



Article

ForExAI: Time Series Inference and News Article Analysis Lead to Profitable Foreign Exchange Trading

Beakal Lemeneh ^{1,†}, Eli Hadad ^{2,†} , Allen George Ajith ^{3,†}, Yanbo Hou ^{4,†}, Charlie Zha ^{5,†}, Massimo Scarozza ^{6,†}, Zakaria Baannou ^{7,†}, Ermiyas Liyeh ^{8,†}, Anthony Tomasic, and Dennis Shasha ^{9,*,†} 

¹ NYU Langone Health, 550 1st Ave, New York, NY 10016 USA; Beakal.Lemeneh@nyulangone.org

² Universidade Presbiteriana Mackenzie - Centro de Ciências Sociais Aplicadas, Rua da Consolação 930, São Paulo, SP - 01302-907 - Brazil; eli.hadad@mackenzie.br

³ Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, NY 10012 USA; allen.ajith@nyu.edu

⁴ University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1 CANADA; y82hou@uwaterloo.ca

⁵ Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, NY 10012 USA; lz2609@nyu.edu

⁶ Baruch College, 55 Lexington Ave, New York, NY 10010 USA; massimo.scarozza@baruchmail.cuny.edu

⁷ International University of Rabat, Technopolis Rabat-Shore, Rabat-Salé Ring Road, 11000 Sala el Jadida, Morocco; zakaria.baannou@uir.ac.ma

⁸ University of Rochester, 500 Joseph C. Wilson Blvd, Rochester, NY 14627 USA; eliyeh@u.rochester.edu

⁹ Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213 USA; tomasic@cs.cmu.edu

¹⁰ Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, NY 10012 USA; des1@nyu.edu

* Corresponding: des1@nyu.edu

Abstract: Forecasting foreign exchange rates over long time periods depends on economic fundamentals. Short-term predictions, by contrast, depend largely on emotions, governmental announcements, the flow of capital, and expert commentary. This paper proposes a suite of novel methods, collectively referred to as *ForExAI*, to predict foreign exchange rates based on time series analysis and news article analysis. The time series analysis is based on classical statistical time series techniques, such as ARIMA, as well as machine learning methods using neural networks. Separately, *ForExAI* uses Large Language Models to analyze news articles based on two kinds of prompts: (i) expert-written based on econometric considerations, (ii) existing prompts documented in the literature. Our findings on time series of exchange rates indicate that there are signals in the time series that can be captured even by simple methods like ARIMA(1,1,1), as well as novel machine learning methods on the time series of foreign exchange rate trades. Further, and this is our main conceptual contribution, a judicious use of the Kelly criterion can enhance both profits and the Sharpe ratio. Finally, an ensemble approach, in some cases, delivers slightly higher profit and also lower volatility, leading to a higher Sharpe ratio. Regarding news article analysis, shorter prompts yield far better results than complex ones derived from expert knowledge. The contribution of this work is to show which methods work to forecast complex time series, as well as the relative virtues of different versions of the Kelly criterion and ensemble approaches. The results make a pragmatic contribution as well. While we measure profits by ignoring transaction costs, one implication is that these algorithms and workflows could be directly deployed by market-makers, since market-makers avoid transaction costs. In addition, our results point to further areas of research for traders.

Keywords: high-frequency trading; foreign exchange; forecasting; transformers; N-BEATS; N-HiTS; ARIMA; Large Language Models; Expert prompts; Zero-shot learning; Event-driven algorithmic trading

JEL Classification: C45; C53; C63; G12



Citation: Lemeneh, B.; Hadad, E.; Ajith, A.G.; Hou, Y.; Zha, C.; Scarozza, M.; Baannou, Z.; Liyeh, E.; Tomasic, A.; Shasha, D. *ForExAI: Time Series Inference and News Article Analysis Lead to Profitable Foreign Exchange Trading*. *Mathematics* **2026**, *1*, 0. <https://doi.org/>

Academic Editor(s): Name

Received: 29 April 2024

Revised: 1 June 2024

Accepted: 4 June 2024

Published:



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Foreign exchange (FX) markets move more than 7.5 trillion United States Dollars (USD) each day, making them the most liquid but also one of the most volatile financial

domains in the world [1]. Price dynamics are driven not only by their own autocorrelated technical patterns but also by exogenous macroeconomic events that are reported through global news outlets.

Statistical forecasting methods such as ARIMA [2] and neural networks capture autoregressive patterns yet ignore the semantic signals embedded in breaking headlines. In contrast, natural language processing (NLP) models that infer sentiment from news or social media often react after large moves have already materialized, so it is a struggle to convert their sentiment to profitable trades.

Motivated by these complementary shortcomings, we present *ForExAI*, a suite of forecasting frameworks that models both univariate time series inference and real-time news article analysis to predict the direction of profitable trades.

We study the currency pairs USD/BRL (US dollar vs. Brazilian Real) and USD/CNY (US dollar vs. Chinese Yuan) markets over a 12-month period testing period. Note that the Brazilian Real trades relatively freely in foreign exchange markets, whereas the Yuan is relatively more regulated [3]. Our results show that *ForExAI* delivers statistically significant gains in total profit and profit-per-trade, for both exchange rates, assuming no transaction costs. Such results are useful both for further investigation for traders and for market makers because they do not have to pay transaction costs.

1.1. Contributions of *ForExAI*

As a system, *ForExAI* makes the following contributions:

- Studies the behavior and profitability of several high-capacity LLM-based time series models, several neural network-based forecasters for time series, mean reversion and trend techniques, as well as several Large Language Models for news sentiment,
- Unifies these approaches through an ensemble model that explicitly learns each model's contribution,
- Assesses the performance taking positions based on two variants of the Kelly criterion as well fixed positions, and
- Compares and contrasts performance with state-of-the-art Natural Language systems for which code is available.

This paper describes the design choices made in the construction of *ForExAI* and experimentally documents the impact of these choices on trading performance.

2. Base Time Series Methods

This section describes the forecasting and classification methods used to generate predictive signals for the trading strategies detailed in Section 5.2, based solely on time series data (i.e., without looking at news) consisting of sequences of exchange rates. We use several time series forecasting models that learn the next-step *mid-price* from some subsequence of prices up to the current time: classical econometric approaches such as ARIMA, (ii) neural network architectures (N-BEATS, N-HiTS, and TCN); and (iii) large pre-trained foundation models (Chronos-Bolt and Toto) applied in a zero-shot forecasting setting without task-specific fine-tuning. The predicted mid-prices from these models serve as the basis for the model-driven trading strategies (Section 5.2), which use the direction of the predicted price movement to determine whether to buy, sell, or hold a currency position.

2.1. Classical Forecasting Models

Traditional econometric models such as ARIMA and GARCH have served as foundational tools for modeling the autocorrelated structure of financial time series [2]. These models excel at capturing linear dependencies and volatility clustering, respectively, but struggle to generalize in the presence of non-stationarity or regime shifts. When embedded in trading strategies, forecasts from an ARIMA(1,1,1) specification typically do not generate profits that are statistically different from random trading, reflecting the near-martingale behavior of financial results.

The Autoregressive Fractionally Integrated Moving Average (ARFIMA) model [4] [5] [6] is an extension of the standard ARIMA class of models, specifically designed to capture long memory and persistence in time series data. While traditional ARMA models are suited for short-memory processes where the influence of past shocks decays exponentially, and ARIMA models handle non-stationarity through integer differencing, many series in finance and economics exhibit a persistence that is stronger than an ARMA process but is still mean-reverting, unlike a unit-root process.

In ARFIMA models with $0 < d < 0.5$, convergence to the unconditional mean occurs slowly due to long memory. Therefore, although ARFIMA forecasts may display greater persistence at short and medium horizons, dynamic multi-step predictions eventually become flat, similarly to dynamic ARIMA models. This behavior is a direct consequence of the long-memory structure described in [4], [5] and [7], rather than a limitation of estimation or model specification. For these reasons, we use a static ARIMA model only.

2.2. Neural Network-based Forecasting Models

We consider two categories of neural network-based forecasting models: (i) models that are trained on the training dataset of the time series dataset, and (ii) large pre-trained foundation models applied in a zero-shot setting without task-specific retraining (though with hyperparameter selection).

2.2.1. Models Requiring Training

The models requiring training include N-BEATS, N-HiTS, and the Temporal Convolution Network (TCN). Our experimental methodology replicates our previous work [8], but we briefly summarize it here for completeness.

N-BEATS: N-BEATS [9] is a deep feedforward forecasting model built from a sequence of residual blocks that iteratively refine predictions. Each block produces two outputs: a “backcast,” which reconstructs the input window, and a “forecast,” which contributes to the final prediction [9]. This residual design helps isolate trend and seasonal components, enabling the model to progressively capture unexplained patterns while maintaining interpretability and strong predictive accuracy [9]. Our implementation uses one block and two stacks, each block consisting of five fully connected layers of width 512 with dropout (0.2).

N-HiTS: N-HiTS [10] extends N-BEATS by introducing a hierarchical structure that models temporal patterns at multiple resolutions. By analyzing the time series at coarse and fine scales, the model captures both short-term fluctuations and longer-term dynamics [10]. It also incorporates interpolation layers to enhance robustness and smoothness in the forecasts [10]. Our implementation has one block and two stacks of five layers each, using feed-forward layers of width 512 and dropout 0.2.

Temporal Convolution Network (TCN): TCNs [11–13] apply one-dimensional convolutions in a causal and dilated fashion, ensuring that each prediction depends only on past inputs while efficiently covering long time horizons. Residual and skip connections stabilize deep architectures, allowing the model to learn complex temporal dependencies without gradient degradation [11–13]. Our implementation uses eight convolutional layers with 64 filters and a kernel size of 3. A dilation base of 2 and weight normalization are applied, with dropout 0.2 for regularization.

2.3. Zero-Shot Models

In contrast to the trained models, Chronos-Bolt and Toto are applied in a zero-shot forecasting setting. These models are large-scale, pre-trained time series foundation models, and we evaluate their out-of-the-box generalization to our data without fine-tuning.

2.3.1. Chronos-Bolt

As described in detail in the related work section Section 15.2, Chronos-Bolt [14] represents an evolution over the original Chronos framework (the Chronos authors request

that the same paper reference be used for both Chronos and Chronos-Bolt) and serves as our initial benchmark in this category. In our experiments, we have used the 204M parameter Transformer-based model.

We ingest raw mid-price sequences, delegating the handling of non-stationarity to the model's intrinsic Instance Normalization layer. Given a historical context window C , the length of which is optimized via hyperparameter search in Section 10, the input is standardized using its local moments. Consequently, the model operates within a normalized space defined by: $\hat{r} = \frac{r - \mu_{context}}{\sigma_{context}}$ where $\mu_{context}$ and $\sigma_{context}$ represent, respectively, the mean and standard deviation of the input window (which is the time period and data that Chronos-Bolt looks at to make a prediction). This adaptive standardization enables the handling of local volatility without need for external scaling.

Chronos-Bolt generates probabilistic forecasts via a regression-based head, targeting a specific set of quantile levels defined during training $\{0.10, 0.20, \dots, 0.90\}$. This allows predictions to extend beyond the observed input range when supported by the underlying trend. For our trading signal, we isolate the median forecast (0.50 quantile) as the deterministic prediction, \hat{r}_{pred} . To recover the final price estimate, the inverse affine transformation is applied using the context statistics: $r_{pred} = \hat{r}_{pred} \cdot \sigma_{context} + \mu_{context}$. Recall that $\mu_{context}$ and $\sigma_{context}$ are derived exclusively from the input window, thus can change as the time series evolves.

2.3.2. Toto Foundation Model for Financial Time Series Forecasting

Toto [15] is a large-scale 115M parameter foundation model for time-series forecasting that, like Chronos-Bolt [14], operates in a zero-shot setting. Toto is designed to generalize across diverse temporal domains without any task-specific fine-tuning. As such, it employs an internal instance normalization strategy similar to that of Chronos-Bolt to handle distributional shifts. Consequently, we provide raw mid-price sequences to the model, relying on its internal mechanism to standardize inputs based on the local context statistics. This approach ensures that the model adapts to the specific volatility of the input window without requiring training set access.

During inference, Toto generates multiple probabilistic samples for each forecast point. In our experiments, we set the sampling parameter to 64 time steps. We then used the median prediction across samples as the final point forecast. We set the time interval to 60 seconds to reflect the one-minute spacing in our forex data.

3. Event-Based News-Reading Forecasting Methods

3.1. News Article Classification Prompts for the Chinese Yuan vs. U.S. Dollar

There are three news prediction prompts for the Chinese Yuan (CNY) vs. U.S. Dollar (USD): a Naive prompt, a Naive+ prompt, and an Expert prompt. All prompts ask the LLM to classify a news article into three categories:

- $y = 1$: Expected CNY appreciation (exchange rate of CNY against USD increase)
- $y = -1$: Expected CNY depreciation (exchange rate of CNY against USD decrease)
- $y = 0$: Neutral sentiment or conflicting signals

The Naive prompt specifies the minimal prompt to predict an article's impact on an exchange rate. This prompt serves as a baseline.

Naive Prompt

Provide a sentiment analysis for the following article regarding the specified currency pair. Indicate whether each headline is positive 1, negative -1, or neutral 0 for the specified currency pair. Your response is only 1, 0, or -1

3.1.1. Naive+ Prompt

The Naive+ prompt extends the Naive prompt with a reference to the relevant currency. This prompt allows for comparing and contrasting with the performance of the Naive prompt. Informally, we hypothesize that additional detail in the prompt will improve performance.

Naive+ Prompt

Act as an expert of forex trading holding USD and CNY. Based on the article {article}, will you buy USD, sell USD or hold in the short term? Answer in one token: buy, sell, or hold.

3.1.2. Expert Prompt

The Expert prompt is derived from discussions with one of the co-authors (E.H.), who is an economist. The prompt is designed to operationalize analyst-like behavior. The prompt is tailored specifically to the CNY/USD currency pair, targeting foreign-exchange analysis in offshore Chinese yuan and U.S. dollars. The Expert prompt includes illustrative examples, which are used to shape outputs and to enforce consistent labeling behavior across tasks.

The Expert prompt's construction follows a fixed, explicit sequence. First, a role is specified to establish the model's standpoint and expected voice. Second, a task is defined to delimit the objective and scope of the response. Third, background information is provided in the form of abbreviations to standardize terminology and reduce ambiguity. Finally, multiple examples are appended, including news items mapped to corresponding labels and lexical cues in which specific words trigger specific labels. These elements are then combined to form the Expert prompt, integrating role specification, task definition, abbreviation-based background, and example-driven guidance into a single, deployable instruction.

Expert Prompt

Role Definition

You are an AI analyst for the FX strategy team at JPMorgan Chase, specializing in foreign exchange trading and multi-market event analysis. Your task is to evaluate news texts immediately, determine their impact on the short-term CNY/USD movement, and return the final impact along with a brief transmission chain explanation. Your analysis should focus on global capital flows, U.S.–China policy dynamics, and cross-market arbitrage perspectives.

Task Requirements

- Analyze the keywords in the news and classify them under the factors listed below.
- Based on the identified keywords and their corresponding transmission paths, determine the impact on CNY:
 - 1 (beneficial for CNY)
 - -1 (detrimental for CNY)
 - 0 (unclear signal or insufficient information)
- The output must be a single line JSON with one field:
 - "impact": the final judgment
- Example output format:


```
{"impact": "1"}
```

Abbreviations

- GDP: Gross Domestic Product
- CPI: Consumer Price Index
- PPI: Producer Price Index

- PMI: Purchasing Managers' Index
- NFP: Non-Farm Payrolls
- VIX: Volatility Index
- Fed: Federal Reserve
- PBoC: People's Bank of China
- RRR: Reserve Requirement Ratio
- FOMC: Federal Open Market Committee
- QE: Quantitative Easing
- QT: Quantitative Tightening

Classification and Labels

General Economic Indicators

- China GDP Increase → label: 1
- China GDP Decrease → label: -1
- Stable China GDP → label: 0
- U.S. GDP Increase → label: -1
- U.S. GDP Decrease → label: 1
- Stable U.S. GDP → label: 0
- China CPI/PPI Higher than Expected → label: -1
- China CPI/PPI Lower than Expected → label: 1
- U.S. CPI Higher than Expected → label: 1
- U.S. CPI Lower than Expected → label: -1
- China PMI > 50 → label: 1
- China PMI < 50 → label: -1
- China PMI = 50 → label: 0

Central Bank & Monetary Policy

- Fed Rate Hike → label: -1
- Fed Rate Cut → label: 1
- PBoC Rate Hike → label: 1
- PBoC Rate Cut → label: -1
- China RRR Cut → label: -1
- China RRR Raise → label: 1
- Fed QE → label: 1
- Fed QT → label: -1

Currency Market Dynamics

- CNY/USD breaks support (CNY appreciates) → label: 1
- CNY/USD breaks resistance (CNY depreciates) → label: -1
- PBoC sets Yuan Fixing stronger than expected → label: 1
- PBoC sets Yuan Fixing weaker than expected → label: -1

U.S. Specific Economic Terms

- Strong NFP → label: -1
- Weak NFP → label: 1
- FOMC Hawkish → label: -1
- FOMC Dovish → label: 1

China Specific Economic Terms

- China Stimulus → label: -1
- China Property Crisis → label: -1
- Capital Outflow Control Tightened → label: 1
- Capital Control Relaxed → label: -1

Geopolitical & Trade Relations

- Trade Tensions Escalate → label: -1
- Trade Agreement Reached → label: 1
- U.S. Tariffs Increased → label: -1
- U.S. Tariffs Lowered → label: 1

Global Economic Context

- Global Slowdown → label: -1
- Global Rebound → label: 1
- High Oil Prices → label: -1
- Low Oil Prices → label: 1

Market Sentiment & Risk Indicators

- Risk-off Sentiment → label: -1
- Risk-on Sentiment → label: 1
- VIX Spikes → label: -1
- VIX Falls → label: 1

Output Format

- Output must be a valid JSON string:
`{"impact": "1"}`
- Only the `impact` field is allowed
- Values: "1", "-1", or "0"

Examples

- News: Offshore Yuan Climbs to Over a 1-Year High...
Output: `{"impact": "1"}`
- News: PBoC to cut RRR by 50bps...
Output: `{"impact": "-1"}`
- News: Trump administration likely to worsen US-China ties...
Output: `{"impact": "-1"}`
- News: Investors await Thursday's GDP data...
Output: `{"impact": "0"}`

Note that these three prompts (Naive, Naive+, Expert) are used to predict the influence of single news articles on exchange rates. In Section 11, we use training data to tune the hold period hyperparameter (in minutes) for fixed-position size trading. Based on this analysis, we select the best-performing news event model for the USD/CNY exchange rate.

3.2. News Article Classification for the Brazilian Real vs. U.S. Dollar

Focusing on the Brazilian Real (BRL)/United States Dollar (USD) exchange rate, we formulate sentiment classification as a three-class problem predicting exchange rate movement direction (up, down, or neutral) given news articles. We assess performance through profitability analysis, comparing trading strategies based on model predictions against market benchmarks.

Given a news article, we use an LLM to infer the expected direction of BRL movement. We formulate this as a three-class classification problem:

$$f_{\text{LLM}}(a) \rightarrow y \in \{\text{Down}, \text{Neutral}, \text{Up}\} \quad (1)$$

where the predicted classes correspond to expected BRL/USD exchange rate movements:

- **Up:** Expected BRL appreciation (USD/BRL rate decreases)
- **Down:** Expected BRL depreciation (USD/BRL rate increases)
- **Neutral:** No significant movement or conflicting signals

3.2.1. Choice of LLM Architecture

In our initial experimentation, we evaluated several LLM architectures, including Llama 3.3 70B Instruct, Mistral Nemo Instruct 2407, and Gemini 2.5 Flash, assessing them across classification accuracy, computational efficiency, and output consistency. We adopted Gemini 2.5 Flash for all experiments based on its strong performance in preliminary evaluations on training data and practical advantages, including API accessibility and ease of deployment.

3.2.2. Prompt Engineering Methodology

We design a systematic prompt engineering strategy developed in collaboration with a financial expert (co-author E.H.), progressing through three distinct approaches with increasing levels of task specification and domain knowledge:

Basic Zero-Shot (P0):

Simple prompts establishing the basic role and task.

Zero Shot Prompt

You are an analyst working to get market movement signals for the BRL/USD foreign exchange market.

Analyze the provided Portuguese news content and assign a sentiment score:

- +1 indicates that BRL is expected to rise (BRL strengthens vs USD)
- -1 indicates that BRL is expected to fall (BRL weakens vs USD)
- 0 indicates neutral sentiment, insignificant or conflicting signals

News Heading: {heading}

News Content: {content}

Analyze step-by-step, then conclude:

- Reasoning: [Analyze the key economic factors and their likely impact on BRL/USD]
- Sentiment: [your score: -1, 0, or +1]

Expert-Guided Zero-Shot (P1):

Comprehensive prompts developed with domain expert input, defining the classification task with financial criteria. These criteria include market drivers such as US Federal Reserve policy, Brazilian fiscal measures, trade balance, and commodity factors, with specific rules for each variable.

Expert-Guided Zero-Shot

You are a sentiment classifier for the foreign exchange market, focusing on BRL/USD. Analyze the provided news input and assign a sentiment score based on actionable insights:

- +1 indicates that BRL is expected to rise.
- -1 indicates that BRL is expected to fall.
- 0 indicates neutral sentiment or conflict in news signals.

Base your assessment on the following actionable insights:

Actionable Insights:

- If US Fed policy signals hawkishness (e.g., interest rate hikes or tightening liquidity), then BRL goes down. (Score: -1)
- If US Fed policy signals dovishness (e.g., interest rate cuts or easing monetary policy), then BRL goes up. (Score: +1)
- If US economic indicators (GDP, unemployment, CPI) are robust and stable, then BRL goes up. (Score: +1)
- If US economic indicators show weakness, such as high unemployment or rising CPI (inflation), then BRL goes down. (Score: -1)

- If Brazil implements proactive fiscal measures (active revenue measures, effective policy negotiations), then BRL goes up. (Score: +1)
- If Brazil shows fiscal indiscipline (policy abandonment, lack of compensatory measures, increased spending), then BRL goes down. (Score: -1)
- If the trade balance and commodity prices are positive (e.g., strong surplus, rising commodity prices), then BRL goes up. (Score: +1)
- If the trade balance and commodity prices are negative (e.g., deficits, falling commodity prices), then BRL goes down. (Score: -1)
- Better than expected Brazil GDP growth announcement: +1
- Worse than expected Brazil GDP growth announcement: -1
- If the Brazilian real depreciates sharply against major currencies, then investor confidence falls and BRL goes down. (Score: -1)
- If there is increased political risk or policy uncertainty, then sentiment turns negative and BRL goes down. (Score: -1)
- If the Central Bank signals a strong, credible commitment to anchor inflation and tighten policy effectively, then BRL goes up. (Score: +1) Otherwise, if its credibility is questioned, then BRL goes down. (Score: -1)
- If fiscal deficits widen or public debt becomes unsustainable, then BRL is expected to fall. (Score: -1)
- If foreign direct investment slows or capital outflows increase, then BRL goes down. (Score: -1)
- If multiple actionable insights conflict or signals are mixed, then neutral sentiment. (Score: 0)

Multiple Indicators:

- A news article may contain multiple indicators pointing to different sentiment signals
- If multiple indicators point in the same direction, this increases confidence in the sentiment score
- When reporting sentiment, list all applicable indicators and their individual scores before providing the final sentiment assessment

Key Market Drivers (for Actionable Insights):

- **1. US/External Factors:** Fed rates/policy changes (hawkish = -1; dovish = +1; mixed = 0); US economic indicators (GDP, unemployment, CPI; strong = +1; weak = -1)
- **2. Brazil Fiscal Measures:** Proactive fiscal actions → +1; Fiscal indiscipline → -1
- **3. Trade & Commodity Factors:** Positive trade balance and rising commodity prices → +1; Negative trade balance and falling commodity prices → -1
- **4. Additional Brazil-Specific Factors:** Sharp depreciation of the real → -1; Increased political risk → -1; Central Bank Credibility (Credible → +1, Lack thereof → -1); Widening deficit → -1; Slowing FDI → -1

News Heading: {heading}

News Content: {content}

Analyze step-by-step, then conclude:

- Reasoning: [Analyze the key economic factors and their likely impact on BRL/USD]
- Sentiment: [your score: -1, 0, or +1]

Expert Framework with Examples (P2):

The expert-guided prompt enhanced with few-shot learning using curated examples per class selected to demonstrate the application of the actionable insights across diverse market scenarios, including fiscal policy changes, monetary policy shifts, and commodity price movements. After initial experimentation on the training data with 1, 3, and 5 examples per class, we selected 1 example per class as it best balanced performance with minimal computational overhead.

Expert Framework with one example per class

You are a sentiment classifier for the foreign exchange market, focusing on BRL/USD. Analyze the provided news input and assign a sentiment score based on actionable insights:

- +1 indicates that BRL is expected to rise.
- -1 indicates that BRL is expected to fall.
- 0 indicates neutral sentiment or conflict in news signals.

Base your assessment on the following actionable insights:

Actionable Insights:

- If US Fed policy signals hawkishness (e.g., interest rate hikes or tightening liquidity), then BRL goes down. (Score: -1)
- If US Fed policy signals dovishness (e.g., interest rate cuts or easing monetary policy), then BRL goes up. (Score: +1)
- If US economic indicators (GDP, unemployment, CPI) are robust and stable, then BRL goes up. (Score: +1)
- If US economic indicators show weakness, such as high unemployment or rising CPI (inflation), then BRL goes down. (Score: -1)
- If Brazil implements proactive fiscal measures (active revenue measures, effective policy negotiations), then BRL goes up. (Score: +1)
- If Brazil shows fiscal indiscipline (policy abandonment, lack of compensatory measures, increased spending), then BRL goes down. (Score: -1)
- If the trade balance and commodity prices are positive (e.g., strong surplus, rising commodity prices), then BRL goes up. (Score: +1)
- If the trade balance and commodity prices are negative (e.g., deficits, falling commodity prices), then BRL goes down. (Score: -1)
- Better than expected Brazil GDP growth announcement: +1
- Worse than expected Brazil GDP growth announcement: -1
- If the Brazilian real depreciates sharply against major currencies, then investor confidence falls and BRL goes down. (Score: -1)
- If there is increased political risk or policy uncertainty, then sentiment turns negative and BRL goes down. (Score: -1)
- If the Central Bank signals a strong, credible commitment to anchor inflation and tighten policy effectively, then BRL goes up. (Score: +1) Otherwise, if its credibility is questioned, then BRL goes down. (Score: -1)
- If fiscal deficits widen or public debt becomes unsustainable, then BRL is expected to fall. (Score: -1)
- If foreign direct investment slows or capital outflows increase, then BRL goes down. (Score: -1)
- If multiple actionable insights conflict or signals are mixed, then neutral sentiment. (Score: 0)

Multiple Indicators:

- A news article may contain multiple indicators pointing to different sentiment signals
- If multiple indicators point in the same direction, this increases confidence in the sentiment score
- When reporting sentiment, list all applicable indicators and their individual scores before providing the final sentiment assessment

EXAMPLES:

BULLISH (+1):

"Para o acumulado de 2023, o consenso é de crescimento de 3,0% do PIB, com projeções entre 2,8% e 3,2%."

Chain-of-Thought: Strong GDP growth consensus attracts foreign investment and strengthens BRL.

Current: +1

BEARISH (-1):

"Em meio aos impactos da guerra na Ucrânia e de uma possível recessão nos Estados Unidos - com impactos sobre a economia global -, o Banco Central continuou a apertar os cintos na política monetária."

Chain-of-Thought: Global economic uncertainty from Ukraine war and US recession fears creates external pressure requiring tighter monetary policy, typically negative for emerging market currencies.

Current: -1

NEUTRAL (0):

"O Copom reúne-se a cada 45 dias. No primeiro dia do encontro, são feitas apresentações técnicas sobre a evolução e as perspectivas das economias brasileira e mundial."

Chain-of-Thought: Procedural information about central bank meetings without policy implications creates no market expectation changes.

Current: 0

Analyze step-by-step, then conclude:

- Reasoning: [Analyze the key economic factors and their likely impact on BRL/USD]
- Sentiment: [your score: -1, 0, or +1]

Similar to our proposed USD/CNY prompts, in Section 11, we use training data to tune the hold period hyperparameter (in minutes) for fixed-position size trading. Based on this analysis, we select the best-performing news-event model for the USD/BRL exchange rate.

3.3. Comparative News-Event Model Baselines

To establish a comparative baseline for news-event trading, we adopt the prompt engineering framework proposed by Fatouros et al. [16], originally evaluated on FinBERT and GPT-3.5. For our experiments, we employ Llama 3.1 8B Instruct [17] as the generative backbone. While Fatouros et al. utilized six distinct prompts, we restrict our evaluation to the four instance-level (an instance is a single headline) prompts listed below. We exclude the remaining two prompts, which group all the headlines of a given day, because we wanted to compare the approaches on single articles. (In [16], the authors report that no single prompt performed best everywhere, so this exclusion doesn't eliminate their best-performing approach.) During inference, we utilize a temperature setting of 0.20, dynamically substituting the template variables (prefixed with \$) with the corresponding ticker symbol and headline text. In instances where the model output deviates from the expected token vocabulary, the sample is assigned a 'neutral' label.

Prompt 1

Act as a financial expert holding \$ticker. How do you feel about the headline \$headline? Answer in one token: positive, negative, or neutral.

Prompt 2

Act as a financial expert. Classify the sentiment for \$ticker based only on the headline \$headline. Answer in one token: positive, negative, or neutral.

Prompt 3

Act as a sentiment analysis model trained on financial news headlines. Classify the sentiment of the headline \$headline. Answer in one token: positive, negative, or neutral.

Prompt 4

Act as an expert at forex trading holding \$ticker. Based only on the headline \$headline, will you buy, sell or hold \$ticker in the short term? Answer in one token: positive for buy, negative for sell, or neutral for hold position.

A second sentiment analysis approach is the prompt engineering framework proposed by Ballinari and Maly [18]. Their study provides a comparative analysis spanning rule-based lexical systems (Loughran-McDonald, VADER), domain-specific encoders (FinBERT), and the generative Llama 3.1 8B Instruct model, evaluating the latter in both zero-shot and task-specific fine-tuned configurations. For our experiments, we utilize the Llama 3.1 8B Instruct model with its official off-the-shelf weights. Following default settings, inference is conducted at a temperature of 0.6. We dynamically populate the template placeholders (denoted by \$) with the specific ticker, title and body text of each news entry to generate directional labels for the target currency.

Note that, following the usage of comparative models in the literature, the trading strategy for these models is "hold until the close of the day." In contrast, for our prompts in Section 11, we treat the hold period as a hyperparameter and tune it for each prompt.

Ballinari and Maly Prompt

Title: \$title

Text: \$text

Instructions:

Objective: For each mentioned currency, answer the following questions:

- What has been the current/past movement of the currency (appreciation, depreciation, or unchanged)?
- What is the future expectation for the currency (appreciation, depreciation, or unchanged)?

You must answer these two questions for each of the following currencies mentioned in the article:

USD_past: "appreciation, depreciation, or unchanged",

USD_future: "appreciation, depreciation, or unchanged",

\$second_currency_past: "appreciation, depreciation, or unchanged",

\$second_currency_future: "appreciation, depreciation, or unchanged"

Output Format:

- Important: Provide your answer in separate rows for each currency as shown above. Do not combine multiple currencies in the same row.

Each currency should have its own line with "_past" or "_future" specified.

Example:

- If the article states, "The USD is expected to appreciate," the output should be:

USD_past: "unchanged",

USD_future: "appreciation"

- If the article states, "USD/\$second_currency depreciated last week," the output should be:

USD_past: "depreciation",

\$second_currency_past: "appreciation"

- If only future movements are mentioned for a currency, the past movement should be labelled as "unchanged" and vice versa.

Currency Pair Interpretation:

- If currencies are discussed in pairs, interpret as follows:
- If "USD/\$second_currency appreciated," label USD_past as "appreciation" and \$second_currency_past as "depreciation".
- If "USD/\$second_currency depreciated," label USD_past as "depreciation" and \$second_currency_past as "appreciation".

Synonyms:

- Recognize the following synonyms for each currency:

- **EUR**: EUR, Euro

- **USD**: USD, Dollar, Dollars, US Dollar, US-Dollar, U.S. Dollar, US Dollars, US-Dollars, U.S. Dollars, Greenback

- **BRL**: BRL, Brazilian Real, Real, Reais, R\$\$, Pila, Prata, Mango, Pau, Conto, Réis, Brazilian currency.

- **CNY**: CNH, CNY, Chinese Yuan, Yuan, Renminbi, RMB, ¥, CN¥, Redback, Kuài, Offshore Yuan, Offshore Chinese Yuan, Offshore Renminbi, People's Currency.

- **JPY**: JPY, Yen, Japanese Yen

- **GBP**: GBP, Pound, Pounds, Sterling, British Pound, British Pounds

- **AUD**: AUD, Australian Dollar, Australian Dollars, Aussie

- **CAD**: CAD, Canadian Dollar, Canadian Dollars

- **CHF**: CHF, Swiss Franc, Swiss Francs, Swissie

- **NZD**: NZD, New Zealand Dollar, New Zealand Dollars, Kiwi

- **NOK**: NOK, Norwegian Krone, Norwegian Kroner

- **SEK**: SEK, Swedish Krona, Swedish Kronor

4. Datasets

We use three main datasets: foreign exchange price data, financial news articles, and a collection of news articles, a subset of which are labeled by an expert. Prices reflect short-term market dynamics (within a few minutes), while news may take longer to have an effect, and its effect may last longer.

4.1. FX price data

The foreign exchange price data were obtained from <https://massive.com> [19], which provides historical and real-time financial market data. For this study, we focus on two major currency pairs involving the U.S. dollar: USD/CNY (U.S. dollar vs. Chinese Yuan) and USD/BRL (U.S. dollar vs. Brazilian Real).

Both the bid (buy) and ask (sell) prices were collected at each trade to compute the mid-price, as defined in Equation 2. The resulting mid-price series serves as the primary signal for both model training and inference during trading simulations. The data were collected at a resolution of at most one minute.

The overall sample covers the period from January 1, 2024 to July 31, 2025. Specifically, the training dataset spans January 1 – June 30, 2024 (used to train the models described in Section 2.2.1); the validation dataset spans July 1 – July 31, 2024 (used to select the best model through a hyperparameter search over the input context length across all models described in Section 2.2); and the test dataset spans August 1, 2024 – July 31, 2025 (used to run the selected model as a forecaster with the trading strategies described in Section 5 to generate the out-of-sample results presented in Section 13 for both currency pairs). For testing, trading was simulated between 9 a.m. and 5 p.m. U.S. Eastern time, when the markets are most active.

4.2. News Articles

Brazilian Real News Articles: The articles used in our experiments span the period from January 12, 2024, to July 30, 2025, comprising a total of 20,543 articles. These articles were scraped, cleaned, and aggregated from Bom Dia Mercado (Good Morning Market, www.bdm.com.br), a Brazilian financial news wire service. The dataset contains articles with corresponding timestamps at minute-level granularity. The dataset was divided as follows: 769 articles from January 12, 2024 through June 30, 2024 were used for training, and the remaining 19,774 articles in the dataset from August 1, 2024 through July 30, 2025 were used for testing.

Chinese Yuan News Articles: The news articles for our experiment were collected between January 19, 2024, and July 30, 2025. In total, 2,683 news articles for the currency pair U.S. Dollar and Yuan were obtained over this period, with timestamps recorded at the minute level. We divide the dataset into a training set containing 826 articles from January 19 to June 30, 2024, and a test set containing 1,857 articles from August 1st, 2024, to July 30, 2025. The news articles themselves were drawn from the LSE Thomson Reuters newswire service [20].

5. Time Series-based Trading Strategies

Time series-based trading strategies look only at the time series of exchange rates, not at news articles or other quantitative indicators. These strategies act on the *mid-price*, defined as the average of the bid and ask quotes:

$$r_t = \frac{\text{askPrice}_t + \text{bidPrice}_t}{2}, \quad (2)$$

where r_t denotes the mid-price of the USD/CNY or USD/BRL exchange rate at time t , askPrice_t is the ask price at time t , and bidPrice_t is the bid price at time t . Here, Currency A is the base currency (USD) and Currency B is the quote currency (CNY or BRL). A “buy” action corresponds to going long on USD (buying Currency A, selling Currency B), while a “sell” action corresponds to going short on USD (selling Currency A, buying Currency B).

The holding time is one minute. In an environment with transaction costs, holding times may be extended if the current position (e.g., long on currency X and short on currency Y) is still suggested by the model at the end of the holding period.

At each time step t , the fractional change in mid-price is computed as:

$$\Delta_t = \frac{r_t - r_{t-1}}{r_{t-1}}. \quad (3)$$

5.1. Rule-Based Strategies

The strategies Mean Reversion and Trend operate purely on previously observed changes in exchange rates. They rely on rule-based heuristics that predict that exchange rate changes will reverse or that they will continue.

5.1.1. Mean Reversion Strategy

The mean reversion strategy assumes that deviations from recent trends are temporary and prices will revert to their previous states. The trading decision is defined as:

$$d_t = \begin{cases} \text{sell} & \text{if } \Delta_t > \Delta_{t-1}, \\ \text{buy} & \text{if } \Delta_t < \Delta_{t-1}, \\ \text{no trade} & \text{otherwise.} \end{cases} \quad (4)$$

5.1.2. Trend Strategy

The trend-following strategy assumes that price movements will persist in the same direction. The decision rule is the opposite of mean reversion:

$$d_t = \begin{cases} \text{sell} & \text{if } \Delta_t < \Delta_{t-1}, \\ \text{buy} & \text{if } \Delta_t > \Delta_{t-1}, \\ \text{no trade} & \text{otherwise.} \end{cases} \quad (5)$$

5.2. Model-Driven Strategies

The second group of strategies consists of machine learning models as described in Section 2.2, either directly or as part of an ensemble system. These models infer the next mid-price based on historical context.

Formally, given the past n timesteps, a model $f(\cdot)$ forecasts the mid-price at $t + 1$:

$$\hat{r}_{t+1} = f(r_t, r_{t-1}, \dots, r_{t-n+1}), \quad (6)$$

where $f(\cdot)$ corresponds to machine learning (ML) models and large language models (LLMs) adapted for time series forecasting, detailed in Section 2.2, and n represents the input context length of the model.

The predicted percentage change is:

$$\hat{\Delta}_{t+1} = \frac{\hat{r}_{t+1} - r_t}{r_t}. \quad (7)$$

The trading rule is as follows:

$$d_t = \begin{cases} \text{buy} & \text{if } \hat{\Delta}_{t+1} > 0, \\ \text{sell} & \text{if } \hat{\Delta}_{t+1} < 0, \\ \text{no trade} & \text{if } \hat{\Delta}_{t+1} = 0. \end{cases} \quad (8)$$

Note that our method applies no threshold beyond zero, because the predicted percentage changes $\hat{\Delta}_{t+1}$ are typically extremely small. This makes it difficult to define a

meaningful non-zero cutoff. Therefore, any positive predicted change is treated as an upward movement and any negative predicted change as a downward movement:

5.3. Ensemble Strategy

The ensemble strategy consolidates the predictive signals produced by the previously described strategies, using a regression model to infer the next price movement directly from the real-valued percentage changes predicted by each strategy.

At each time step t , let $\hat{\Delta}_{t+1}^{(\text{MR})}$ denote the predicted percentage change generated by the mean reversion strategy. The trend strategy is not included, as its predicted price change is exactly the negative of mean reversion and would therefore add no new information. Let $\hat{\Delta}_{t+1}^{(i)}$ denote the predicted percentage change produced by the i -th forecasting model, where $i \in \{1, \dots, m\}$ and the forecasting models are described in Section 2.2.

The ensemble feature vector at time t is constructed by concatenating these real-valued predictions:

$$\mathbf{x}_t = [\hat{\Delta}_{t+1}^{(\text{MR})}, \hat{\Delta}_{t+1}^{(1)}, \hat{\Delta}_{t+1}^{(2)}, \dots, \hat{\Delta}_{t+1}^{(m)}]. \quad (9)$$

A regression function $g(\cdot)$, implemented as a Ridge Linear Regressor, is trained to map this feature vector to the actual percentage change in mid-price observed at the next timestep:

$$\hat{\Delta}_{t+1}^{(\text{ens})} = g(\mathbf{x}_t). \quad (10)$$

The model is trained using historical data, where the supervision signal is the realized mid-price change:

$$\Delta_{t+1} = \frac{r_{t+1} - r_t}{r_t}, \quad (11)$$

computed using mid-prices outside market hours to avoid lookahead bias.

The resulting predicted change $\hat{\Delta}_{t+1}^{(\text{ens})}$ is then converted into a trading action. As with the model-driven strategies, no explicit threshold is applied because predicted changes are typically extremely small in magnitude:

$$d_t = \begin{cases} \text{buy} & \text{if } \hat{\Delta}_{t+1}^{(\text{ens})} > 0, \\ \text{sell} & \text{if } \hat{\Delta}_{t+1}^{(\text{ens})} < 0, \\ \text{no trade} & \text{if } \hat{\Delta}_{t+1}^{(\text{ens})} = 0. \end{cases} \quad (12)$$

6. Time Series-based Trading: Kelly-Based Position Sizing

The Kelly fraction is computed dynamically to adjust position sizes based on the observed performance of each trading strategy.

For each strategy (e.g., mean reversion, Chronos–Bolt) we maintain and update after every trade:

- *numWins*, *numLosses*: counts of winning and losing trades,
- *totalGains*: sum of positive profits,
- *totalLosses*: sum of absolute losses.

Thus, *numWins*, *numLosses*, *totalGains*, and *totalLosses* all change after each new trade.

From these running statistics, after trade t we estimate:

$$\hat{p}_t = \frac{\text{numWins}_t}{\text{numWins}_t + \text{numLosses}_t}, \quad (13)$$

$$\hat{q}_t = 1 - \hat{p}_t, \quad (14)$$

$$\bar{G}_t = \frac{\text{totalGains}_t}{\text{numWins}_t}, \quad (15)$$

$$\bar{L}_t = \frac{\text{totalLosses}_t}{\text{numLosses}_t}, \quad (16)$$

$$b_t = \frac{\bar{G}_t}{\bar{L}_t}. \quad (17)$$

Here b_t is the empirical win–loss ratio based on realized trades.

To incorporate the model's prediction for the current trade, we also keep track of predicted gains. Let

$$\hat{g}_t = \text{model-predicted gain for trade } t, \quad (18)$$

$$\bar{g}_t = \frac{1}{t} \sum_{s=1}^t \hat{g}_s \quad (19)$$

be the running average of all predicted gains up to and including trade t (even for trades that ultimately lose). We then define a prediction-based weight

$$h_t = \frac{\hat{g}_t}{\bar{g}_t}, \quad (20)$$

The Kelly fraction for trade t is then

$$f_t^{\text{Kelly}} = \hat{p}_t - \frac{\hat{q}_t}{h_t \cdot b_t}. \quad (21)$$

In practice, we separate the procedure into a short “warm-up” phase and a Kelly-driven phase. For the first $N = 120$ trades, we do not trust the early estimates of \hat{p}_t , \bar{G}_t , and \bar{L}_t (because there are too few wins and losses), so we use a fixed, small fraction,

$$f_t = 0.5\% = 0.005, \quad t \leq N, \quad (22)$$

solely to accumulate enough data to stabilize numWins , numLosses , totalGains , and totalLosses .

After this warm-up period ($t > N$), we switch to the model-weighted Kelly fraction (21) to size positions:

$$f_t = \begin{cases} f_t^{\text{Kelly}}, & \text{if } f_t^{\text{Kelly}} > 0, \\ 0, & \text{if } f_t^{\text{Kelly}} \leq 0. \end{cases} \quad (23)$$

The Kelly formula can produce negative values when either the probability of losing is high, or the win–loss ratio is unfavorable. In those cases, we set the Kelly fraction to 0 and trade nothing.

6.1. Active vs. Passive Kelly

We call a strategy that changes \hat{p}_t , \hat{q}_t , \bar{G}_t , and \bar{L}_t only when equation model actually trades *Active Kelly*. That is, only when Equation 23 is in the case $f_t^{\text{Kelly}} > 0$, do any of \hat{p}_t , \hat{q}_t , \bar{G}_t , or \bar{L}_t .

However, even when a method under Kelly trades nothing, it could still keep track of whether its predictions were correct or not. That would influence \hat{p}_t and \hat{q}_t . Thus, if a prediction were correct, that could increase \hat{p}_t even though no trade happens. However, for the purpose of the calculations of \bar{G}_t and \bar{L}_t , numWins_t and numLosses_t are affected

only by situations in which an actual trade based on the prediction takes place. We call a strategy that changes \hat{p}_t and \hat{q}_t , though not \bar{G}_t , and \bar{L}_t even when Kelly suggests a zero bet *Passive Kelly*.

Formally, we keep track of another two variables $numWinsAll_t$ and $numLossesAll_t$ that are updated every time a prediction is made, regardless of whether $f_t^{Kelly} > 0$. In contrast, $numWins_t$ and $numLosses_t$ are updated only when $f_t^{Kelly} > 0$. The update rules then proceed as follows:

- $numWins, numLosses$: counts of winning and losing trades,
- $numWinsAll, numLossesAll$: counts of winning and losing predictions

$$\hat{p}_t = \frac{numWinsAll_t}{numWinsAll_t + numLossesAll_t}, \quad (24)$$

$$\hat{q}_t = 1 - \hat{p}_t, \quad (25)$$

$$\bar{G}_t = \frac{totalGains_t}{numWins_t}, \quad (26)$$

$$\bar{L}_t = \frac{totalLosses_t}{numLosses_t}, \quad (27)$$

$$b_t = \frac{\bar{G}_t}{\bar{L}_t}. \quad (28)$$

That is, only the calculation for \hat{p}_t changes, which causes the calculation for \hat{q}_t to change. The calculations for \bar{G}_t and \bar{L}_t remain the same because we don't want to dilute the gain calculations with zero-position-sized trades.

7. News Article-time Trading: Position Size and Hold Period

Our LLM models for news analysis predict only the direction of a currency exchange rate change based on the content of the article. These predictions are converted into a fixed-sized and fixed-hold period trading strategy. We use a fixed-sized bet rather than a variable Kelly-based method because news article analysis predicts only a direction, not a predicted gain, so using the Kelly criterion is not appropriate. For time series predictions, our models predict exchange rate prices. These price predictions are converted into a variable-sized trading and a variable-hold strategy.

7.1. Fixed Position Size and Hold for News Analysis

In the fixed position scheme, each trade (long or short) is executed with a constant position size of

$$q_{fixed} = 10,000 \text{ USD}. \quad (29)$$

There are three other policy decisions with respect to fixed-sized and fixed-hold trading on news articles.

1. If multiple events occur during the same time period, a single trade is executed on the basis of the majority vote. If there is no majority, no trade is executed.
2. If an event occurs while an existing trade is executing, a new independent trade is executed. Thus, multiple trades may be executed simultaneously and may be in different directions.
3. At the end of the trading day (5 pm), all trades are closed, regardless of the remaining hold time. In other words, no trades are held beyond 5 pm.

8. Evaluation Metrics for the Experiments Based Purely on the Exchange Rate Time Series

For time series trading, to obtain stable and comparable estimates across strategies and currency pairs, each trading strategy was executed $J = 10$ times, where each run ($j = 1, \dots, J$) represents a complete trading simulation. This repetition helps account for

stochastic variation due to model initialization and sampling effects in the models and configurations selected from Section 10, ensuring that the reported results reflect consistent, expectation-based performance rather than outcomes from a single random realization. We then aggregate the resulting outcomes.

For News article trading, we execute the trading simulation only once, since all parameters are fixed.

We report three evaluation metrics: the *average cumulative profit* (ACP), the *average profit per trade* (APPT), and the *monthly Sharpe ratio* (MSR) computed from per-trade returns.

Let $\pi_{t,s}^{(j)}$ denote the profit of strategy s at time t in run $j \in \{1, \dots, J\}$. Because of randomness, some runs may bet at time t and others may not. So, let the total number of trades over all runs (between 1 and J) and all times (between 1 and T) be $TotTrade(J, T)$. The average profit per trade (APPT) for strategy s for all runs and all times up to T is then:

$$APPT_s = (\sum_{t=1}^T \sum_{j=1}^J \pi_{t,s}^{(j)}) / TotTrade(J, T). \quad (30)$$

The average cumulative profit (ACP_s) for method s is simply the total profit divided by J :

$$ACP_s = (\sum_{t=1}^T \sum_{j=1}^J \pi_{t,s}^{(j)}) / J. \quad (31)$$

For the Monthly Sharpe ratio, profits are first aggregated at the monthly level. Let \mathcal{T}_m denote the set of time indices belonging to month m , and define the monthly profit for run j as

$$P_{m,s}^{(j)} = \sum_{t \in \mathcal{T}_m} \pi_{t,s}^{(j)}. \quad (32)$$

We then compute the average monthly profit across runs:

$$\bar{P}_{m,s} = \frac{1}{J} \sum_{j=1}^J P_{m,s}^{(j)}. \quad (33)$$

Let μ_s and σ_s denote the mean and standard deviation of $\bar{P}_{m,s}$ across all months. The Monthly Sharpe ratio is defined as

$$MSR_s = \frac{\mu_s}{\sigma_s}. \quad (34)$$

9. Validation Measurements of Trading Strategies

9.1. Statistical Significance of Average Cumulative Profits

To evaluate whether the observed profits of each trading strategy reflect predictive power rather than random chance, we performed a non-parametric significance analysis. For a given strategy M and currency C , we first collected the trade directions produced by M on the test dataset across all runs, including *buy*, *sell*, and *no trade*. From these directions, we estimated the empirical probabilities p_{buy} , p_{sell} , and p_{nt} (with $p_{\text{buy}} + p_{\text{sell}} + p_{\text{nt}} = 1$). We then constructed a randomized trader that, at each time point t , sampled an action from this categorical distribution: it bought currency C with probability p_{buy} , sold with probability p_{sell} , and placed no trade with probability p_{nt} . This design preserves both the overall directional bias and the trading frequency of M on the test set, while removing any dependence on future price movements. For Kelly-based strategies, the bet size at time t was computed using the same execution rules as in M , based on the randomized trader's performance history up to time t .

We simulated $N = 10,000$ such randomized trading trajectories for each strategy to construct a null distribution of profits, denoted as $\{P_{\text{null}}^{(i)}\}_{i=1}^N$. Each trajectory was evaluated

using the same execution framework as used in the experiments described in Section 13. The observed final average cumulative profit of method M is denoted by P_M^{obs} .

The Monte Carlo p -value is then computed as

$$pValue(M) = \frac{1 + \sum_{i=1}^N \mathbf{1}[P_{\text{null}}^{(i)} \geq P_M^{\text{obs}}]}{N + 1}, \quad (35)$$

where $\mathbf{1}[\cdot]$ is the indicator function. Intuitively, this p -value calculation counts the number of times the random trading process yields a higher profit than method M 's profit and divides that by N , the number of random trajectories. (The $+1$ correction is to avoid p -values of 0.) A smaller $pVal(M)$ indicates stronger evidence that the observed profit by method M is statistically significantly better than a random strategy.

To further assess the robustness of M 's performance, we computed a 95% confidence interval for its mean profit using nonparametric bootstrap resampling. Specifically, we resampled with replacement from the individual trade profits achieved by M to generate bootstrap replicates of the mean profit. This procedure provides an empirical estimate of uncertainty around the observed performance, following standard resampling procedures in statistical inference [21,22].

10. Model Tuning Procedures and Results for Forecasting Models

We adopted the same training process for all forecasting models described in Section 2.2.1. Each model (N-BEATS, N-HiTS, and TCN) was trained using the Adam optimizer [23] with a learning rate of 10^{-4} , a batch size of 1024, and for 50 epochs on a single NVIDIA A100 GPU. The mean squared error (MSE) loss function was used across all models.

All forecasting models, including the zero-shot foundation models, share two key hyperparameters: the prediction length and the size of the input context window (C). Since our objective is to predict the next mid-price, the prediction length was fixed to one step. To identify the optimal input context size, we evaluated several values, $C \in 16, 32, 64, 128, 256$.

After training the models described in Section 2.2.1 using various input context lengths, we evaluated their performance on the validation dataset (July 2024) alongside the zero-shot models, Chronos-Bolt and Toto.

For each currency pair, we ran a complete trading simulation using the model-driven and ensemble strategies described in Section 5.2 and 5.3, respectively. The goal was to identify, for each model, the input context length that produced the highest average total profit on the validation data. This context length was then selected as the optimal configuration for that model and used in the final trading experiments on the test dataset.

To ensure a fair comparison across models and account for randomness in model initialization (for trained models) and sampling (for zero-shot models), each trading simulation was repeated ten times for every model-context pair. The resulting total profits were averaged to obtain stable and representative performance estimates.

The resulting best parameter values on the validation dataset are shown for USD/CNY under active Kelly in Table 1 and under passive Kelly in Table 2, and for USD/BRL under active Kelly in Table 3 and under passive Kelly in Table 4. For our experimental results, each forecasting model is paired with its optimal input chunk length as selected on the validation split. It is important to note, however, that these validation-optimal configurations do not necessarily correspond to the models that ultimately perform best on the out-of-sample test period. As shown later in the trading-performance tables (Tables 6, 7, 9, and 10), models such as Toto, Ensemble, and Chronos-Bolt compete closely on the validation data, whereas ARIMA emerges as the strongest performer over the full one-year test horizon, though its Sharpe Ratio remains low.

Model	Input Chunk Length				
	16	32	64	128	256
ARIMA	N/A	1542.38	734.46	1047.33	1468.21
N-BEATS	-279.77	-181.15	-109.63	-129.55	-395.10
N-HITS	-7.08	-6.95	-6.84	-6.61	-6.62
TCN	-6.92	-171.7	105.27	-44.46	-154.21
Toto	-374.01	-972.29	-333.57	282.25	1337.59
Chronos-Bolt	1610.27	-78.81	-6.12	667.73	4862.15
Ensemble	N/A	326.02	223.22	808.15	3312.3

Table 1. Active Kelly (see Section 6.1 for definition) average cumulative profit (see Section 8) results on validation data (July 2024) for US dollar to Chinese Yuan as a function of input chunk length (how far back the model looks). Note that the average is over different random seeds. For ARIMA, the chunk length means the number of data points to fit the ARIMA(1,1,1) model. The **Chronos-Bolt** model with an input chunk length of 256 achieved the highest total average profit.

Model	Input Chunk Length				
	16	32	64	128	256
ARIMA	N/A	2584.48	845.55	1244.75	2027.81
N-BEATS	-233.60	-155.91	-254.34	-164.64	-280.95
N-HITS	4.47	34.94	28.24	19.51	14.96
TCN	-32.66	-67.69	61.55	-0.36	-91.55
Toto	-351.29	-1670.79	-467.70	220.64	1442.90
Chronos-Bolt	2307.03	125.86	-102.08	683.88	4252.93
Ensemble	N/A	361.52	116.26	795.55	14.96

Table 2. Passive Kelly (see Section 6.1 for definition) average cumulative profit (see Section 8) results on validation data (July 2024) for US dollar to Chinese Yuan as a function of input chunk length. Note that the average is over different random seeds. For ARIMA, the chunk length means the number of data points to fit the ARIMA(1,1,1) model. The **Chronos-Bolt** model with an input chunk length of 256 achieved the highest total average profit.

Model	Input Chunk Length				
	16	32	64	128	256
ARIMA	N/A	-1446.05	11484.59	13182.42	13881.76
N-BEATS	4229.07	2752.76	3178.23	2888.63	4289.54
N-HITS	211.06	-91.79	-130.59	-105.4	-124.67
TCN	-211.13	1141.68	270.76	-640.03	-121.4
Toto	7210.09	12510.99	11549.02	4273.77	-1949.14
Chronos-Bolt	-4164.37	10256.85	14146.14	8594.79	7424.0
Ensemble	N/A	1593.17	23375.10	21033.06	8110.11

Table 3. Active Kelly (see Section 6.1 for definition) average cumulative profit (see Section 8) results on validation data (July 2024) for US dollar to Brazilian Real as a function of input chunk length. Note that the average is over different random seeds. For ARIMA, the chunk length means the number of data points to fit the ARIMA(1,1,1) model. The **Ensemble** model with an input chunk length of 64 achieved the highest total average profit.

Model	Input Chunk Length				
	16	32	64	128	256
ARIMA	N/A	623.45	9504.99	20101.04	16270.43
N-BEATS	4686.94	3693.28	3721.51	4389.44	4190.89
N-HiTS	-309.91	-553.92	-590.15	114.14	865.49
TCN	-731.37	-179.77	-161.00	-723.53	-683.75
Toto	6899.56	11502.12	8554.53	4559.66	-835.94
Chronos-Bolt	-4106.82	8682.09	15463.43	7696.98	7566.64
Ensemble	N/A	1290.38	26062.88	25571.07	14067.19

Table 4. Passive Kelly (see Section 6.1 for definition) average cumulative profit (see Section 8) results on validation data (July 2024) for US dollar to Brazilian Real as a function of input chunk length. Note that the average is over different random seeds. For ARIMA, the chunk length means the number of data points to fit the ARIMA(1,1,1) model. The **Ensemble** model with an input chunk length of 64 achieved the highest total average profit.

11. Choice of LLM Architectures and Prompts for News Analysis Based on Training/Validation Data

Choosing a particular LLM architecture (model, parameters, etc.) is still an art, since different LLM architectures do far more than compute the next probable token.

11.1. Model Parameter Setting for USD/BRL

After experimenting with various models on the training data (the dataset was divided as follows: 3524 articles from November 21, 2024, through December 31, 2024), we chose Gemini 2.5 Flash to generate predictions for USD/BRL foreign exchange. We adopted a sampling temperature of 0.1 to minimize output variance and achieve deterministic outputs. We experimented with 8-bit post-training quantization to reduce the memory footprint and the inference latency for predictions. However, experiments yielded inconsistent classification behavior and reduced accuracy, leading us to choose full-precision inference for signal reliability.

11.2. Model and Parameter Setting for USD/CNY

We chose DeepSeek-V3 Chat as our default LLM for USD/CNY label generation. This decision was driven by API cost efficiency and strong Chinese-text capability, which we expected to better capture China-related news signals relevant to CNY. Accordingly, we used DeepSeek-V3 Chat (Dec 2024) with default parameters as the baseline model in our USD/CNY prediction.

In our experiments, we used the default values for this model. According to the documentation, frequency_penalty defaults to 0, presence_penalty defaults to 0, temperature defaults to 1, and top_p defaults to 1.

11.3. Additional Considerations

We initially explored both translation-based and direct processing approaches. We began by translating Portuguese financial news to English and verifying classification performance, but observed superior results when providing Portuguese content directly to the LLMs. Consequently, we adopted direct Portuguese and Chinese processing to preserve semantic nuances and domain-specific terminology prevalent in economic news, as this approach yielded better sentiment classification accuracy compared to translated news articles.

To allow comparisons with the literature, we include results for FinBERT (trained on English-language financial texts [24]), Chinese FinBERT2 [25], and FinBERT-PT-BR for Brazilian Portuguese [26].

11.4. Selecting the Optimal Prompt and Fixed-Size Trade Hold Period

Exchange	Model	Hold Period (Minutes)							
		1	2	5	10	20	40	80	120
USD/CNY	Expert	-5.53	-13.28	-42.14	-36.78	-79.25	-117.61	-145.80	-192.96
	Naive	-17.23	-35.79	-19.75	21.01	38.56	38.50	167.32	218.18
	Naive+	-12.17	-19.92	-21.23	-30.29	-50.08	-33.16	2.38	-10.82
	FinBERT2	-13.49	-58.97	-47.62	-58.64	6.77	-14.39	65.93	77.78
USD/BRL	P0	23.31	-9.69	-59.19	-175.66	-259.24	-673.85	-1051.15	-719.30
	P1	56.34	61.51	-26.14	-86.73	-88.58	-474.47	-871.27	-634.21
	P2	53.02	22.70	-22.21	-107.01	-119.06	-503.92	-697.18	-681.36
	FinBERT2	-33.44	-66.48	-107.91	-170.88	94.54	33.21	267.09	830.09

Table 5. Cumulative profit over the training data (January 1, 2024 to June 30, 2025) for various LLM prompt strategies for different hold periods as measured by total profit on the training data, using the fixed position sizing. Note that the data size is larger here than for the time series analysis of Section 10 because no training is necessary for news article analysis based on prompts. The **Naive** prompt for USD/CNY (hold period: 120 minutes) and **FinBERT2** for USD/BRL (hold period: 120 minutes) achieved the highest total profits.

As part of the hyperparameter tuning, we performed a tuning procedure with different fixed hold periods to determine the best hold period for the fixed-size position trading strategy (Table 5). Note that LLM trading always uses a fixed position size trading strategy because the LLM models do not generate a probability of success, a required element to compute the Kelly position size.

Comparing LLM tuning with forecast model tuning, we note that the magnitude of profit (loss) of LLM trading is small compared to forecasting models. Our explanation for this difference is the frequency of trading. News events occur multiple times per day, but forecasting models may trade every minute, depending on trading patterns.

When comparing the various prompts for trading, profit (loss) generally increases (decreases) with the hold period. Our explanation for this trend is that increasing the hold time allows for more time for the currency to move in some direction.

Considering the USD/CNY prompts of Naive and Naive+, the hypothesis of including more details in a simple prompt does not appear to be confirmed, as the more specific Naive+ prompt performs worse than the Naive prompt. Following the same logic, and contrary to our expectations, the longer and more detailed Expert prompt performs the worst of all the prompts across most of the hold periods. The performance of FinBERT2 is surprisingly uneven, considering that the model was trained on Chinese financial text. Finally, the best combination of prompt and hold period for the USD/CNY is the Naive prompt with a hold period of 120 minutes. We select this prompt and hold period for the remaining experiments with USD/CNY.

Considering the USD/BRL prompts, we note that the addition of examples to a detailed expert prompt has a mixed impact during short hold periods. But all expert prompts perform poorly over longer hold periods. Finally, FinBERT2 strongly outperforms all variations of expert prompts over a longer hold period. For the remaining USD/BRL experiments, we use the FinBERT2 model with a 120 minute hold period for USD/BRL trading.

12. Experimental Methodology for Test Data

Model selection and input context length choices for all forecasting models follow the procedure described in Section 10. After identifying the optimal context length for each model based on performance on the validation dataset, we conducted out-of-sample trading simulations on the test dataset for both currency pairs (USD/CNY and USD/BRL). All models were trained and validated using data from **January 1, 2024 to July 31, 2024**,

while all reported trading results are evaluated on the held-out test period spanning **August 1, 2024 to July 31, 2025**.

Trading starts with \$1,000,000 in holdings. A fixed Kelly fraction (see Eq 22) is used initially, before sufficient trading has occurred.

13. Comparison of Trading Strategies Results for USD/CNY

13.1. Active and Passive Kelly Position Size Trading Results

Figures 1 and 2 show the ACP trajectories for the USD/CNY test dataset using active and passive Kelly respectively. These graphs include all trading strategies described in Section 5 and the forecasting models from Section 2.2, using their best-performing configurations identified in Section 10. All simulations were conducted with active and passive Kelly-based position sizing, described in Section 6.1, and averaged over ten independent runs to ensure result stability. News event models are incompatible with the Kelly criterion because they do not produce a probability of success for the prediction based on the news article. Fixed position size trading for news event models is discussed in the next section.

Table 6 and 7 summarize, for each trading strategy, the final average cumulative profit (ACP_T), the average profit per trade (APPT) and the risk-adjusted performance measured by the monthly Sharpe ratio (MSR) for the USD/CNY test dataset for active and passive Kelly respectively.

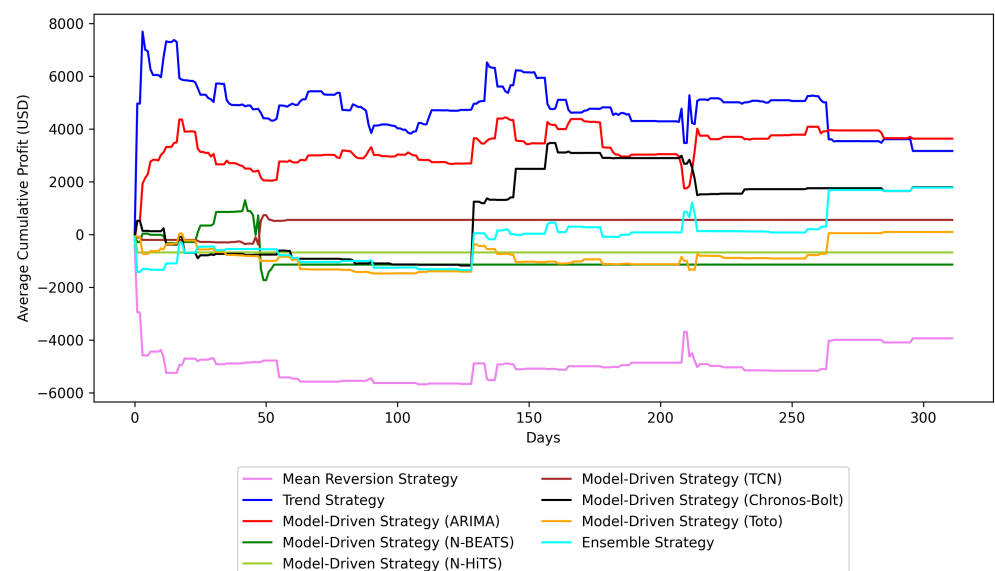


Figure 1. Average cumulative profit (ACP) trajectories for all trading strategies using active Kelly (Section 6.1) bet sizing. Results are shown for US dollar vs. Chinese Yuan. Note that the accumulated profit is less than 1% over the whole year. Please see the analysis in the text.

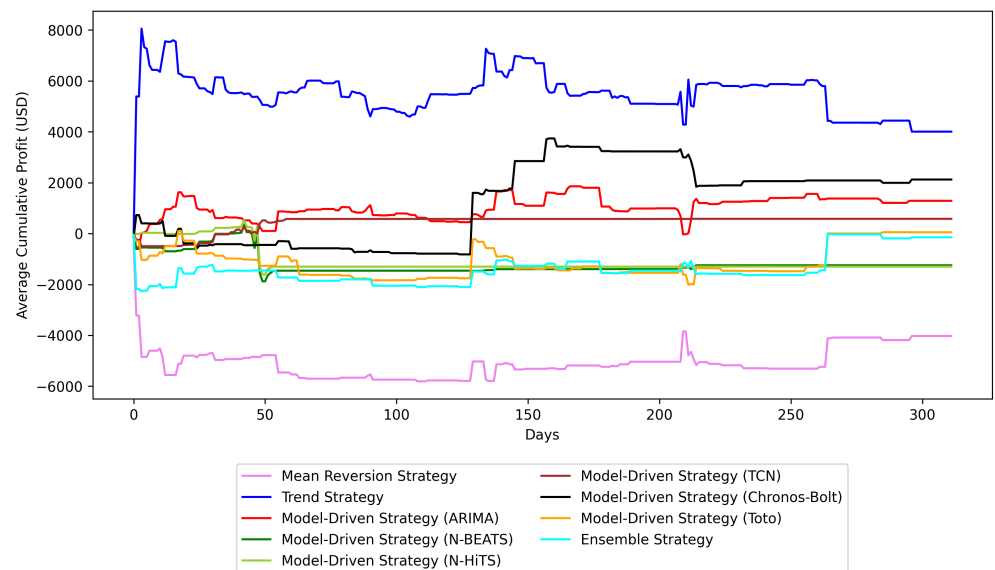


Figure 2. Average cumulative profit (ACP) trajectories for all trading strategies using passive Kelly (Section 6.1) bet sizing. Results are shown for US dollar vs. Chinese Yuan. As above, note that the accumulated profit is less than 1% over the whole year. Please see analysis in text.

Trading Strategy	ACP (USD)	APPT (USD/trade)	MSR	p-Value	95% CI
Mean Reversion	-3935.20	-8.114	-0.232	1.0000	[-9767.47 - 1978.9]
Trend	3168.96	1.252	0.148	0.0157	[-8266.30 - 14792.42]
Model-driven (ARIMA)	3636.40	3.190	0.305	0.0014	[-6039.73 - 13297.87]
Model-driven (N-BEATS)	-1138.79	-0.245	-0.184	0.9203	[-6078.73 - 3822.78]
Model-driven (N-HITS)	-677.78	-5.510	-0.302	1.0000	[-2248.56 - 68.35]
Model-driven (TCN)	555.50	0.236	0.181	0.2369	[-3272.52 - 4325.84]
Model-driven (Toto)	97.87	0.203	0.016	0.3005	[-5628.58 - 5896.36]
Model-driven (Chronos-Bolt)	1794.55	1.850	0.167	0.0213	[-4366.70 - 7966.85]
Ensemble	1779.10	2.417	0.251	0.0487	[-4819.04 - 8464.80]

Table 6. Active Kelly (Section 6.1) trading performance of all trading strategies on the US dollar vs. Chinese Yuan test dataset (August 1, 2024 to July 31, 2025). For time series-based approaches, **ARIMA** achieves the lowest p-value, the highest average cumulative profit, average profit per trade, and monthly Sharpe Ratio, but its monthly Sharpe Ratio is too low to be acceptable.

Trading Strategy	ACP (USD)	APPT (USD/trade)	MSR	p-Value	95% CI
Mean Reversion	-4026.04	-7.879	-0.235	1.0000	[-10342.47 - 2173.03]
Trend	4007.35	1.612	0.181	0.0004	[-7055.61 - 15197.29]
Model-driven (ARIMA)	1288.54	1.967	0.237	0.0436	[-6331.85 - 8599.05]
Model-driven (N-BEATS)	-1246.67	-0.323	-0.270	0.9431	[-5438.41 - 2946.7]
Model-driven (N-HiTS)	-1299.83	-0.332	-0.262	0.9850	[-4402.13 - 1870.96]
Model-driven (TCN)	579.38	0.101	0.188	0.2740	[-5841.05 - 7082.00]
Model-driven (Toto)	50.76	0.086	0.006	0.3526	[-6482.33 - 6724.88]
Model-driven (Chronos-Bolt)	2124.77	2.295	0.204	0.0108	[-3868.09 - 8123.84]
Ensemble	-147.20	-0.227	-0.020	0.9637	[-6331.09 - 6033.38]

Table 7. Passive Kelly (Section 6.1) trading performance of all trading strategies on the US dollar vs. Chinese Yuan test dataset (August 1, 2024 to July 31, 2025). For time series-based approaches, **Trend** achieves the highest average cumulative profit and the lowest p-value. **Chronos-Bolt** achieves the highest average profit per trade. **ARIMA** achieves the highest monthly Sharpe ratio, but it's still too low to be acceptable.

13.2. Discussion of Results for Kelly Position Size for USD/CNY

The graphs and tables (Figure 1 and Table 6) for the Active Kelly criterion (Section 6.1) approach show that: (i) the profit and/or loss of any trading method is minimal (less than 1% of the initial \$1 million), (ii) all methods have a high p-value, and (iii) their confidence interval straddles zero.

The positive and negative jumps at the beginning reflect that the mid-price exchange rate shows unusually high volatility early in the test data set (Figure 3). To warm up the model, we use a small fixed Kelly fraction (see Eq 22) for the initial 120 minutes of trading. This rule interacts with the high volatility of exchange prices at the beginning of the test data trading, producing a spike.

Although time series forecasting for USD/CNY using the Passive Kelly (Section 6.1) criterion (Figure 2 and Table 7) have different results, with the Trend trading strategy producing the highest average cumulative profit, the general results of Active and Passive Kelly have the same general structure and the same difficulty with confidence intervals crossing into overall losses. In addition, the total profit for the best strategies remains small compared to the initial holding size.

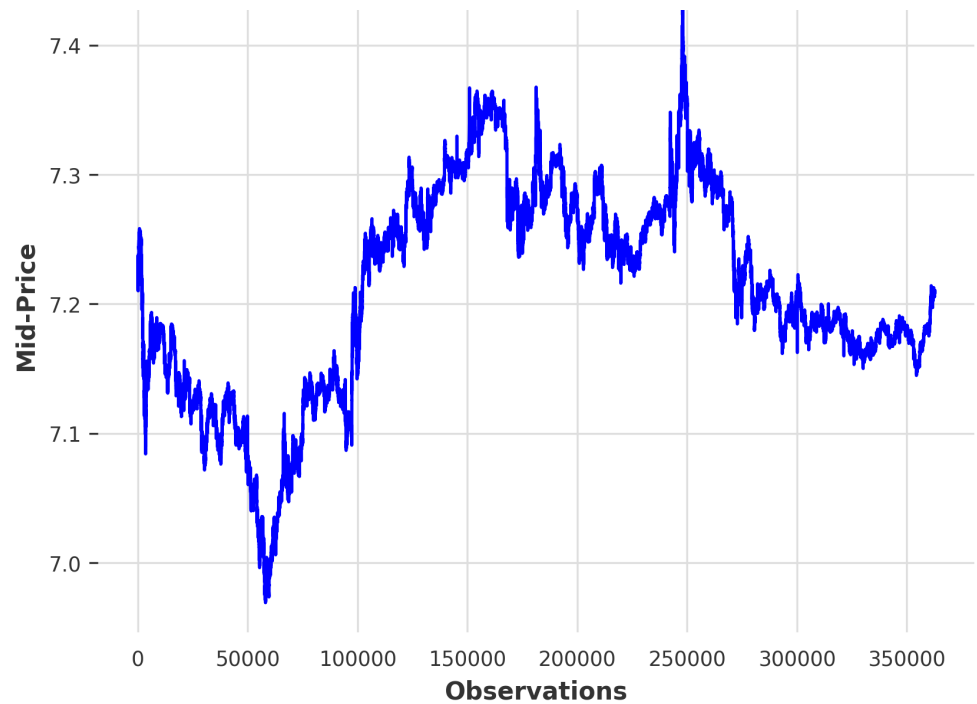


Figure 3. The actual mid-price value for USD/CNY for the test data.

13.3. Fixed Position Size Trading for USD/CNY

The results for the fixed position size trading include both pure time series approaches and news-based approaches.

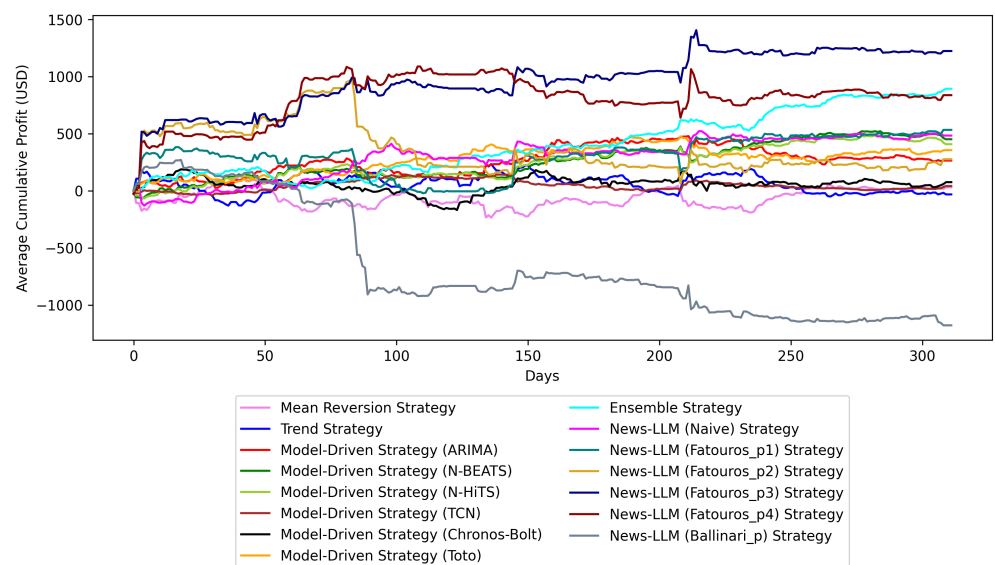


Figure 4. Average Cumulative profit (ACP) trajectories for all trading strategies using fixed bet sizing. Results are shown for US dollar vs. Chinese Yuan.

Trading Strategy	ACP (USD)	APPT (USD/trade)	MSR	p-Value	95% CI
Mean Reversion	31.27	0.000	0.028	0.4582	[-529.88 - 582.17]
Trend	-31.27	-0.000	-0.028	0.5451	[-573.80 - 533.05]
Model-driven (ARIMA)	264.94	0.002	0.257	0.1807	[-294.65 - 826.59]
Model-driven (N-BEATS)	451.52	0.004	0.744	0.0264	[-111.01 - 987.27]
Model-driven (N-HiTS)	408.00	0.003	0.534	0.0324	[-159.93 - 967.52]
Model-driven (TCN)	42.10	0.000	0.089	0.3928	[-510.61 - 591.63]
Model-driven (Toto)	356.05	0.003	0.407	0.1138	[-187.63 - 914.64]
Model-driven (Chronos-Bolt)	75.92	0.001	0.077	0.4047	[-476.84 - 639.52]
Ensemble	893.65	0.007	1.100	0.0006	[341.74 - 1448.32]
News-LLM (Naive)	483.65	0.493	0.572	0.0451	[-84.52 - 1053.63]
News-LLM (Fatouros_p1)	533.51	0.703	0.273	0.1021	[-226.85 - 1301.77]
News-LLM (Fatouros_p2)	276.60	0.241	0.092	0.3588	[-710.89 - 1268.83]
News-LLM (Fatouros_p3)	1223.79	1.313	0.559	0.0056	[373.01 - 2094.00]
News-LLM (Fatouros_p4)	837.44	0.933	0.376	0.1122	[36.49 - 1680.81]
News-LLM (Ballinari_p)	-1177.20	-1.507	-0.414	0.9966	[-1951.79 - -401.07]

Table 8. Trading performance of Fixed position size trading performance of all trading strategies on the USD/CNY test dataset. The news-event strategy using **Fatouros_p3** model achieved the highest total average profit and average profit per trade. The **Ensemble** approach has a low p-value and wholly positive confidence interval. The modest Sharpe Ratio, however, indicates a lot of variance. None of the news-based systems has an acceptable Sharpe ratio.

13.4. Discussion of Results for Fixed Position Size for USD/CNY

The results for the models for the USD/CNY exchange rate using fixed position sizing (Figure 4 and Table 8) again show minimal profits and losses compared to the \$1 million of initial capital, mostly under 0.1%. That said, among the time series forecasting models, the Ensemble model has the highest average cumulative profit, highest average profit per trade, and the highest monthly Sharpe ratio and low p-value. The confidence interval range shows a small but consistent profit. The Ensemble model appears to learn a profitable signal over time. Note that the better-performing models also steadily increase average cumulative profitability over time.

The results for the news event models for the USD/CNY exchange rate using fixed position sizing (Figure 4 and Table 8) show that, among the news event models, the Fatouros_p3 model has the highest average cumulative profit, the highest average profit per trade, and the second highest monthly Sharpe ratio. (The Naive model has the highest Sharpe ratio.) The p-val of Naive is high, but for Fatouros_p3, the p-value is well under 0.05, indicating a possible signal. Also, the confidence interval of Fatouros_p3 is fully in the profitable range.

Comparing the texts of the different prompts, we note that slight variations in the wording result in vastly different performances. Generally, the shorter the prompt, the better the performance. The Fatouros prompts all use only the headline, and the trading strategy opens a position when the news article arrives and closes the position at the end of the day. The Ballinari prompt is a long (expert) description of economic factors that uses the same strategy of closing at the end of the day, but performs poorly.

13.5. Overall Conclusion of the Chinese Yuan vs. US Dollar Exchange Rate Experiments

Although some methods did find a signal in the data, the overall performance of all the methods tested indicates that the signal is weak. We hypothesize that likely governmental manipulation of the Yuan-dollar exchange rate [3] defeats prediction models.

14. Comparison of Trading Results for USD/BRL

14.1. Active and Passive Kelly Position Size Trading Results

Figures 5 and 6 show the ACP trajectory for the USD/BRL test dataset using active and passive Kelly respectively. These plots include all trading strategies described in Section 5 and the forecasting models from Section 2.2, using their best-performing configurations identified in Section 10. All simulations were conducted with active and passive Kelly-based position sizing, described in Section 6.1, and averaged over ten independent runs to ensure result stability.

Table 9 and 10 summarize, for each trading strategy, the final average cumulative profit (ACP_T), the average profit per trade (APPT), and the risk-adjusted performance measured by the monthly Sharpe ratio (MSR) for the USD/BRL test datasets for active and passive Kelly respectively.

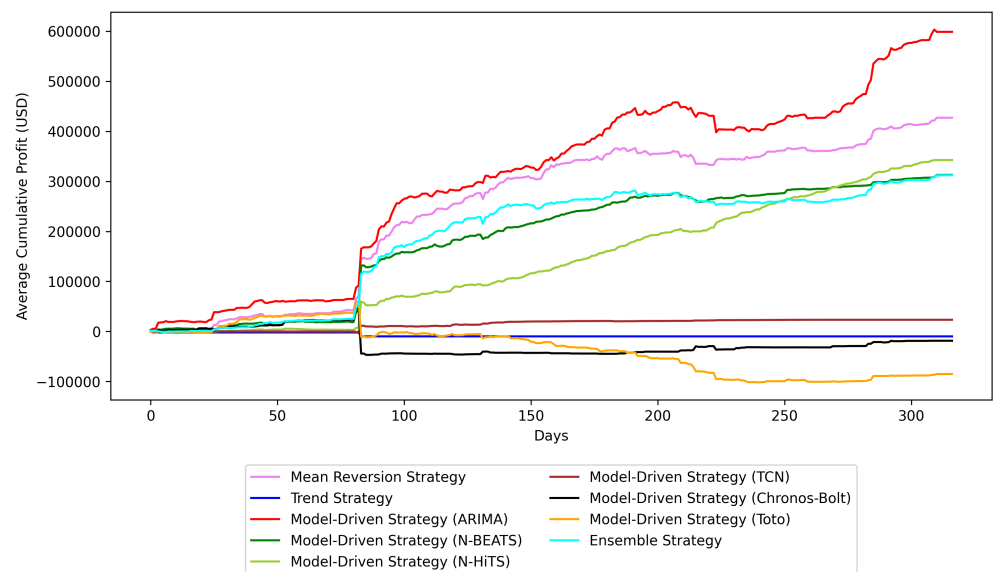


Figure 5. Average cumulative profit (ACP) trajectories for all trading strategies using Active Kelly (Section 6.1) bet sizing. Results are shown for US dollar vs. Brazilian Real. Please see the analysis in the text.

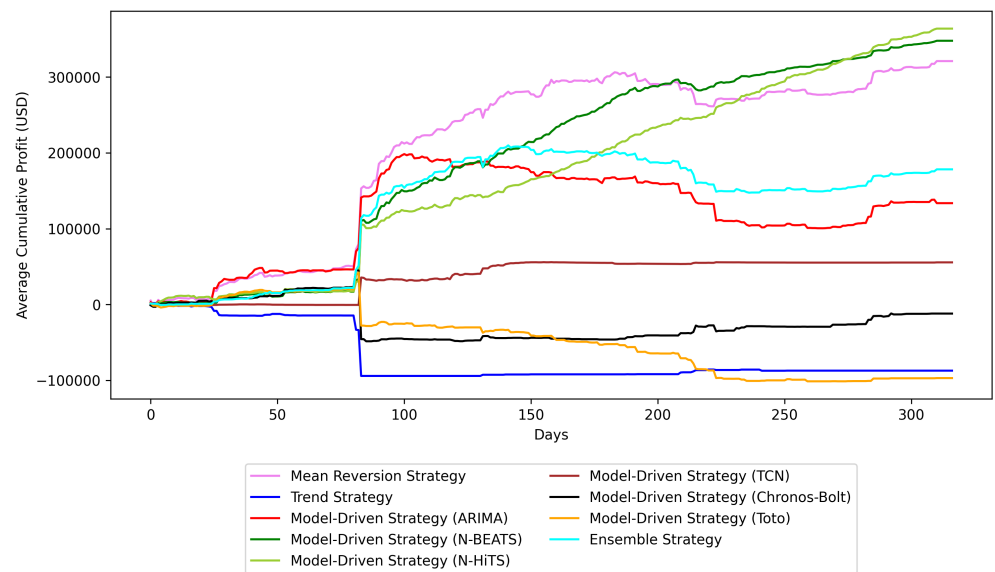


Figure 6. Average cumulative profit (ACP) trajectories for all trading strategies using Passive Kelly (Section 6.1) bet sizing. Results are shown for US dollar vs. Brazilian Real. Please see analysis in text.

Trading Strategy	ACP (USD)	APPT (USD/trade)	MSR	p-Value	95% CI
Mean Reversion	426939.78	41.551	0.654	0.0001	[250771.24 - 635922.18]
Trend	-9818.05	-76.704	-0.330	0.9853	[-27478.00 - -6.30]
Model-driven (ARIMA)	598519.53	67.721	0.711	0.0001	[410060.16 - 804488.2]
Model-driven (N-BEATS)	312758.05	14.738	0.620	0.0001	[151165.36 - 506172.83]
Model-driven (N-HITS)	342419.59	5.213	1.130	0.0001	[236999.77 - 470077.01]
Model-driven (TCN)	23349.10	3.137	0.533	0.0011	[-12362.68 - 92608.39]
Model-driven (Toto)	-85016.10	-17.322	-0.322	1.0000	[-260799.95 - 46495.75]
Model-driven (Chronos-Bolt)	-18585.75	-17.177	-0.068	0.9864	[-199396.80 - 103277.19]
Ensemble	311932.03	48.901	0.550	0.0001	[139531.70 - 513951.80]

Table 9. Active Kelly (Section 6.1) trading performance of all trading strategies on the US dollar vs. Brazilian Real test dataset (August 1, 2024 to July 31, 2025). For time series based approaches, **ARIMA** achieves the highest average cumulative profit and average profit per trade. The monthly Sharpe Ratio and p-value indicates that **N-HITS** may be a viable trading strategy

Trading Strategy	ACP (USD)	APPT (USD/trade)	MSR	p-Value	95% CI
Mean Reversion	321006.24	41.468	0.526	0.0001	[164594.84 - 536500.05]
Trend	-87132.54	-176.025	-0.291	0.9995	[-236625.39 - -877.77]
Model-driven (ARIMA)	133704.20	44.958	0.214	0.0001	[-4614.92 - 309216.61]
Model-driven (N-BEATS)	347815.26	12.225	0.718	0.0001	[203941.03 - 522141.80]
Model-driven (N-HITS)	363831.70	7.517	0.953	0.0001	[236404.80 - 522046.69]
Model-driven (TCN)	55637.03	2.975	0.419	0.0384	[-47782.35 - 180431.77]
Model-driven (Toto)	-96940.65	-49.563	-0.440	1.0000	[-264723.77 - 19118.93]
Model-driven (Chronos-Bolt)	-11893.92	-9.813	-0.043	0.9787	[-199142.35 - 114017.81]
Ensemble	178288.17	47.866	0.326	0.0001	[27476.97 - 368867.17]

Table 10. Passive Kelly (Section 6.1) trading performance of all trading strategies on the US dollar vs. Brazilian Real test dataset (August 1, 2024 to July 31, 2025). For time series based approaches, the **Ensemble** strategy achieves the highest average profit per trade. **N-HITS** achieves the highest average cumulative profit and monthly Sharpe ratio, but it's still too low to be acceptable.

14.2. Discussion of Results for Kelly Position Size for USD/BRL

The results for the forecasting models for the USD/BRL exchange rate using the Active Kelly (Section 6.1) criterion for position sizing (Figure 5 and Table 10) show profits of approximately 60% for some methods (like ARIMA) with low p-values and consistently positive confidence intervals. While the results show that the ARIMA model has the highest average cumulative profit and highest average profit per trade. On the other hand, N-HITS has the highest monthly Sharpe ratio. The results for Passive Kelly show that the same methods doing well, but N-HITS dominates the other methods.

We note that the time series of average cumulative profit also shows several spikes (Figure 7). Nothing in foreign exchange is smoothly predictable. The mid-price exchange rate shows unusually high volatility at the spike, which some models can exploit. By and large though, the models work well in both the spike regions and the spike-free regions.

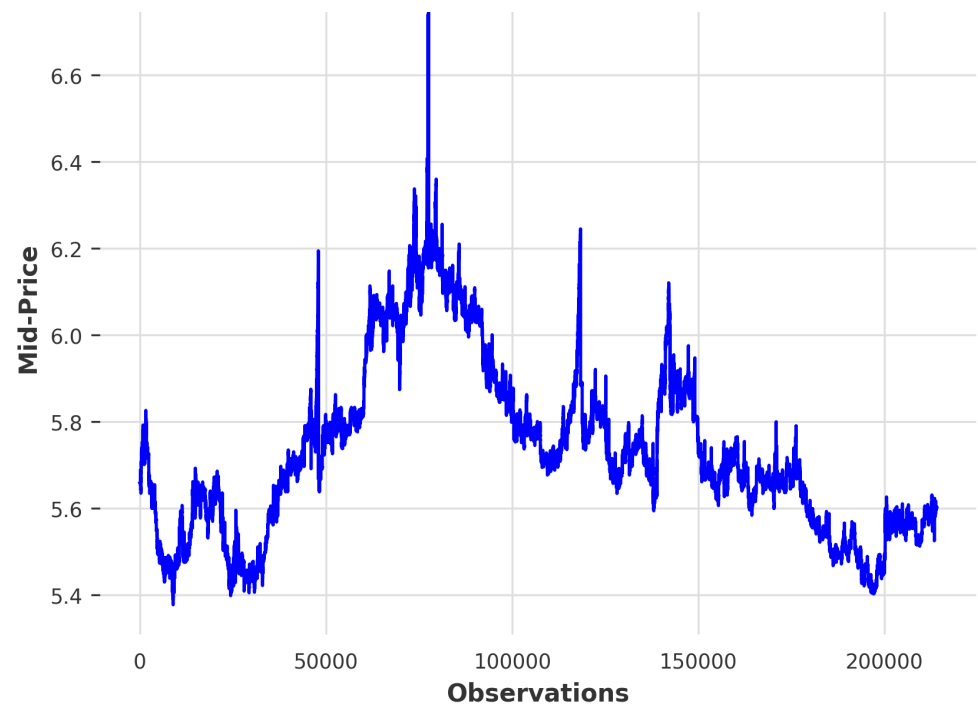


Figure 7. The actual mid-price value for USD/BRL for the test data.

14.3. Fixed Position Size Trading Results for USD/BRL

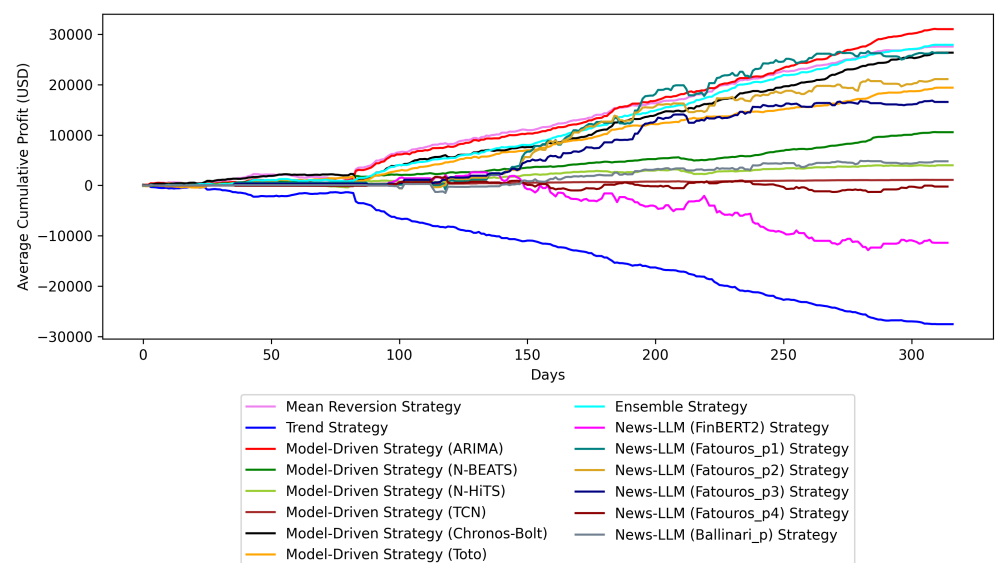


Figure 8. Average Cumulative profit (ACP) trajectories for all trading strategies using fixed bet sizing. Results are shown for US dollar vs. Brazilian Real.

Trading Strategy	ACP (USD)	APPT (USD/trade)	MSR	p-Value	95% CI
Mean Reversion	27535.23	0.301	1.185	0.0001	[23891.47 - 31461.00]
Trend	-27535.23	-0.301	-1.185	1.0000	[-31443.57 - -23887.67]
Model-driven (ARIMA)	30996.89	0.331	1.199	0.0001	[26710.08 - 34156.58]
Model-driven (N-BEATS)	10550.02	0.112	0.990	0.0001	[6666.64 - 14182.39]
Model-driven (N-HiTS)	3989.35	0.042	0.558	0.0024	[286.18 - 7830.53]
Model-driven (TCN)	1090.89	0.012	0.594	0.3711	[-2664.94 - 4906.57]
Model-driven (Toto)	19386.22	0.206	1.356	0.0001	[15157.92 - 22623.78]
Model-driven (Chronos-Bolt)	26331.8	0.280	1.360	0.0001	[21929.69 - 29233.43]
Ensemble	27889.36	0.296	1.261	0.0001	[24077.89 - 31670.24]
News-LLM (FinBERT2)	-11387.18	-1.742	-0.479	0.0021	[-17659.26 - -5150.56]
News-LLM (Fatouros_p1)	26262.53	5.489	0.830	0.0001	[18856.47 - 33564.22]
News-LLM (Fatouros_p2)	21082.86	4.865	0.818	0.0007	[14172.44 - 27993.36]
News-LLM (Fatouros_p3)	16552.93	3.783	0.819	0.0001	[9607.71 - 23435.50]
News-LLM (Fatouros_p4)	-232.31	-0.180	-0.025	0.0024	[-3754.81 - 3317.84]
News-LLM (Ballinari_p)	4780.90	3.482	0.654	0.4354	[795.20 - 8742.08]

Table 11. Fixed position size trading performance of all trading strategies on the USD/BRL test dataset. **ARIMA** achieves the highest average cumulative profit, and MSR. The monthly Sharpe Ratio and p-value indicates that **Chronos-Bolt** may be a viable trading strategy.

14.4. Discussion of Fixed-Size Position for USD/BRL

The results for the forecasting models for the USD/BRL exchange rate using fixed position sizing (Figure 4 and Table 8) show that, among the time series forecasting models, the ARIMA model has the highest average cumulative profit and highest average profit per trade, followed by the Ensemble model. Chronos-Bolt has the highest monthly Sharpe ratio. The confidence intervals for these models are also profitable. The overall structure of the average cumulative profits timeseries is remarkable similar to the active and passive Kelly criterion versions of trading strategies for USD/BRL.

The results for the news event models for the USD/BRL exchange rate using fixed position sizing show that, contrary to the USD/CNY results, among the news event models, the Fatouros_p1 model has the highest average cumulative profit, the highest average profit per trade, and the highest monthly Sharpe ratio. The confidence interval is also profitable. Surprisingly, FinBERT2 performs relatively poorly, even though the model is trained on financial information. Comparing the texts of the different prompts and the results of the news event models, we note that again the shorter prompts appear to perform better. The surprisingly good results of zero-shot LLM-based prompt trading necessitate further investigation to establish a definitive understanding of the relationship between prompts and trading performance in financial systems.

15. Related Work

15.1. Neural Network-Based Forecasting Models for Time Series Data

With the advent of deep learning, neural architectures such as LSTMs [27,28], Temporal Convolutional Networks [11–13], and Transformer-based sequence models [24,29–32] have demonstrated improved performance by learning long-range dependencies and non-linear interactions without requiring strict parametric assumptions. In foreign-exchange settings specifically, such models have been shown to predict short-term price movements and volatility surfaces with greater precision than purely statistical baselines [33].

15.2. Large Language Models Applied to Time Series Data

Recent efforts have sought to create foundational models for time series by adapting the Transformer-based architectures that have been successful in natural language directly to time series. A notable example is Chronos, a model family based on the T5 architecture [34]. To apply a language model to continuous time series, Chronos transforms the data into a sequence of discrete tokens. This transformation is achieved through a two-step process: the time series (in our case, the training part of the time series) is first normalized using mean-scaling, and then its values are quantized into a fixed set of bins, where the number of bins defines the model's vocabulary size. The model is then trained on these tokenized sequences using a standard cross-entropy loss, enabling it to perform probabilistic zero-shot forecasts on unseen data [14].

Chronos-Bolt, an update of the Chronos model, forgoes tokenization to incorporate patching, a method also used by models like PatchTST, where the input time series is segmented into smaller contiguous non-overlapping sub-sequences or "patches" [35–37]. In Chronos-Bolt, each patch is composed of 16 data points. These patches are fed to the model as input tokens. This technique seems to improve the model's ability to learn local patterns and significantly reduces computational load, thereby allowing it to process longer historical contexts.

In a similar vein, Toto, a foundation model developed by Datadog, leverages large-scale pretraining for zero-shot forecasting [38]. Toto is a Transformer model trained on one trillion time series data points, with datasets from both Datadog's internal observability metrics and publicly available time series collections [38]. Because Toto is trained on such a wide variety of time series, it appears to learn domain-agnostic temporal representations and capture multiscale trends [38].

15.3. News-Based Forecasting

The pure time series forecasting approaches operate in isolation from macroeconomic signals and news events. By contrast, much work has utilized natural language processing to extract predictive signals from news articles, financial disclosures, and social media networks. Early lexicon-based or bag-of-words approaches extracted polarity scores from news and social media to forecast equity or currency moves [39–41].

The advent of pretrained language models has markedly increased representational power. FinBERT2 [24] adapts BERT [30] to financial sentiment classification, while FinGPT [32] extends the capabilities of GPT [31] to a broad spectrum of financial tasks. Studies have suggested that sentiment extracted from highly credible outlets can serve as a leading indicator, often anticipating abnormal returns around macro announcements [42]. However, textual signals alone often lack the temporal precision to align cleanly with minute-level price movements, and their predictive power diminishes in periods of low news volume or market overreaction. Moreover, many sentiment-driven models are reactive because the news articles on which they are based capture only post-event shifts rather than anticipating them.

Recognizing these complementary weaknesses, recent studies combine price-based features with textual sentiment. Some fuse scalar sentiment scores with lagged returns in linear regressions or shallow networks [43]. Others embed news headlines jointly with technical indicators in deep architectures [44]. Although these hybrids tended to outperform

single-modality baselines and showed greater robustness during volatility spikes, some authors criticized them for relying on naive early concatenation, lack principled temporal alignment, or for only providing limited interpretability [33].

Following these developments and criticisms, recent work evaluated LLMs in financial sentiment classification tasks. Fatouros et al. [16] compared FinBERT [24], an encoder-only transformer adapted to the financial domain, with ChatGPT, a decoder-only model, in a zero-shot classification setting. Their evaluation used 2,291 manually annotated foreign exchange news headlines collected from ForexLive and FXStreet over an 86-day period (January - May 2023). The dataset spanned five major currency pairs (AUDUSD, EURCHF, EURUSD, GBPUSD, USDJPY). For 4 out of the 6 prompts that were tested, each headline was labeled according to its expected short-term directional impact. As for the rest, a daily sentiment based on the aggregate of the day's headlines was generated. We delve into the first four prompts and obtain sentiment-driven trading simulation results for the USDBRL and USDCNY tickers in Table 9 and 6 respectively.

More recently, Ballinari and Maly [18] explored the effectiveness of fine-tuning large language models for this task by adapting a Llama 3.1 8B model to a curated dataset of 251,845 FX analysis articles from Investing.com, DailyFX, and FXStreet. Their work demonstrates that domain-specific fine-tuning yields sentiment classifiers that outperform traditional lexicon-based methods and established benchmarks like FinBERT. A key contribution is the explicit distinction between past-oriented (descriptive) and future-oriented (predictive) sentiment within news text. The approach was validated through backtested trading strategies. A daily sentiment signal was derived from a simple majority vote over individual article sentiments, defaulting to a neutral stance when the counts of positive (+1) and negative (-1) articles were close. This signal generated superior risk-adjusted returns compared to baseline methods, including the non-finetuned version of Llama 3.1 8B. Simulating trades using the prompt and off-the-shelf model weights are shown in Tables 6 and 9 and for the US Dollar vs. Yuan and US Dollar vs. Brazilian Real, respectively.

16. Conclusions

Folk myths about foreign exchange trading come in two forms: (i) foreign exchange trading is for fools, because currencies are all manipulated, or (ii) fluctuations are entirely random. This work appears to validate myth (i) in an exchange rate that is likely manipulated (Chinese Yuan vs. US dollar). However, the work invalidates (ii) for uncontrolled (or imperfectly controlled) currencies (Brazilian Real vs. US dollar).

In the uncontrolled case, short-term (few minute) exchange rate changes can be predicted using both simple historical techniques (mean reversion and ARIMA) as well as machine learning techniques (like NHiTs) on the time series of exchange rates. The experiments also suggest that short zero-shot LLM prompts for news headlines (Fatouros_p3) can perform well.

We have also found two techniques that enhance these results: (i) hyperparameter tuning, especially the amount of history used in forecasting, matters a lot; and (ii) using the Kelly criterion based on the size of predicted profits and the size of historically realized profits proves to be helpful. The paper explored two versions of the Kelly Criteria: one which adjusts the amount to trade only based on the results of previous trades done by a trading strategy (Active Kelly) and one which gathers information to influence future trades even when the trading strategy does not trade (Passive Kelly). Active Kelly seems to work better, though we think both should be considered in future work.

The main future work is to understand *why* these forecasting models work well on time series. Speculations include: the herd behavior of traders may result in an underlying pattern which can be detected thanks to the large size of the modeling corpus. However, those are merely speculations. A similar question is to understand why simple news-related prompts based on headlines perform better than prompts based on financial expertise.

We invite researchers to use our software for other currencies and trading strategies. The code and pointers to the public data will be available upon publication.

Institutional Review Board Statement: Not applicable

Author Contributions: Conceptualization: E.H., A.T., D.S. Methodology: E.H., A.T., C.Z., D.S., B.L., E.L., Z.B., Y.H., A.G.A., Software: A.G.A., Z.B., Y.H., B.L., E.L., A.T., C.Z., Validation: D.S., E.H., A.T., B.L., Formal Analysis: E.H., A.T., D.S., B.L., Investigation: E.H., A.T., C.Z., D.S., B.L., E.L., M.S., Z.B., Y.H., A.G.A., Data Curation: B.L., Z.B., A.G.A., M.S., C.Z., Y.H., Writing – original draft preparation: E.H., A.T., D.S., B.L., E.L., Z.B., A.G.A., M.S., Writing – review and editing: E.H., A.T., D.S., B.L., Z.B., Y.H., A.G.A., Visualization: B.L., Supervision: E.H., A.T., D.S., Funding Acquisition: D.S.

Funding: NYU Wireless

Data Availability Statement: The software and data are available at https://github.com/bkmulusew/ml_fx_trading

Acknowledgments: We thank Mustafa Kocak and Sohail Hodarkar for helpful comments.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Triennial Central Bank Survey of foreign exchange and Over-the-counter (OTC) derivatives markets in 2022, 2022.
2. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time series analysis: forecasting and control*; John Wiley & Sons, 2015.
3. von Beschwitz, B. Internationalization of the Chinese Renminbi: Progress and Outlook. *FEDS Notes* **2024**, pp. 30–3.
4. Granger, C.W.; Joyeux, R. An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis* **1980**, *1*, 15–29.
5. Hosking, J. Fractional differencing. *Biometrika*, *68* (1), 165–176, 1981.
6. Beran, J. *Statistics for long-memory processes*; Routledge, 2017.
7. Baillie, R.T. Long memory processes and fractional integration in econometrics. *Journal of econometrics* **1996**, *73*, 5–59.
8. Hadad, E.; Hodarkar, S.; Lemeneh, B.; Shasha, D. Machine Learning-Enhanced Pairs Trading. *Forecasting* **2024**, *6*, 434–455. <https://doi.org/10.3390/forecast6020024>.
9. Oreshkin, B.N.; Carpov, D.; Chapados, N.; Bengio, Y. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437* **2019**. <https://doi.org/10.48550/arXiv.1905.10437>.
10. Challu, C.; Olivares, K.G.; Oreshkin, B.N.; Ramirez, F.G.; Canseco, M.M.; Dubrawski, A. Nhits: Neural hierarchical interpolation for time series forecasting. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2023, Vol. 37, pp. 6989–6997. <https://doi.org/10.1609/aaai.v37i6.25854>.
11. Lea, C.; Flynn, M.D.; Vidal, R.; Reiter, A.; Hager, G.D. Temporal convolutional networks for action segmentation and detection. In Proceedings of the proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 156–165.
12. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* **2018**.
13. Wan, R.; Mei, S.; Wang, J.; Liu, M.; Yang, F. Multivariate Temporal Convolutional Network: A Deep Neural Networks Approach for Multivariate Time Series Forecasting. *Electronics* **2019**, *8*. <https://doi.org/10.3390/electronics8080876>.
14. Ansari, A.F.; Stella, L.; Turkmen, C.; Zhang, X.; Mercado, P.; Shen, H.; Shchur, O.; Rangapuram, S.S.; Arango, S.P.; Kapoor, S.; et al. Chronos: Learning the Language of Time Series, 2024, [arXiv:cs.LG/2403.07815].
15. Cohen, B.; Khwaja, E.; Wang, K.; Masson, C.; Ramé, E.; Doubli, Y.; Abou-Amal, O. Toto: Time Series Optimized Transformer for Observability, 2024, [arXiv:cs.LG/2407.07874].
16. Fatouros, G.; Soldatos, J.; Kouroumali, K.; Makridis, G.; Kyriazis, D. Transforming sentiment analysis in the financial domain with ChatGPT. *Machine Learning with Applications* **2023**, *14*, 100508.
17. Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Vaughan, A.; et al. The Llama 3 Herd of Models, 2024, [arXiv:cs.AI/2407.21783].
18. Ballinari, D.; Maly, J. FX sentiment analysis with large language models. Working Papers 2025-11, Swiss National Bank, 2025. <https://doi.org/None>.
19. Massive. Massive: Real-time and Historical Market Data APIs, 2025. Accessed: 2025-08-29.
20. London Stock Exchange Group and Thomson Reuters. LSE Thomson Reuters Newswire Service, 2025. Accessed: 2025-08-29.
21. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman & Hall/CRC, 1993.
22. Good, P. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*; Springer Series in Statistics, 2005.
23. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization, 2017, [arXiv:cs.LG/1412.6980].
24. Araci, D. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063* **2019**.
25. Xu, X.; Wen, F.; Chu, B.; Fu, Z.; Lin, Q.; Liu, J.; Fei, B.; Li, Y.; Zhou, L.; Yang, Z. FinBERT2: A Specialized Bidirectional Encoder for Bridging the Gap in Finance-Specific Deployment of Large Language Models. In Proceedings of the Proceedings of the 31st

- ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2, New York, NY, USA, 2025; KDD '25, p. 5117–5128. <https://doi.org/10.1145/3711896.3737219>.
26. Leme, L. FinBERT-PT-BR: Financial BERT for Brazilian Portuguese. <https://huggingface.co/lucas-leme/FinBERT-PT-BR>, 2023. Accessed: 2024.
 27. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
 28. Fischer, T.; Krauss, C. Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research* **2018**, *270*, 654–669.
 29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems* **2017**, *30*.
 30. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 2019, pp. 4171–4186.
 31. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I.; et al. Improving language understanding by generative pre-training, 2018.
 32. Liu, X.Y.; Wang, G.; Yang, H.; Zha, D. Fingpt: Democratizing internet-scale data for financial large language models. *arXiv preprint arXiv:2307.10485* **2023**.
 33. 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* **2020**, *90*, 106181.
 34. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, 2023, [[arXiv:cs.LG/1910.10683](https://arxiv.org/abs/2303.18223)].
 35. Das, A.; Kong, W.; Sen, R.; Zhou, Y. A decoder-only foundation model for time-series forecasting. *arXiv preprint arXiv:2310.10688* **2023**.
 36. Zuo, Z.; Liu, Z.; Zheng, Z.; Huang, Z.; Zhu, S.; Yu, Z.; Kwok, J.T. AutoTimes: A foundation model for time-series forecasting. *arXiv preprint arXiv:2401.12345* **2024**.
 37. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers, 2023, [[arXiv:cs.LG/2211.14730](https://arxiv.org/abs/2311.14730)].
 38. Cohen, B.; Khwaja, E.; Doubli, Y.; Lemaachi, S.; Lettieri, C.; Masson, C.; Miccinilli, H.; Ramé, E.; Ren, Q.; Rostamizadeh, A.; et al. This Time is Different: An Observability Perspective on Time Series Foundation Models, 2025, [[arXiv:cs.LG/2505.14766](https://arxiv.org/abs/2505.14766)].
 39. Loughran, T.; McDonald, B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of finance* **2011**, *66*, 35–65.
 40. Nassirtoussi, A.K.; Aghabozorgi, S.; Wah, T.Y.; Ngo, D.C.L. Text mining for market prediction: A systematic review. *Expert Systems with Applications* **2014**, *41*, 7653–7670.
 41. Jeon, Y.; McCurdy, T.H.; Zhao, X. News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Journal of Financial Economics* **2022**, *145*, 1–17.
 42. Tetlock, P.C. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance* **2007**, *62*, 1139–1168.
 43. Zhang, W.; Skiena, S. Trading strategies to exploit blog and news sentiment. In Proceedings of the Proceedings of the international AAAI conference on web and social media, 2010, Vol. 4, pp. 375–378.
 44. Ding, X.; Zhang, Y.; Liu, T.; Duan, J. Deep learning for event-driven stock prediction. In Proceedings of the Ijcai, 2015, Vol. 15, pp. 2327–2333.

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