QUANTITATIVE METHODS IN ECONOMICS
Volume XXVII, No. 1, 2026, pp. 1 – 26

Received: 02.02.2026
Accepted: 25.03.2026

BEATING THE MACHINE: EVALUATING GEMINI LLM AND SEASONAL RANDOM WALK MODELS FOR EARNINGS PER SHARE FORECASTING IN POLAND

Wojciech Kuryłek https://orcid.org/0000-0003-0692-3300
University of Warsaw, Faculty of Management, Warsaw, Poland
e-mail: wkurylek@wz.uw.edu.pl

Abstract. Emerging markets like Poland face limited analyst coverage, with only 20% of listed companies receiving professional scrutiny, necessitating automated forecasting methods. This study investigates whether Large Language Models can outperform traditional statistical approaches in predicting earnings per share (EPS). We compare Google's Gemini LLM against the seasonal random walk (SRW) model using quarterly EPS data from 267 Polish companies (2010-2019). Performance is evaluated using Mean Arctangent Absolute Percentage Error (MAAPE), with robustness checks using RMSE and MAE metrics. Results reveal a notable discrepancy: SRW achieves superior MAAPE scores, while Gemini excels on RMSE and MAE measures. This metric-dependent performance suggests model selection should align with specific error tolerance requirements. Despite Gemini's sophisticated chain-of-thought reasoning mimicking analyst approaches, simpler models prove equally effective in markets with unsophisticated EPS dynamics. These findings challenge assumptions about LLM superiority in financial forecasting and highlight the continued relevance of parsimonious statistical methods in emerging markets.

Keywords: earnings per share, seasonal random walk, Large Language Model, LLM, Gemini, Warsaw Stock Exchange

JEL codes: C01, C02, C12, C14, C58, G17

INTRODUCTION

The fundamental determination of corporate equity valuations is intrinsically linked to the product of the Price-to-Earnings (P/E) ratio and Earnings Per Share (EPS), constituting a critical component in investment evaluation methodologies. Precise

https://doi.org/10.22630/MIBE.2026.27.1.1

prognostication of these financial indicators, with particular emphasis on EPS, remains imperative as they provide essential quantitative data regarding an organization's future performance, thereby informing market valuation estimations and establishing parameters for audit expectations. The significant disparity between mature financial markets, exemplified by the United States with its comprehensive analyst coverage, and developing markets such as Poland, where merely one-fifth of listed entities receive professional analytical scrutiny, necessitates the implementation of computational and algorithmic approaches for EPS prediction in these less-covered economies.

This study addresses two primary research questions: (1) Can Large Language Models trained on extensive textual data provide more accurate EPS forecasts than traditional time series models for Polish listed companies? (2) How does the choice of error metric influence the comparative assessment of LLM versus statistical model performance? The hypothesis of Gemini superiority is partially supported: rejected when evaluated by MAAPE but supported under RMSE and MAE metrics. This work represents the inaugural application of LLMs combining textual and time series inputs for EPS forecasting. The deployment of these advanced architectures for earnings prediction constitutes uncharted territory in academic research, filling an important scholarly void. The outcomes illuminate the capabilities of AI-based techniques for financial analysis, particularly benefiting markets with sparse analyst attention.

The research herein distinctively investigates the utilization of a Large Language Model (LLM) for temporal sequence prediction of EPS values. Large Language Models demonstrate considerable capacity for processing and analyzing extensive unstructured datasets, representing sophisticated instruments for financial examination. The investigation specifically employs Google's Gemini-2.0-flash-lite model, incorporating comprehensive prompting techniques and formalized output structures. The analytical framework encompasses quarterly EPS information from 267 corporations listed on the Polish securities exchange, with temporal boundaries extending from the global financial crisis of 2008-2009 through the COVID-19 pandemic of 2020.

This investigation adopts the mean arctangent absolute percentage error (MAAPE) as introduced by Kim and Kim (2016) as an evaluation criterion, rather than exclusively utilizing the conventional mean absolute percentage error (MAPE) methodology, which exhibits susceptibility to distortion when denominators approach zero, thus enabling more accurate assessment of predictive performance.

The scholarly contribution of this work lies in its pioneering application of LLM technology to financial forecasting, specifically in the domain of EPS prediction, representing one of the earliest research endeavors in this field. The implementation of such sophisticated computational frameworks for this predictive task constitutes an unexplored area within the current academic literature, thereby addressing a significant research lacuna. The empirical findings offer meaningful perspectives regarding the efficacy of artificial intelligence-driven methodologies

for financial prognostication. Furthermore, the research enhances comprehension of EPS forecasting mechanisms within emerging economic contexts, with particular attention to the Polish market, a domain that has received comparatively limited academic attention relative to established markets such as the United States. Multiple error measurement techniques, chronological periods, and statistical verification procedures are utilized to substantiate the conclusions, while simultaneously elucidating practical ramifications for investment tactics involving Polish equities.

The purpose of this study is to evaluate whether Large Language Models, specifically Google's Gemini, can provide superior earnings per share (EPS) forecasting accuracy compared to traditional statistical methods for companies listed on the Warsaw Stock Exchange. This investigation aims to determine if the sophisticated capabilities of LLMs offer meaningful advantages over simpler approaches like the seasonal random walk model in markets characterized by limited analyst coverage and relatively unsophisticated EPS dynamics.

LITERATURE REVIEW

The trajectory of Earnings per Share (EPS) forecasting techniques commenced in the 1960s with algorithmic approaches, predominantly utilizing autoregressive integrated moving average (ARIMA) frameworks. These methodologies were investigated by numerous scholars including Ball and Watts [1972], Watts [1975], Griffin [1977], Foster [1977], and Brown and Rozeff [1979]. The efficacy of these analytical structures showed considerable variation; certain investigations indicated that basic random walk models often demonstrated comparable performance to more intricate systems, while alternative research yielded divergent conclusions. Analogous examinations for the Polish financial environment were subsequently conducted by Kuryłek (2023a).

The prominence of ARIMA frameworks persisted due to their predictive precision, as documented by Lorek [1979] and Bathke and Lorek [1984], until the latter part of the 1980s when a prevailing academic consensus emerged suggesting that forecasts generated by financial analysts exceeded the accuracy of time series model predictions [Brown et al., 1987]. The research of Conroy and Harris [1987] emphasized that analysts' projections demonstrated superior performance for immediate temporal horizons, though this advantage diminished across extended timeframes. This academic perspective remained largely unchallenged until contemporary scholarship [Lacina et al., 2011; Bradshaw et al., 2012; Pagach and Warr, 2020; Gaio et al., 2021] began to reevaluate the purported superiority of analyst forecasts compared to time series methodologies.

Concurrent with ARIMA developments, exponential smoothing techniques for EPS prediction have been explored since the late 1960s by various researchers including Elton and Gruber [1972], Ball and Watts [1972], Johnson and Schmitt [1974], Brooks and Buckmaster [1976], Ruland [1980], Brandon et al. [1987], and

Jarrett [2008], yielding inconsistent outcomes. Within the Polish market context, Kuryłek [2023b] conducted parallel investigations into these methodologies.

Multivariate cross-sectional frameworks incorporating fundamental indicators derived from financial ratios have demonstrated superior performance compared to firm-specific and common-structure ARIMA models in earnings forecasting. This superiority has been documented in research by Lorek and Willinger [1996], Lev and Thiagarajan [1993], Abarbanell and Bushee [1997], Cao and Gan [2009], Cao and Parry [2009], Ahmadpour et al. [2015], Ball and Ghysels [2017], Hou et al. [2012], and Li and Mohanram [2014]. Nevertheless, Lev and Sougiannis [2010] observed the constrained long-term effectiveness of estimate-based accounting elements. Conceptual frameworks for earnings prediction were established by Ohlson [1995, 2001], Pope-Wang [2005, 2014], and Li [2011], with Harris and Wang [2019] confirming the enhanced accuracy and reduced bias exhibited by Pope and Wang's [2005] model.

Contemporary research has increasingly concentrated on artificial neural networks for EPS forecasting, producing diverse results. Atiya et al. [1997] were pioneers in applying neural networks based on fundamental characteristics for stock price prediction, consistently outperforming alternative models. Cao et al. [2004] conducted comparative analyses of neural feedforward networks (MLP), demonstrating their superior accuracy relative to other forecasting methodologies. Conversely, Lai and Li [2006] reported that an ANN framework exhibited the least satisfactory accuracy for EPS predictions. Research by Cao and Parry [2009] illustrated that univariate neural network models surpassed linear regression frameworks and that genetic algorithms excelled in determining network weights. Cao and Gan [2009] verified the enhanced performance of neural network models optimized with genetic algorithms for predicting EPS of Chinese listed enterprises. Gupta et al. [2013] identified an optimal multiperceptron architecture for stock market price forecasting, emphasizing the significance of EPS and public sentiment. Ahmadpour et al. [2015] successfully employed standard multilayer perceptron (MLP) neural networks, achieving substantially higher accuracy with extracted rules compared to pure MLP frameworks. Chen et al. [2020] assessed various methodologies for EPS prediction, including decision trees and radial basis function networks, demonstrating the superior accuracy of ensemble approaches. Elend et al. [2020] contrasted Long Short-Term Memory (LSTM) networks with Temporal Convolution Networks (TCNs) for EPS prediction, with LSTM significantly enhancing accuracy compared to naive persistent models. Xiaoqiang [2022] furnished a comprehensive overview of machine learning techniques applicable to financial ratio forecasting, including EPS.

In the most recent academic landscape, Large Language Models (LLMs) have garnered significant scholarly attention. Cao et al. [2024] demonstrated that LLMs, which extract more comprehensive and predictive content from earnings conference calls, outperform traditional analytical benchmarks in forecasting stock price volatility. Kong et al. [2024] elucidated both the capabilities and limitations of

finance-specific large language models in addressing complex tasks, while identifying critical challenges including data quality issues, modeling complexities, and ethical considerations. Research by Abe et al. [2024] examined how LLMs can predict price movements in stock and bond portfolios utilizing economic indicators, revealing that LLM-based strategies, particularly when combined with mode ensemble techniques, outperform buy-and-hold strategies in terms of Sharpe ratio during periods of increasing consumer price index (CPI). Sarker (2024) expressed concerns regarding the trustworthiness of LLMs, characterizing these models as opaque systems. Cao and Wang [2024] emphasized the principal challenges encountered by LLMs in time series prediction contexts, observing that predictive accuracy significantly decreases when confronted with heterogeneous time series data and traditional signals containing both periodic and trend components, as well as when signals comprise complex frequency elements. Abdelsamie and Wang [2024] conducted comparative analyses of market prediction accuracy between LLM-based systems and human expertise within financial analysis domains, demonstrating that their specialized model achieved superior accuracy and efficiency in financial forecasting compared to human predictions, particularly in dynamic, information-rich contexts. Nevertheless, their research acknowledged that limitations in nuanced contextual comprehension and adaptability persist, underscoring the enduring value of human expertise. It merits emphasis that the existing literature contains no studies specifically dedicated to EPS forecasting utilizing LLMs.

The inaugural investigation of statistical forecasting methodologies for EPS among Polish listed companies was conducted by Kurylek [2023a]. This research, which employed SARIMA-type models, determined that the seasonal random walk (SRW) most effectively captures the behavior of Earnings Per Share in Poland, with no alternative model demonstrating superior performance. This finding received further corroboration for models within the exponential smoothing family [Kuryłek, 2023b]. In subsequent research, Kurylek (2024a) illustrated that even contemporary time series models developed by major technology organizations – including Facebook's Prophet, LinkedIn's SilverKite, Amazon's DeepAR, and Google's TFT – fail to exceed the performance of the SRW model in the Polish market context. Similarly, various Artificial Neural Network architectures, whether implemented in time series or multivariate configurations, demonstrated no enhancement in results [Kuryłek, 2024b; Kuryłek, 2024c]. Additionally, Kuryłek [2025] investigated the potential application of natural language processing (NLP) techniques, including FastText and FinBERT word embeddings combined with the gradient-boosting decision tree algorithm XGBoost, for EPS forecasting in Poland. This investigation concluded that the implementation of these sophisticated NLP methodologies may not be justified, as the SRW model provides a more accurate representation of market behavior. These investigations collectively indicate that neither statistical frameworks nor more advanced machine learning and deep learning methodologies have demonstrated the capacity to outperform the elementary seasonal random walk

(SRW) model in Poland. The current research examines the application of Gemini LLM (Google DeepMind, 2023) to EPS forecasting.

DATA AND METHODS

Time Series Data

Following its European Union integration in 2004, the Polish stock market demonstrated remarkable expansion, reaching a capitalization of $197 billion by late 2021 with 774 companies listed. However, analytical coverage remains deficient compared to Western European and American markets, with merely 20% of listed entities receiving analyst evaluation as of 2019. This analytical deficit highlights the necessity for implementing machine learning and statistical methodologies to forecast crucial financial metrics. The current investigation concentrates on earnings per share (EPS) information extracted from the financial analysis system EquityRT.

The investigation analyzes EPS patterns for Warsaw Stock Exchange corporations spanning from Q1 2010 through Q4 2019, positioned between two major disruptive events: the 2008-2009 financial downturn and the 2020 COVID-19 outbreak. Global market volatility has intensified substantially since 2020, initially due to the pandemic and subsequently aggravated by energy instability following the Ukrainian conflict in 2022. This context establishes 2019 as the final year of comparative stability, rendering it an appropriate foundation for model calibration and validation.

The deliberate and strategic selection of this stable timeframe was intentional. A statistical or machine learning approach, including any Large Language Model, that performs inadequately during periods of stability would likely produce unreliable outcomes in volatile conditions. Stable periods offer optimal environments where patterns are more discernible, noise is reduced, and variable relationships maintain relative consistency. Models failing under stable circumstances presumably lack capacity to identify essential dependencies, trends, or fundamental data structures. During environmental volatility, these limitations become amplified, as regime shifts, sudden disruptions, and nonlinear dynamics introduce complexities beyond simple historical extrapolation. Models exclusively based on historical trends may inadequately interpret rapid changes, generating significant errors. Additionally, volatility typically enhances uncertainty, emphasizing requirements for flexible models capable of distinguishing between temporary fluctuations and enduring structural alterations. Models demonstrating fragility in stable environments become even less reliable under unpredictable circumstances, underscoring the importance of establishing dependable baselines before application in challenging scenarios.

For forecasting purposes, data encompassing Q1 2010 through Q4 2018 (36 quarters total) serves as model training material, while Q1 2019 through Q4 2019 is reserved for out-of-sample validation. The forecasting framework includes

projections ranging from one to four quarters ahead, with additional validation utilizing 2017 and 2018 samples. This dataset provides substantial accounting information, even for publicly traded entities issuing quarterly financial statements, representing at minimum nine complete years of EPS records. Firms that discontinued publishing financial reports within the study horizon were removed from the sample. Following the application of full-time window coverage and the exclusion of stock splits and reverse splits, the final sample consisted of 267 companies. Subsequently, these EPS time series undergo transformation into textual format and incorporation into prompts for Large Language Model processing.

Textual Data

The Elektroniczny System Przekazywania Informacji (ESPI) operates as an electronic disclosure framework through which public entities on the Warsaw Stock Exchange's (WSE) primary market fulfill mandatory reporting obligations. This system facilitates transparent and immediate dissemination of critical information encompassing financial outcomes, substantial ownership modifications, organizational developments, and additional relevant occurrences that might influence securities valuation or investment determinations. The Polish Financial Supervision Authority (Komisja Nadzoru Finansowego, KNF) maintains regulatory oversight regarding ESPI reporting compliance among WSE-listed corporations, thereby enforcing market transparency standards and investor protection measures.

Concurrently, the Elektroniczna Baza Informacji (EBI) functions as a complementary digital information dissemination infrastructure predominantly utilized by corporations trading on NewConnect – an alternative exchange platform administered by the Warsaw Stock Exchange (WSE) designed for emerging enterprises and developmental ventures, particularly those in formative phases or technology sectors. The EBI framework ensures punctual and transparent communication of fundamental corporate information to the investment community and general public. While operational management of the EBI system resides with the WSE, the KNF exercises supervisory authority concerning compliance with disclosure requirements, thus maintaining market integrity and safeguarding investor interests.

HTML5 (Hypertext Markup Language version 5) serves as the technological foundation for both ESPI and EBI systems, ensuring report accessibility, legibility, and interactivity across diverse computational devices and internet browsers. The Python software package html2text performs conversion of individual corporate HTML documents into simplified, comprehensible ASCII text reports. These textual documents are chronologically arranged for each corporate entity per quarter and subsequently consolidated into unified quarterly textual compilations. The resultant textual data is thereafter incorporated into prompts that subsequently serve as input for Large Language Models (LLM).

Models

The seasonal random walk model (SRW)

The SRW can be described as:

$$EPS_t = EPS_{t-4} + \varepsilon_t \text{ where } \varepsilon_t \text{ are IID and } \varepsilon_t \sim N(0, \sigma^2) \quad (1)$$

The predictive methodology employs the earnings value from the corresponding quarter of the previous year as its prognostic output, thus circumventing any requirement for parametric estimation procedures. It implies that for most of the stocks, the best predictor of the future EPS is the EPS in the analogous quarter of the previous year. This particular forecasting technique functions as a comparative standard, whose efficacy surpassing that of alternative temporal sequence analytical models within the Polish financial marketplace has been substantiated through empirical investigations conducted by Kuryłek (2023a, 2023b, 2024). The research corpus established by Kuryłek across multiple publications demonstrates the relative advantages of this parameter-free approach when applied specifically to Polish market conditions.

Large Language Models (LLMs)

Large Language Models (LLMs) represent a significant advancement in the natural language processing (NLP) domain. The developmental trajectory of these computational systems has witnessed a transformation from rudimentary frameworks to intricate architectures. Contemporary models are engineered to interpret and generate human-like linguistic outputs through the utilization of comprehensive datasets and sophisticated neural network configurations.

The fundamental processing methodology encompasses the segmentation of textual inputs into tokens (lexical units or sub-lexical components), subsequently transformed into numerical representations – a dual procedure termed tokenization and word embedding. Sequential information is preserved via positional encoding, which integrates the ordinal arrangement of tokens, thus maintaining syntactic integrity. The structural framework consists of multiple hierarchical arrangements of self-attention mechanisms and feed-forward recurrent neural networks, with each successive layer enhancing the representational fidelity of input tokens. As Gomez (2017) elucidates, self-attention mechanisms allocate relative importance to diverse tokens within input sequences, effectively capturing extended contextual dependencies. This capability directs the model's focus toward contextually pertinent input segments, culminating in the conceptualization of transformer architecture, initially formulated as an encoder-decoder paradigm.

Transformers were conceptualized with encoder-decoder dualities for specialized linguistic tasks such as translation. However, contemporary LLMs adopt a decoder-exclusive variant optimized for autoregressive textual production. This streamlined transformer iteration processes sequential data while minimizing computational complexity. The architecture employs self-attention mechanisms alongside feed-forward recurrent neural networks for the interpretation and

generation of sequential data. This configuration demonstrates particular efficacy in generating contextually coherent output sequences predicated on antecedent tokens – LLMs undergo training to anticipate subsequent linguistic elements across extensive textual corpora. In contrast to comprehensive attention mechanisms in encoders, decoders implement masked self-attention, ensuring that positional predictions are exclusively informed by preceding tokens, a critical feature for autoregressive functionalities. The dependence on previously generated sequences for contextual comprehension may present challenges with extensive sequential inputs, potentially resulting in diminished retention of initial information as sequence processing advances.

The training methodology incorporates vast textual corpora to predict subsequent tokens in sequences, thereby acquiring linguistic patterns, grammatical structures, factual information, and rudimentary reasoning capabilities. Training datasets typically encompass diverse textual sources including literary works, academic publications, digital content, and miscellaneous textual repositories. The training protocol leverages unstructured textual data, utilizing inherent contextual information for supervisory guidance.

Despite technological advancements, LLMs exhibit certain limitations. The assimilation of extensive sequential data may result in the propagation of inherent biases present within training corpora, with bias mitigation remaining an active research domain. While LLMs demonstrate proficiency in generating innovative textual formats, occasional factual inaccuracies - termed "hallucinations" - may manifest in generated content. Furthermore, substantial computational resources and memory allocation are requisite, particularly during the training phase.

The GPT (Generative Pre-trained Transformer) series represents a preeminent exemplar within the LLM category. In 2018, OpenAI introduced the inaugural GPT model, employing a transformer-based architecture featuring unsupervised pre-training followed by task-specific optimization. The subsequent iteration, GPT-2, emerged in 2019 with expanded parametric dimensions and augmented training data, achieving superior performance across various NLP tasks without specialized fine-tuning. Released in 2020, GPT-3 further expanded to 175 billion parameters, exhibiting exceptional linguistic comprehension and generative capabilities with minimal additional training. The parametric scale of GPT-4, introduced in 2023, remains undisclosed.

In their comparative analysis, Hou and Lian (2024) examine three prominent large language models - ChatGPT, Mistral, and LLaMA - assessing their performance across dimensions of computational efficiency, linguistic precision, and ethical alignment. Research findings reveal distinctive strengths and specific constraints associated with each model: ChatGPT demonstrates superiority in linguistic accuracy, LLaMA exhibits enhanced adaptability across multilingual contexts, while Mistral introduces novel approaches to complex linguistic processing. This comparative evaluation provides substantive insights into contemporary LLM capabilities.

Current research endeavors aim to further expand these models, including investigations into multi-trillion parameter architectures. Despite persistent challenges including computational demands and inherent biases, LLMs continue their evolutionary trajectory, fostering innovations in machine-based linguistic comprehension and generation. These computational systems, exemplified by models such as GPT, have transformed the NLP landscape through the implementation of transformer architectures, facilitating advanced linguistic interpretation and production. Notwithstanding ongoing technical and ethical challenges, these models continually expand the boundaries of machine-facilitated natural language processing.

Gemini (GEMINI)

The research employs Google DeepMind's Gemini (version 2.0-flash-lite), a series of multimodal large language models (LLMs) introduced in 2023. Forecasts for earnings per share (EPS) are generated through API interfaces, necessitating 1,068 distinct API calls (representing 267 companies across 4 quarters) to produce comprehensive annual projections. These models were estimated individually using the data of each company.

Prompts

Input texts or instructions, known as prompts, are provided to the LLM to elicit specific outputs. While user prompts address particular tasks, system prompts establish the overarching operational framework across all interactions, functioning as the foundation for the AI's behavioral patterns and output generation.

```
system_prompt = f"""
You are an AI expert specializing in making financial predictions based on
historical quarterly data, and all information available until {Date:%Y-
%m-%d}.
Your task is to analyze the provided historical data and make a prediction
of
Earnings Per Share (EPS) exactly one year ahead.

Do the following steps:
0. Convert all relevant information into the Polish Zloty, if this
information
   is provided in other currency.
1. Find factors that will influence the future revenues of companies and
make
   the forecast of future revenues. -> Revenues
2. Find factors that will influence the future costs of companies and make
   the forecast of future costs. -> Costs
3. Calculate the tax factor for the last quarters by dividing Net Profit
by
   Gross Profit (Net Profit / Gross Profit) and make the forecast of tax
   factor in the future. -> Tax Factor
4. Find the information about number of Shares Outstanding in the last
quarter.
   Assume that Shares Outstanding should remain constant in the future.
   -> Shares Outstanding
5. Calculate the future EPS as (Revenues - Costs) * Tax Factor / Shares
Outstanding
"""
```

```
user_prompt = f"""
Forecast EPS for the anonymized ticker: {anonymized_ticker}
Available historical EPS data: {historical_EPS_data}
ESPI reports for the available last 5 quarters:
ESPI report 0: date of information_cut: {ending_date_0:%Y-%m-%d}
   espi reports: {espi_reports_0}
...
ESPI report 4: date of information_cut: {ending_date_4:%Y-%m-%d}
   espi reports: {espi_reports_4}
Examples of quarterly data and forecasts made on the basis of them:
Example 0: ... Example 4: ...
Make the forecast accurately and consistently. Do not hallucinate.
REMEMBER: Return ONLY valid JSON with the structure specified.
"""
```

The implemented system prompt instructs the AI to generate EPS predictions utilizing all available information up to a specified date, projecting one year forward with EPS forecasts decomposed into revenues, costs, and tax components. Tickers and names of companies are intentionally anonymized to prevent from the data leakage between the LLM's training data and the test period. The user prompt incorporates five forecast examples utilizing actual EPS figures, along with historical EPS data and ESPI reports spanning from five quarters to one year prior to the forecast date. These prompts implement Python's f-strings (string literals with an 'f' prefix containing values in curly braces {} that are interpolated into the text). Multiline text elements are denoted using triple quotes.

Structured output

The model generates structured/JSON formatted output utilizing the pydantic library in Python. Beyond EPS projections, the AI provides reasoning supporting each forecast and a certainty metric ranging from 0 to 1, indicating confidence levels. As Chong and Kao (2025) observe, in volatile financial markets where professional judgment frequently guides decision-making, the quality of reasoning can supersede predictive accuracy in importance. Consequently, explanatory model reasoning was deemed essential, implemented as a chain of thought for each prediction. This technique enables Large Language Models to decompose complex problems into sequential steps, explicitly articulating intermediate reasoning before reaching conclusions, thereby enhancing problem-solving accuracy and making logical processes transparent to users. Additionally, the model assigns a certainty score between 0 and 1 to each forecast.

Setting temperature

The temperature parameter critically balances predictability and creativity in generated text. Lower settings prioritize established patterns for more deterministic results, while higher values promote exploration, enhancing output diversity. Temperature was calibrated to minimize Mean Arctangent Absolute Percentage Error (MAAPE) for 2019, with comparative testing at settings of 0.0, 0.2, and 0.4, yielding MAAPE values of 0.743, 0.725, and 0.731 respectively. The empirical analysis identified 0.2 as the optimal temperature parameter.

Mean Arctangent Absolute Percentage Error (MAAPE)

For a specific corporate entity i, the empirical earnings per share (EPS) throughout the quarterly sequence of 2019 may be denoted as A1, ..., A4. The corresponding prognostic values for these temporal intervals facilitate the computation of forecast precision via the absolute percentage error (APE) for any quarterly period j in 2019 for any given entity i, as expressed by the mathematical formulation:

$$APE_j^i = \left|\frac{A_j^i - F_j^i}{A_j^i}\right| \tag{2}$$

Despite its prevalent application, the absolute percentage error (APE) exhibits a notable limitation: the metric becomes mathematically indeterminate or infinite when actual values approximate zero – a phenomenon not uncommon in earnings prediction contexts. Furthermore, when actual numerical outcomes are exceptionally minimal, particularly below unity, the resultant percentage discrepancies can become disproportionately amplified, creating statistical anomalies. This methodological constraint is exacerbated when actual values equal zero, producing mathematically undefined or infinite APE calculations. Addressing this methodological deficiency, Kim and Kim (2016) proposed the arctangent absolute percentage error (AAPE), offering a more methodologically sound alternative for such scenarios in predictive analytics.

$$AAPE_j^i = \begin{cases} 0 & if\ A_j^i = F_j^i = 0 \\ arctan\left(\left|\frac{A_j^i - F_j^i}{A_j^i}\right|\right) & otherwise \end{cases} \tag{3}$$

This methodological refinement utilizes the arctangent function's mathematical property of transforming values spanning from negative to positive infinity into the bounded interval [–π/2, π/2]. Thus, the Mean Arctangent Absolute Percentage Error (MAAPE) for the j-th quarterly period across the entirety of I corporate entities within the analytical sample can be mathematically expressed as:

$$MAAPE_j = \frac{1}{I}\sum_{i=1}^{I} AAPE_j^i = \frac{1}{I}\sum_{i=1}^{I} arctan\left(\left|\frac{A_j^i - F_j^i}{A_j^i}\right|\right) \tag{4}$$

The preference for MAAPE (Mean Arctangent Absolute Percentage Error) over MAPE (Mean Absolute Percentage Error) is justified by the inclusion of corporate entities with actual profitability approaching zero within the examined dataset. In scenarios where even a singular observation approximates zero, while remaining observations exhibit substantially higher values, MAPE can escalate to extraordinarily elevated magnitudes, approaching mathematical infinity. This statistical phenomenon distorts the mean calculation process, effectively diminishing the statistical significance of other observational data points.

The statistical test

To assess the statistical relevance of Mean Arctan Absolute Percentage Error disparities between models, Wilcoxon's [1945] nonparametric methodology has been implemented. This analytical approach serves as a comparative examination technique for paired samples with interdependence, eschewing distributional presuppositions beyond symmetrical characteristics and independent difference scores. The application of the Wilcoxon procedure in validation contexts, particularly for establishing statistical significance in error differentials across Ensemble Prediction System models, was comprehensively examined by Ruland [1980]. For analytical purposes, distinct probability value tables are constructed quarterly (spanning the first through fourth quarters) as well as for the aggregated quarterly information.

$$H_o: AAPEs\ of\ a\ pair\ of\ models\ are\ the\ same \quad (5)$$

The null hypothesis posits equality between Arctan Absolute Percentage Errors for model pairs. Probability values below the predetermined alpha threshold of 0.05 necessitate rejection of this null hypothesis in each analytical instance. This significance criterion maintains widespread recognition and validation within statistical literature, as substantiated by Ruland [1980] among other scholarly sources.

RESULTS

Empirical Findings

The empirical analysis presented in Table 1 demonstrates that the seasonal random walk (SRW) methodology yields superior performance, as measured by the Mean Arctangent Absolute Percentage Error (MAAPE) criterion, relative to the GEMINI framework. This pattern of superiority manifests across nearly all quarterly periods and throughout the entirety of 2019, with the exception of the second quarter where Gemini marginally outperforms SRW.

Table 1. Summary statistics on forecast errors for 2019 quarters

model	Q1 MAAPE	Q2 MAAPE	Q3 MAAPE	Q4 MAAPE	Total MAAPE
SRW	0.658	0.702	0.653	0.736	0.687
GEMINI	0.746	0.700	0.666	0.787	0.725

Source: own calculations

Statistical validation through Wilcoxon testing was implemented to determine the significance of performance differentials between the aforementioned methodologies, with corresponding p-values for each temporal segment catalogued in Table 2. The analytical evidence reveals an absence of statistically meaningful disparities in forecast accuracy between SRW and GEMINI methodologies during

three quarterly periods. However, a statistically significant divergence emerges specifically in the first quarter of 2019, as well as when considering the comprehensive annual performance. Consequently, the empirical findings present an mixed pattern.

Table 2. P-values of the Wilcoxon test of forecast errors for SRW and GEMINI in 2019

	Q1	Q2	Q3	Q4	ALL
p-value	0.008	0.972	0.801	0.224	0.025

Source: own calculations

Robustness Checks

Longitudinal examination spanning the years 2017, 2018, and 2019 reveals that the seasonal random walk (SRW) methodology consistently exhibited superior predictive accuracy compared to the GEMINI approach, as documented in Table 3. Table 4 indicates that while statistically significant differences in error metrics between these methodologies were observed in 2019, comparable statistical significance was not detected in either 2017 or 2018.

Table 3. Summary statistics on forecast errors for whole years 2017–2019

model	2017 MAAPE	2018 MAAPE	2019 MAAPE
SRW	0.686	0.711	0.687
GEMINI	0.703	0.728	0.725

Source: own calculations

Table 4. P-values of paired Wilcoxon test of forecast errors for whole years 2017–2019

	2017	2018	2019
p-value	0.064	0.338	0.025

Source: own calculations

The robustness verification protocol incorporated two additional prominent performance metrics. Table 5 presents an assessment utilizing alternative error quantification methodologies: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). This comprehensive evaluation encompasses all quarterly segments of 2019. To ensure comparative equity, Consumer Price Index (CPI) adjustments were implemented to standardize the present values of future nominal errors with contemporary error measurements.

Contrary to initial expectations, as illustrated in Table 5, the GEMINI methodology generates the most favorable results when evaluated via RMSE and MAE criteria, with the seasonal random walk model achieving marginally inferior performance. The Wilcoxon test, as presented in Table 6, confirms that the error differentials between these two methodological approaches exhibit statistically significant variations.

Table 5. Summary statistics on forecast errors for RMSE and MAE in all quarters 2019

	SRW	GEMINI
RMSE	0.937	0.861
MAE	0.705	0.655

Source: own calculations

Table 6. P-values of paired Wilcoxon test of forecast errors for RMSE and MAE in 2019

	RMSE	MAE
p-value	0.011	0.011

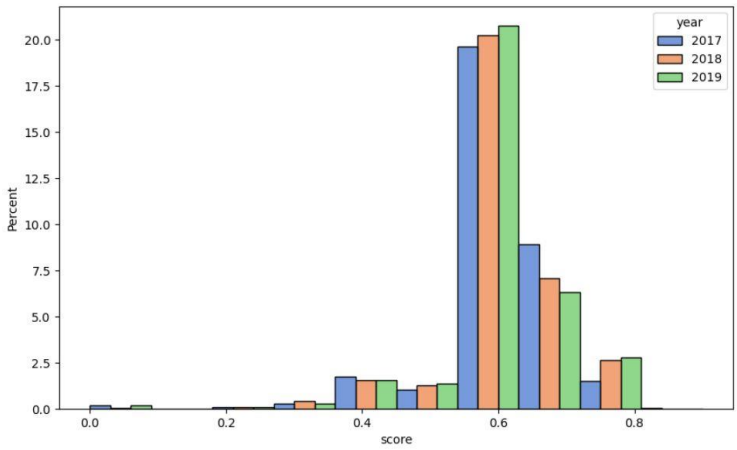
Source: own calculations

In conclusion, the empirical evidence suggests that from a MAAPE perspective, the elementary random walk model represents the optimal methodological choice for the Polish market context when compared to the artificial intelligence-based GEMINI approach. Conversely, alternative error metrics including MAE and RMSE indicate an inverse relationship regarding methodological superiority.

Certainty of forecasts

The predictive confidence metric for generative artificial intelligence systems, specifically large language models (LLMs), is quantified on a scale between zero and unity, where the extremes denote absolute uncertainty and complete conviction, respectively. This numerical assessment enables stakeholders to evaluate the dependability of model-generated content, particularly in contexts where decisions of consequence must be made.

Figure 1. Distribution of forecast certainty scores



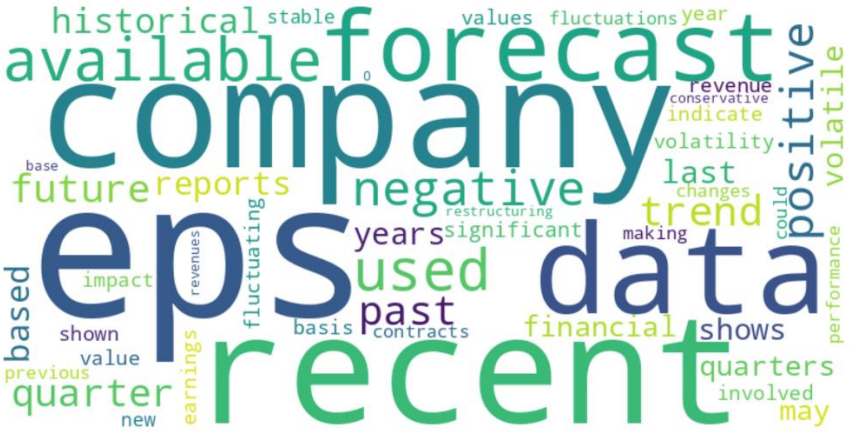
Source: own calculations

As demonstrated in Figure 1, the frequency distribution of predictive confidence metrics from the Gemini system throughout the 2017-2019 timeframe reveals that values are predominantly accumulated in the vicinity of 0.6 across all three annual periods. Substantial frequency concentration is also observed in the 0.7-0.8 range, whereas minimal distributional presence is evident in the lower confidence spectrum beneath 0.4, signifying that the model infrequently produces predictions characterized by low self-assessed reliability. The distributional characteristics suggest that the LLM consistently generates prognostications with intermediate to elevated confidence levels, with an observable progression toward more densely clustered confidence metrics over the three-year observational interval.

Chain of thought

The implementation of sequential reasoning mechanisms within sophisticated neural language frameworks facilitates the generation of intermediary analytical phases, thereby providing transparent explanatory pathways regarding predictive determinations. This methodological approach augments comprehensibility by deconstructing intricate prognostications into coherent cognitive progressions that stakeholders can scrutinize and assess. Lexical frequency visualization techniques may be employed to represent these sequential reasoning pathways across numerous forecasts by consolidating and exhibiting the most recurrent lexical units from the model's analytical stages, with dimensional prominence indicating higher occurrence rates. This methodology assists in identifying recurring patterns, conceptual frameworks, or thematic elements upon which the model relies when formulating predictions, thus providing a comprehensive overview of its analytical behavior.

Figure 2. The most frequent words in chain of thought represented by word-cloud

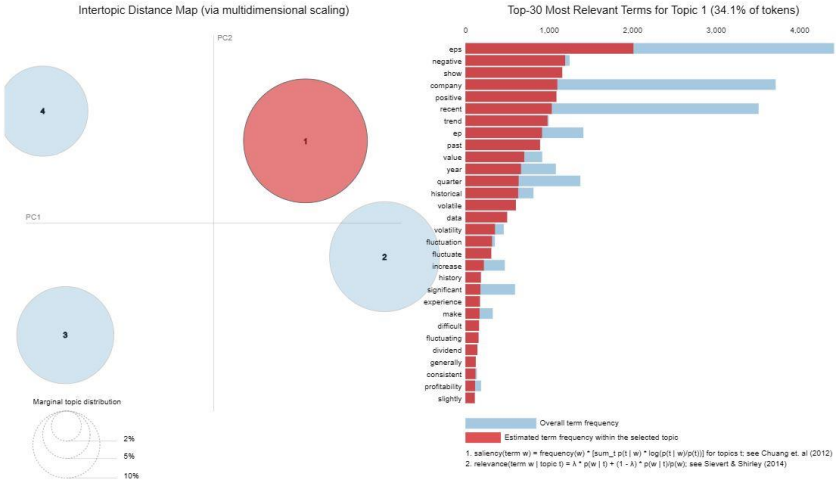


Source: own calculations

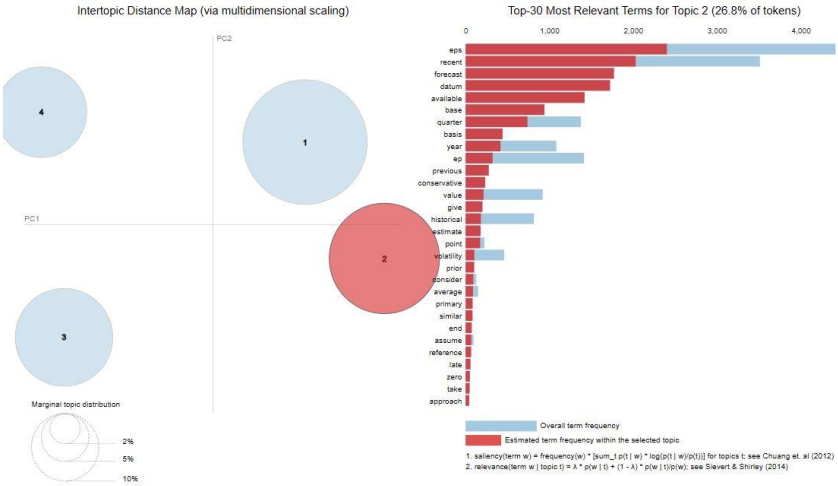
This word cloud highlights key financial forecasting terminology with “company“, “forecast“, “EPS“, “recent“, and “data“, appearing most prominently, suggesting a focus on corporate earnings predictions using recent data. The presence of terms like “volatility“, “fluctuations“, “negative“, “positive“, and “trend“ indicates analytical concerns about instability of earnings and potential swings in company performance. The inclusion of time-related words such as “historical,“ “quarter,“ “past,“ and “future“ demonstrates the temporal framework essential to financial analysis.

Figure 3. Topic model visualization in chain of thought

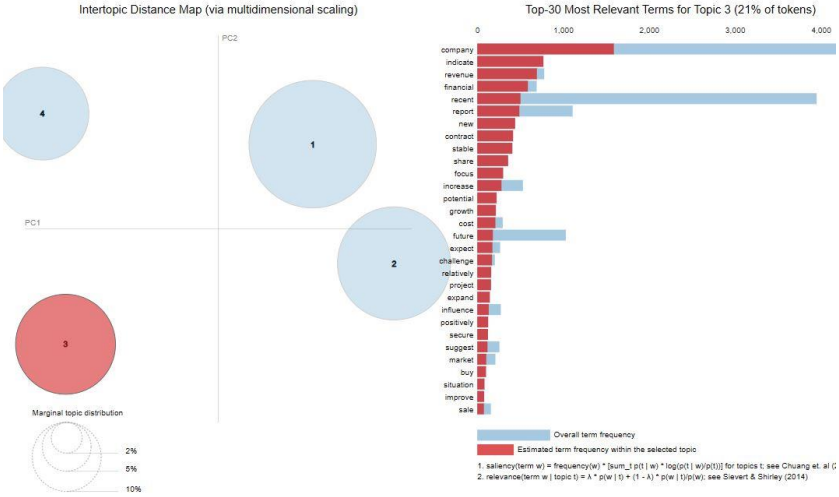
Topic 1



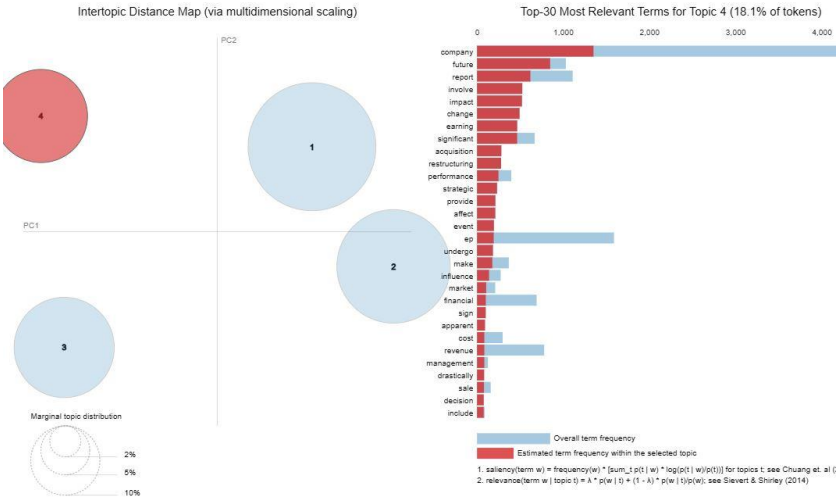
Topic 2



Topic 3



Topic 4



Source: own calculations

This graphical representation is called a Topic Model Visualization, specifically combining an Intertopic Distance Map (left side) with a Term Relevance Bar Chart (right side). The left visualization uses multidimensional scaling to display topic clusters in a two-dimensional space coming from Principal Component Analysis, showing their relative relationships, while the right side displays the most salient or relevant words for the selected topic (e.g. Topic 1) with their frequency distributions. This type of visualization is commonly used in topic modeling approaches like LDA (Latent Dirichlet Allocation) to help interpret and validate

discovered topics within a corpus. Latent Dirichlet Allocation (LDA) is a generative probabilistic model that assumes each document is a mixture of topics and each topic is a distribution over words. In topic modeling, LDA is used to uncover hidden thematic structures in large text corpora by identifying groups of words that frequently occur together, thereby revealing the underlying topics within the data.

Topic 1 appears to focus on EPS volatility and market trends. The most relevant terms include "eps", "negative", "show", "company", "positive", "recent", and "trend" suggesting this topic relates to the directional movement and fluctuations in earnings reports. This topic represents the foundational analysis of historical EPS patterns and volatility. Topic 2 centers on EPS forecast methodology and the data inputs used, containing terms like "forecast", "datum", "available", "basis", "previous", "conservative", "estimate", "reliability", and "quarter". Topic 3 relates to company fundamentals and financial outlook, mainly based on its revenues. Prominent terms include "company", "indicate", "revenue", "financial", "recent," "report," and "contract". This topic represents the contextual business factors that influence EPS expectations. Topic 4 focuses on corporate events and strategic changes. Key terms include "company," "future," "report," "involve," "impact," "change," "earnings," "acquisition," and "restructuring," suggesting this topic covers significant corporate activities that can materially affect EPS outcomes. Summarizing, the chain of thought for EPS forecasts thus progresses from: (1) analyzing historical EPS patterns and volatility, to (2) applying forecasting methodologies using available data, to (3) contextualizing within broader company financial fundamentals, to (4) adjusting for potential corporate events and strategic changes that might impact future earnings.

Discussion

This empirical investigation reveals that the seasonal random walk (SRW) model provides superior accuracy in representing Earnings per Share (EPS) patterns among Polish public companies. According to extensive research conducted by Kurylek [2023a, 2024a, 2024b, 2024c, 2025], this approach demonstrated exceptional performance compared to alternative models examined. The large language model (LLM) Gemini forecasts underperform relative to SRW when evaluated using the Mean Arctangent Absolute Percentage Error (MAAPE) metric. These conclusions remain consistent across multiple temporal ranges during robustness verification.

Interestingly, when alternative evaluation metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are employed, the Gemini model exhibits statistically significant superior performance compared to SRW. This discrepancy across different error metrics can be attributed to their distinct mathematical characteristics. The MAAPE metric likely imposes less severe penalties on outliers compared to RMSE and MAE, suggesting that SRW performs better in relative/percentage terms - particularly for companies with small EPS values where percentage errors can be very large - while Gemini produces lower absolute errors overall (MAE, RMSE), indicating generally accurate predictions

with occasional significant outliers that receive substantial penalties under the arctangent metric. This statistical divergence emphasizes how model selection should be determined by specific error tolerance requirements of financial applications - SRW being preferable when avoiding extreme outliers is crucial, while Gemini offers advantages when average prediction accuracy across the entire dataset is prioritized.

The examined period (2010-2019) represents a relatively stable market environment characterized by minimal disruptions and a horizontal WIG index trend. Under these conditions, SRW's fundamental assumption - that the upcoming quarter will closely mirror the corresponding quarter from the previous year - becomes increasingly reliable. Consequently, sophisticated models like GEMINI may offer minimal advantages in identifying stable patterns. Furthermore, the structural characteristics of emerging markets such as Poland may inherently favor less complex forecasting approaches. These outcomes suggest that simpler models may capture the maximum achievable predictive accuracy from historical data and effectively reflect the relatively straightforward EPS dynamics of Polish listed companies. The superior performance of less sophisticated models in the Polish context may be attributed to the simplicity of the underlying stochastic processes, consistent with the relatively unsophisticated nature of Polish public enterprises.

The observed inferior performance of GEMINI aligns with Cao and Wang [2024], who determined that LLMs' predictive accuracy significantly diminishes when confronted with diverse time series data and traditional signals containing both periodic and trend components, which EPS time series certainly exhibit. Additionally, these findings partially corroborate Abdelsamie and Wang's [2024] observation that LLMs continue to demonstrate limitations in nuanced contextual understanding and adaptability, highlighting the persistent value of human expertise.

Gemini LLM's forecast certainty score provides valuable insight into its confidence regarding EPS predictions, with scores predominantly clustering around 0.6-0.7, indicating moderate to high confidence despite variable accuracy. This metric functions as a practical instrument for investment decision-making, potentially enabling users to prioritize predictions with higher certainty scores while treating lower-confidence forecasts with appropriate caution. The chain-of-thought reasoning demonstrates that Gemini's forecasting methodology follows a logical progression from analyzing historical EPS patterns and volatility to implementing forecasting methodologies, contextualizing within broader company fundamentals, and ultimately adjusting for potential corporate events. Word cloud analysis of these reasoning chains emphasizes key financial terminology centered around volatility and trend indicators, demonstrating the model's focus on temporal frameworks essential to financial analysis. Topic modeling further illuminates four distinct dimensions of Gemini's analytical approach: historical EPS volatility assessment, forecast methodology application, company fundamental analysis, and consideration of strategic corporate events potentially impacting future earnings. This multifaceted reasoning structure indicates that despite accuracy limitations, Gemini employs a

sophisticated analytical framework mirroring the comprehensive evaluation process utilized by human financial analysts.

Given that EPS behavior follows a seasonal random walk pattern over shorter time series, and considering that stock prices derive from EPS multiplied by the price-to-earnings (P/E) multiple, one may infer that stock prices exhibit at minimum equivalent randomness to EPS. Accurately forecasting stock prices even one quarter in advance becomes exceptionally challenging when EPS behavior demonstrates random walk characteristics. During shorter timeframes where EPS remains constant, stock price forecasting essentially becomes P/E multiple prediction. Consequently, forecasting P/E multiples for periods shorter than a quarter - intervals between quarterly financial reports - could be particularly relevant for investment decisions. The seasonal random walk forecast effectively replicates values from identical quarters in previous years, suggesting that for extended forecast horizons, the P/E multiple may exert greater influence than forthcoming company earnings (EPS) when predicting future prices. This corresponds with economic theory positing that P/E multiples are influenced by anticipated future earnings growth, prospective interest rates, and market sentiment or risk premium, whereas EPS forecasts relate exclusively to imminent earnings. In both short-term and long-term investment contexts, consensus indicates that P/E multiples possess greater predictive significance than EPS forecasts.

CONCLUSIONS

The prediction of Earnings per Share (EPS) holds particular relevance within emerging financial markets such as Poland, where analyst coverage of publicly traded entities remains sparse. This investigation examines the performance capabilities of Large Language Models (LLM), specifically the Gemini architecture, in prognosticating EPS temporal sequences. An examination of quarterly EPS information from 267 Polish corporate entities spanning 2010-2019 demonstrates that the Seasonal Random Walk (SRW) methodology yielded superior results with minimal error rates as quantified by the Mean Arctangent Absolute Percentage Error. Such findings potentially reflect the relatively rudimentary organizational structures prevalent among Polish listed corporations. Notably, alternative error assessment frameworks including Mean Absolute Error and Root Mean Square Error reveal contradictory patterns, potentially attributable to these metrics' heightened sensitivity to statistical outliers.

Confidence indicators generated by the Gemini model predominantly exhibit moderate to substantial certainty values (0.6-0.7), potentially offering investment decision guidance despite inconsistent accuracy performance. A chain-of-thought examination reveals sophisticated reasoning processes that emulate human financial analytical approaches – proceeding from historical pattern recognition through methodological implementation, fundamental analysis, and strategic event

consideration – indicating that LLMs provide analytical sophistication even when failing to surpass simpler statistical frameworks in accuracy measurements.

The functional implications of these findings suggest that within abbreviated temporal sequences, employing methodologies of greater complexity than SRW for EPS forecasting within Poland may be unwarranted. However, dependency on SRW for EPS modeling suggests that projected equity valuations may exhibit considerable stochasticity, complicating accurate predictions. Consequently, forecasting price-to-earnings multiples may prove more relevant than EPS predictions for future equity valuation estimations, particularly within abbreviated investment timeframes where EPS demonstrates relative stability.

Subsequent scholarly inquiry might explore correlations between forecasting accuracy and organizational magnitude, with sectoral analysis potentially informing optimal EPS prediction model selection. Investigation of time series transformations to normalize EPS distributions could yield significant insights. Moreover, assessing the performance of diverse prediction methods and analyst projections during recessionary periods, such as the 2008-2009 financial crisis or COVID-19 pandemic, may yield meaningful findings in further research. Detecting cyclical patterns with the SRW approach may provide insights for investment strategies, potentially questioning the assumptions of the 'weak form' of the Efficient Market Hypothesis (EMH). Future studies could evaluate whether such strategies are capable of outperforming the market.

ABBREVIATIONS

In the text the following abbreviations are used:
SRW – Seasonal Random Walk model
GEMINI – the Gemini model by Google DeepMind
MAAPE – Mean Arctangent Absolute Percentage Error
MAE – Mean Absolute Error
RMSE – Root Mean Square Error

STATEMENTS AND DECLARATIONS

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.
Funding: The authors have received no funding from any source in the preparation of this work.
Data availability statement: Data will be available on request.

REFERENCES

Abarbanell J., Bushee B. (1997) Fundamental Analysis, Future EPS, and Stock Prices. Journal of Accounting Research, 35(1), 1-24. https://doi.org/10.2307/2491464

Abdelsamie M., Wang H. (2024) Comparative Analysis of LLM-based Market Prediction and Human Expertise with Sentiment Analysis and Machine Learning Integration. 2024 7th International Conference on Data Science and Information Technology (DSIT), 1-6. https://doi.org/10.1109/dsit61374.2024.10881868

Abe Y., Matsuo S., Kondo R., Hisano R. (2024) Leveraging Large Language Models for Institutional Portfolio Management: Persona-Based Ensembles. 2024 IEEE International Conference on Big Data (BigData), 4799-4808. https://doi.org/10.1109/bigdata62323.2024.10825362

Ahmadpour A., Etemadi H., Moshashaei S. (2015) Earnings Per Share Forecast using Extracted Rules from Trained Neural Network by Genetic Algorithm. Computational Economics, 46(1), 55-63. https://doi.org/10.1007/s10614-014-9455-6

Atiya A., Shaheen S., Talaat N. (1997) An Efficient Stock Market Forecasting Model using Neural Networks. IEEE International Conference on Neural Networks – Conference Proceedings. http://dx.doi.org/10.1109/ICNN.1997.614231

Ball R., Ghysels E. (2017) Automated Earnings Forecasts: Beat Analysts or Combine and Conquer? Management Science, 64(10), 4936–4952. https://doi.org/10.1287/mnsc.2017.2864

Ball R., Watts R. (1972) Some Time Series Properties of Accounting Income. The Journal of Finance, 27(3), 663-681. http://dx.doi.org/10.1111/j.1540-6261.1972.tb00991.x

Bathke Jr. A. W., Lorek K. S. (1984) The Relationship between Time-Series Models and the Security Market's Expectation of Quarterly Earnings. The Accounting Review, 59(2), 163-176.

Bradshaw M., Drake M., Myers J., Myers L. (2012) A Re-Examination of Analysts' Superiority over Time-Series Forecasts of Annual Earnings. Review of Accounting Studies, 17(4), 944-968. http://dx.doi.org/10.1007/s11142-012-9185-8

Brandon Ch., Jarrett J. E., Khumawala S. B. (1987) A Comparative Study of the Forecasting Accuracy of Holt-Winters and Economic Indicator Models of Earnings Per Share for Financial Decision Making. Managerial Finance, 13(2), 10-15. http://dx.doi.org/10.1108/eb013581

Brooks L. D., Buckmaster D. A. (1976) Further Evidence of the Time Series Properties of Accounting Income. The Journal of Finance, 31(5), 1359-1373. http://dx.doi.org/10.1111/j.1540-6261.1976.tb03218.x

Brown L. D., Griffin P. A., Hagerman R. L., Zmijewski M. E. (1987) Security Analyst Superiority Relative to Univariate Time-Series Models in Forecasting Quarterly Earnings. Journal of Accounting and Economics, 9(1), 61-87. http://dx.doi.org/10.1016/0165-4101(87)90017-6

Brown L. D., Rozeff M. S. (1979) Univariate Time-Series Models of Quarterly Accounting Earnings Per Share: A Proposed Model. Journal of Accounting Research, 17(1), 179-189. http://dx.doi.org/10.2307/2490312

Cao Q., Gan Q. (2009) Forecasting EPS of Chinese Listed Companies using a Neural Network with Genetic Algorithm. 15th Americas Conference on Information Systems 2009, AMCIS 2009, 2791-2981.

Cao Q., Parry M. (2009) Neural Network Earnings Per Share Forecasting Models: A Comparison of Backward Propagation and the Genetic Algorithm. Decision Support Systems, 47(1), 32-41. https://doi.org/10.1016/j.dss.2008.12.011

Cao Q., Schniederjans M. J., Zhang W. (2004) Neural Network Earnings Per Share Forecasting Models: A Comparative Analysis of Alternative Methods. Decision Sciences, 35(2), 205-237. https://doi.org/10.1111/j.00117315.2004.02674.x

Cao R., Wang Q. (2024) An Evaluation of Standard Statistical Models and LLMs on Time Series Forecasting. 2024 IEEE International Conference on Future Machine Learning and Data Science (FMLDS), 533-538. https://doi.org/10.1109/fmlds63805.2024.00098

Cao Y., Chen Z., Pei Q., Lee N., Subbalakshmi K. P., Ndiaye P. M. (2024) ECC Analyzer: Extracting Trading Signal from Earnings Conference Calls using Large Language Model for Stock Volatility Prediction. Proceedings of the 5th ACM International Conference on AI in Finance, 257-265. https://doi.org/10.1145/3677052.3698689

Chen Y., Chen S., Huang H., Sangaiah A. (2020) Applied Identification of Industry Data Science using an Advanced Multi-Componential Discretization Model. Symmetry, 12(10), 1-28. https://doi.org/10.3390/sym12101620

Chong C. Y., Kao H.-Y. (2025) Rationale-Driven Predictions for Stock Movements: A Multi-model Integration and Stack Generalization Approach. Technologies and Applications of Artificial Intelligence, 124-138. https://doi.org/10.1007/978-981-96-4589-3_9

Conroy R., Harris R. (1987) Consensus Forecasts of Corporate Earnings: Analysts' Forecasts and Time Series Methods. Management Science, 33(6), 725-738. http://dx.doi.org/10.1287/mnsc.33.6.725

Elend L., Kramer O., Lopatta K., Tideman S. (2020) Earnings Prediction with Deep Learning. German Conference on Artificial Intelligence (Künstliche Intelligenz) KI 2020: Advances in Artificial Intelligence, 267-274. http://dx.doi.org/10.1007/978-3-030-58285-2_22

Elton E. J., Gruber M. J. (1972) Earnings Estimates and the Accuracy of Expectational Data. Management Science, 18(8), B409-B424. http://dx.doi.org/10.1287/mnsc.18.8.B409

Foster G. (1977) Quarterly Accounting Data: Time-Series Properties and Predictive-Ability Results. The Accounting Review, 52(1), 1-21.

Gaio L., Gatsios R., Lima F., Piamenta Jr. T. (2021) Re-Examining Analyst Superiority in Forecasting Results of Publicly-Traded Brazilian Companies. Revista de Administracao Mackenzie, 22(1), eRAMF210164. https://doi.org/10.1590/1678-6971/eramf210164

Gomez A. N., Kaiser L., Jones L., Parmar N., Polosukhin I., Vaswani A., Shazeer N., Uszkoreit J. (2017) Attention Is All You Need. arXiv preprint arXiv:1706.03762.

Google DeepMind. (2023) Gemini: A Family of Highly Capable Multimodal Models. arXiv. https://arxiv.org/abs/2312.1180

Griffin P. (1977) The Time-Series Behavior of Quarterly Earnings: Preliminary Evidence. Journal of Accounting Research, 15(1), 71-83. http://dx.doi.org/10.2307/2490556

Gupta R., Khirbat G., Singh S. (2013) Optimal Neural Network Architecture for Stock Market Forecasting. Proceedings – 2013 International Conference on Communication Systems and Network Technologies, CSNT 2013, 557-561. https://doi.org/10.1109/csnt.2013.120

Harris R. D. F., Wang P. (2019) Model-Based Earnings Forecasts vs. Financial Analysts' Earnings Forecasts. British Accounting Review, 51(4), 424-437. https://doi.org/10.1016/j.bar.2018.10.002

Hou K., van Dijk M., Zhang Y. (2012) The Implied Cost of Capital: A New Approach. Journal of Accounting and Economics, 53(3), 504-526. https://doi.org/10.1016/j.jacceco.2011.12.001

Hou G., Lian Q. (2024) Benchmarking of Commercial Large Language Models: ChatGPT, Mistral, and Llama. ResearchSquare. https://doi.org/10.21203/rs.3.rs-4376810/v1

Jarrett J. E. (2008) Evaluating Methods for Forecasting Earnings Per Share. Managerial Finance, 16, 30-35. http://dx.doi.org/10.1108/eb013647

Johnson T. E., Schmitt T. G. (1974) Effectiveness of Earnings Per Share Forecasts. Financial Management, 3(2), 64-72. http://dx.doi.org/10.2307/3665292

Kim S., Kim H. (2016) A New Metric of Absolute Percentage Error for Intermittent Demand Forecasts. International Journal of Forecasting, 32(3), 669-679. http://dx.doi.org/10.1016/j.ijforecast.2015.12.003

Kong Y., Nie Y., Dong X., Mulvey J. M., Poor H. V., Wen Q., Zohren S. (2024) Large Language Models for Financial and Investment Management: Models, Opportunities, and Challenges. The Journal of Portfolio Management, 51(2), 211-231. https://doi.org/10.3905/jpm.2024.1.646

Kuryłek W. (2023a) The Modeling of Earnings Per Share of Polish Companies for the Post-Financial Crisis Period using Random Walk and ARIMA Models. Journal of Banking and Financial Economics, 1(19), 26-43. http://dx.doi.org/10.7172/2353-6845.jbfe.2023.1.2

Kuryłek W. (2023b) Can Exponential Smoothing Do Better than Seasonal Random Walk for Earnings Per Share Forecasting in Poland? Bank Credit, 54(6), 651-672.

Kuryłek W. (2024a) Can We Profit from BigTechs' Time Series Models in Predicting Earnings Per Share? Evidence from Poland. Data Science in Finance and Economics, 4(2), 218-235. http://dx.doi.org/10.3934/DSFE.2024008

Kuryłek W. (2024b) Artificial Neural Networks and Gradient-Boosting Decision Trees in Time Series Forecasting of Earnings Per Share in Poland. Eastern European Economics, 1-22. https://doi.org/10.1080/00128775.2024.2429137

Kuryłek W. (2024c) If Multilayer Perceptron Network May Help in Multivariate EPS Forecasting. Evidence from Poland. Quantitative Methods in Economics, 25(3), 107-123. https://doi.org/10.22630/mibe.2024.25.3.10

Kuryłek W. (2025) Are Natural Language Processing Methods Applicable to EPS Forecasting in Poland? Data Science in Finance and Economics, 5(1), 35-52. https://doi.org/10.3934/dsfe.2025003

Lacina M., Lee B., Xu R. (2011) An Evaluation of Financial Analysts and Naïve Methods in Forecasting Long-Term Earnings. [in:] K. D. Lawrence R. K. Klimberg (Eds.), Advances in Business and Management Forecasting (pp. 77-101) Bingley, UK: Emerald. http://dx.doi.org/10.1108/S1477-4070(2011)0000008009

Lai S., Li H. (2006) The Predictive Power of Quarterly Earnings Per Share based on Time Series and Artificial Intelligence model. Applied Financial Economics, 16(18), 1375-1388. http://dx.doi.org/10.1080/09603100600592752

Lev B., Thiagarajan S. (1993) Fundamental Information Analysis. Journal of Accounting Research, 31(2), 190–215. http://doi.org/10.2307/2491270

Lev B., Li S., Sougiannis T. (2010) The Usefulness of Accounting Estimates for Predicting Cash Flows and Earnings. Review of Accounting Studies, 15(4), 779-807. https://doi.org/10.1007/s11142-009-9107-6

Li K. K. (2011) How Well Do Investors Understand Loss Persistence? Review of Accounting Studies, 16(3), 630-667. https://doi.org/10.1007/s11142-011-9157-4

Li K. K., Mohanram P. (2014) Evaluating Cross-Sectional Forecasting Models for the Implied Cost of Capital. Review of Accounting Studies, 19(3), 1152-1185. https://doi.org/10.1007/s11142-014-9282-y

Lorek K. S. (1979) Predicting Annual Net Earnings with Quarterly Earnings Time-Series Models. Journal of Accounting Research, 17(1), 190-204. http://dx.doi.org/10.2307/2490313

Lorek K. S., Willinger G. L. (1996) A Multivariate Time-Series Model for Cash-Flow Data. Accounting Review, 71, 81-101.

Ohlson J. A. (1995) Earnings, Book Values, and Dividends in Equity Valuation. Contemporary Accounting Research, 11(2), 661-687. https://doi.org/10.1092/7tpj-rxqn-tqc7-ffae

Ohlson J. A. (2001) Earnings, Book Values, and Dividends in Equity Valuation: An Empirical Perspective. Contemporary Accounting Research, 18(1), 107-120. https://doi.org/10.1092/7tpj-rxqn-tqc7-ffae

Pagach D. P., Warr R. S. (2020) Analysts Versus Time-Series Forecasts of Quarterly Earnings: A Maintained Hypothesis Revisited. Advances in Accounting, 51, 1-15. http://dx.doi.org/10.1016/j.adiac.2020.100497

Pope P. F., Wang P. (2005) Earnings Components, Accounting Bias and Equity Valuation. Review of Accounting Studies, 10(4), 387-407. https://doi.org/10.1007/s11142-005-4207-4

Pope P., Wang P. (2014) On the Relevance of Earnings Components: Valuation and Forecasting Links. Review of Quantitative Finance and Accounting, 42, 399-413. https://doi.org/10.1007/s11156-013-0347-y

Ruland W. (1980) On the Choice of Simple Extrapolative Model Forecasts of Annual Earnings. Financial Management, 9(2), 30-37. http://dx.doi.org/10.2307/3665165

Sarker I. H. (2024) LLM Potentiality and Awareness: A Position Paper from the Perspective of Trustworthy and Responsible AI Modeling. https://doi.org/10.36227/techrxiv.170905626.67078570/v1

Watts R. L. (1975) The Time Series Behavior of Quarterly Earnings. Working paper, Department of Commerce, University of New Castle.

Wilcoxon F. (1945) Individual Comparisons by Ranking Methods. Biometrics, 1, 80-83. http://dx.doi.org/10.2307/3001968

Xiaoqiang W. (2022) Research on Enterprise Financial Performance Evaluation Method based on Data Mining. 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI) https://doi.org/10.1109/icetci55101.2022.9832404