transformers (time series)	Transformer, Informer, Temporal Fusion Transformer	baselines in study	error metrics	transformer-based models compared for FX forecasting

Study	Year	Stream	Main method	Baseline/ comparator (as reported)	Metrics reported (examples)	Summary of reported outcome
Fatouros et al.	2023	LLM (finance NLP)	ChatGPT in financial sentiment workflows	traditional NLP methods	NLP evaluation metrics	discusses performance and implications for finance NLP
Yang et al.	2025	LLM (FX)	ChatGPT-based text processing for FX analysis/prediction	alternative sentiment or NLP baselines	NLP and forecasting metrics	reports benefits of LLM-derived text signals in tested designs

Source: authors' extraction from the publications listed in References.

Taken together, Table 2 supports three main points. First, deep learning provides the most repeated evidence of improved forecasting results, especially when models are built to capture sequence structure and changing time dependence. Second, text-based information is often linked to additional gains, which suggests that a wider information set can improve forecasting performance. Third, large language model-related research is clearly growing, but it is still too new and too varied to support a strong general conclusion. Therefore, the evidence supports the main argument of the paper, while also showing why the results have to be interpreted carefully.

Conclusion

This paper analyzed whether forecasting systems based on artificial intelligence can improve the prediction of foreign exchange rates when compared with classical econometric models. The review was based on fifteen peer-reviewed studies indexed in Web of Science. The main finding is that artificial intelligence methods are often linked to better forecasting results than classical econometric benchmarks. This is especially visible in studies that use deep learning, multimodal models, transformers and methods connected to large language models.

Still, these results should be interpreted carefully. The advantages are not the same in all cases. They depend on the currency pair, the forecast horizon, the period analyzed, the type of data used and the benchmark chosen for comparison. Therefore, the review does not show that artificial intelligence fully solves the problem of exchange rate forecasting. It shows instead that these methods offer a wider and often more useful set of tools than econometric models with fixed forms, especially in many of the tested settings.

One important conclusion is that classical econometric models continue to face clear forecasting limits. Since the study of Meese and Rogoff (1983), researchers have repeatedly shown that it is difficult to beat simple benchmarks such as the random walk. Later work also confirms that exchange rate predictability is often weak and depends heavily on the context, the period and the currency analyzed (Rossi, 2013). Even so, classical econometrics remains important. It provides the main reference point for testing newer models. In this sense, classical models should not be treated as irrelevant. They are still useful because they offer interpretation, economic logic and a standard against which more complex methods can be judged.

The reviewed evidence suggests that artificial intelligence models become especially useful when exchange rate movements are nonlinear, when relations between variables change over time, or when important information is difficult to capture through standard macroeconomic and financial indicators. Machine learning methods can bring improvements because they are more

flexible than many traditional models. Deep learning methods tend to show stronger results because they are better suited to time-series data and can learn patterns that change across time. Models that use text and numerical data together can also bring useful information from news, sentiment and market commentary, which may influence exchange rates but are hard to measure through numerical variables alone (Windsor & Cao, 2022; Liu et al., 2024).

The paper also contributes by organizing the reviewed studies into four main groups: classical econometrics, machine learning, deep learning and methods based on transformers or large language models. This structure helps clarify the main pattern in the literature. Machine learning usually brings moderate improvements. Deep learning is more often linked to stronger gains. Multimodal models add value by combining numerical data with text. The most recent studies on transformers and large language models show potential, especially for extracting signals from unstructured financial text. However, this area is still new and needs more testing before stronger conclusions can be drawn (Fatouros et al., 2023; Yang et al., 2025).

Another conclusion concerns the difficulty of comparing the existing studies. The reviewed papers use different datasets, currencies, periods, forecast horizons, benchmarks and performance measures. Because of this, the evidence is clearer as a general pattern than as a fully standardized proof. Future research should use more common benchmark datasets, clearer econometric and machine learning baselines, similar reporting rules and stronger robustness checks. It would also be useful to test models across both normal market periods and crisis periods. Without these steps, it remains difficult to know whether better results come from the model itself or from the specific design of the study.

The findings also matter for practical forecasting. Artificial intelligence methods deserve attention in financial applications, especially when the task involves nonlinear movements, time dependence or text-based information. However, a model should not be judged only by lower forecast errors. In real financial settings, the model also has to be understandable, reliable and usable under operational constraints. Predictions must remain useful after considering transaction costs, implementation limits and model risk. For this reason, accuracy is important, but it is not the only condition for practical use.

This study also has limitations. First, the review includes only English-language publications indexed in Web of Science, so relevant studies from other databases or languages may have been left out. Second, the reviewed papers are too different in terms of data, model design and evaluation methods to allow a direct statistical comparison. Third, publication bias may influence the overall picture, since studies with positive results are more likely to be published. Finally, the literature on large language models in foreign exchange forecasting is still limited and recent. Any conclusion about their long-term value should therefore be treated with caution (Fatouros et al., 2023; Yang et al., 2025).

Overall, the paper concludes that artificial intelligence-based forecasting systems show real potential to improve foreign exchange predictability compared with classical econometric models. Their advantage appears stronger when they are used for nonlinear relations, time-based patterns and information environments that include text. Even so, the literature does not yet support the idea that one single method is always superior. Further progress will depend on better comparison standards, clearer benchmarks and careful testing of whether the reported improvements remain stable, understandable and economically useful across different foreign exchange settings.

Acknowledgments:

ERASMUS-EDU-2023-EUR-UNIV, Project 101124676 - EELISA, funded by the European Union, <https://eelisa.eu/>. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them. The authors also acknowledge the use of DeepL version: Translate and for the translation of the manuscript from Romanian to English, last accessed on April 20, 2026. The authors meticulously reviewed and edited the content to ensure accuracy and take full responsibility for the final manuscript.

References

- Bormpotsis, C., Sedky, M., & Patel, A. (2023). Predicting forex currency fluctuations using a novel bio-inspired modular neural network. *Big Data and Cognitive Computing*, 7(3), 152.
- Chi, D. T. K., Kien, H. N. T., & Nguyen, T. Q. (2025). Enhancing forex market forecasting with feature-augmented multivariate LSTM models using real-time data. *Knowledge-Based Systems*, 287, 110238.
- Dautel, A. J., Härdle, W. K., Lessmann, S., & Seow, H.-V. (2020). Forex exchange rate forecasting using deep recurrent neural networks. *Digital Finance*, 2, 69–96.
- Fatouros, G., Soldatos, J., Kyriazis, D., Koumoulis, N., Papanikolaou, A., Alexandris, G., & Stathis, A. (2023). Transforming sentiment analysis in the financial domain with ChatGPT. *Machine Learning with Applications*, 14, 100498.
- Ghahremani, S., & Nguyen, U. T. (2025). Prediction of foreign currency exchange rates using an attention-based long short-term memory network. *Machine Learning with Applications*, 20, 100648.
- Jaworski, K. (2018). Density forecasts of emerging markets exchange rates using Monte Carlo simulation with regime switching. In *Contemporary Trends and Challenges in Finance*, pp. 13–21.
- Liu, S., Huang, Q., Li, M., & Wei, Y. (2024). A new LASSO-BiLSTM-based ensemble learning approach for exchange rate forecasting. *Engineering Applications of Artificial Intelligence*, 127, 107305.
- Mabrouk, A., Abdallah, A., Sedky, M., & Alsharif, M. H. (2025). Semi-supervised deep learning framework for forex forecasting using financial news. *Expert Systems with Applications*, 238, 121981.
- Maneejuk, P., & Srichaikul, W. (2021). Forecasting foreign exchange markets: Further evidence using machine learning models. *Soft Computing*, 25, 7887–7898.
- Meese, R., & Rogoff, K. (1983). Exchange rate models do not outperform a random walk. *Journal of International Economics*, 14(1–2), 3–24.
- Nazareth, N., & Reddy, Y. V. R. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications*, 213, 118857.
- Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, 51(4), 1063–1119.
- Windsor, E., & Cao, W. (2022). Improving exchange rate forecasting via a new deep multimodal fusion model. *Applied Intelligence*, 52(14), 16701–16717.
- Yang, K., Deng, R. X., Wei, Y., & Wang, S. (2025). The power of ChatGPT in processing text: Evidence from analysis and prediction in the exchange rate markets. *Financial Innovation*, 11(1), 78.
- Zhao, J., & Yan, H. (2024). Prediction of currency exchange rate based on Transformer, Informer and Temporal Fusion Transformer. *Journal of Risk and Financial Management*, 17(8), 332.