

*Research article*

## **Can we profit from BigTechs' time series models in predicting earnings per share? Evidence from Poland**

**Wojciech Kuryłek\***

Faculty of Management, University of Warsaw, 1/3 Szturmowa Street, 02-678 Warsaw, Poland

\* **Correspondence:** Email: [wkurylek@wz.uw.edu.pl](mailto:wkurylek@wz.uw.edu.pl).

**Abstract:** Forecasting earnings for publicly traded companies is of paramount significance for investments, which is the background of this research. This holds particularly true in emerging markets where the coverage of these companies by financial analysts' predictions is limited. This research investigation delves into the prediction inaccuracies of cutting-edge time series forecasting algorithms created by major technology companies such as Facebook, LinkedIn, Amazon, and Google. These techniques are employed to analyze earnings per share data for publicly traded Polish companies during the period spanning from the financial crisis to the pandemic shock. My objective was to compare prediction errors of analyzed models, using scientifically defined error measures and a series of statistical tests. The seasonal random walk model demonstrated the lowest error of prediction, which might be attributable to the overfitting of complex models.

**Keywords:** earnings per share; financial forecasting; machine learning; deep learning; Prophet by Facebook; SilverKite by LinkedIn, DeepAR by Amazon; TFT by Google; Warsaw Stock Exchange

**JEL Codes:** C01, C02, C12, C14, C58, G17

---

**Abbreviations:** RW: Random walk; SRW: Seasonal random walk; MAAPE: Mean arctangent absolute percentage error

## 1. Introduction

The prediction of company stock prices relies on the multiplication of two key components: Earnings per share (EPS) and the Price-to-Earnings (P/E) multiple. Accurate forecasting of these elements is paramount for making sound investment decisions. EPS forecasts, in particular, play a critical role by providing valuable numerical insights into a company's financial performance. In advanced markets such as the United States, a significant proportion of companies benefit from extensive coverage by financial analysts. However, in emerging markets like Poland, this coverage extends to only about 20% of companies. This discrepancy underscores the necessity of employing statistical models for EPS prediction, a need that began to be addressed in the 1960s. The initial class of models that garnered researchers' attention were the autoregressive integrated moving average (ARIMA) models. The outcomes of these investigations yielded mixed results, with some studies suggesting the efficacy of the basic random walk model, indicating that more complex models did not consistently outperform it. Conversely, other studies arrived at divergent conclusions. Furthermore, from the late 1960s onward, researchers embarked on an extensive exploration of various approaches employing exponential smoothing techniques for EPS prediction. These investigations produced also mixed results, with some advocating for the use of exponential smoothing techniques, while others opposed them. However, over time, a consensus gradually emerged among researchers, pointing to the superiority of ARIMA-type models, which consistently delivered the most accurate forecasts. This consensus persisted until the late 1980s when a new prevailing belief emerged, suggesting that forecasts generated by financial analysts surpassed those produced by time series models. However, many researchers observed something opposite. More recently, the advent of machine learning and deep learning has ushered in new opportunities for experimentation. Artificial neural networks (ANNs) have been applied to various financial problems, including EPS forecasting.

I aim to compare the predictive capabilities of four cutting-edge algorithms developed by major technology companies, namely Facebook, LinkedIn, Amazon, and Google, with the traditional seasonal random walk model. These methods are applied in univariate time series settings to predict EPS. The analysis encompasses quarterly EPS data for 267 companies listed on the Polish stock exchange, covering the period from the 2008–2009 financial crisis to the 2020 pandemic shock. For forecast testing, the years 2017–2019 are selected, aligning with the sample used in Kuryłek's previous studies (2023a, 2023b).

In lieu of relying on the conventional mean absolute percentage error (MAPE) metric, which may yield extreme values when the denominator is small, we opt for the mean arctangent absolute percentage error (MAAPE) metric, as proposed by Kim and Kim (2016).

In summary, I pursue the following primary objective. It seeks to identify methods for forecasting EPS in situations where such predictions are scarce, particularly in emerging markets like Poland. Unfortunately, it occurred in the previous research that old-fashioned statistical methods were not useful in this respect. I also aim to assess whether recently developed state of the art time series models by BigTechs can effectively predict EPS and outperform a simplistic seasonal random walk model. Notably, this article presents a comparative analysis of these models using the same dataset, a unique contribution to the field. Hence, the paper extends the existing evidence that the naïve random walk model outperforms more advanced statistical models in out-of-sample predictions (Brandon et al., 1983; Gerakos and Gramacy 2013; Grigaliūnienė, 2013).

## 2. Literature review

The exploration of Earnings Per Share (EPS) forecasting has been a subject of extensive research in academic literature since the late 1960s, with a primary focus on US companies. Researchers have delved into various forecasting models, encompassing both simple random walk models and more intricate autoregressive integrated moving average (ARIMA) models (Ball and Watts, 1972; Watts, 1975; Griffin, 1977; Foster, 1977; Brown and Rozeff, 1977, 1979). The outcomes of these studies have produced mixed results, with some studies advocating for the basic random walk model, asserting that more complex models did not consistently outperform it, while others have drawn different conclusions. However, over time, a consensus gradually emerged among researchers, indicating that ARIMA-type models generally provided the most accurate forecasts, as evidenced by the works of Lorek (1979) and Bathke and Lorek (1984). Lorek and Willinger (1996) found that multivariate cross-sectional model outperforms firm-specific and common-structure ARIMA models. Later, Finger's (1994) study has examined the value relevance of earnings as measured by their ability to predict both earnings and cash flow. Earnings were found to be a significant predictor of themselves for most of the sample firms. Out-of-sample forecasts showed that random walk models outperformed individually estimated earnings models for one-year ahead forecasts. This occurred because the choice of an appropriate model was dependent on the business context as suggested by Lorek and Willinger (2007). According to them, the random walk-with-drift model provided more accurate forecasts for a sample of high-technology firms, and certain ARIMA-type models were more accurate for a sample of regulated companies and financial institutions.

Moreover, since the late 1960s and beyond, researchers have undertaken comprehensive explorations of various methods employing exponential smoothing for EPS forecasting. For instance, Elton and Gruber (1972) found that additive exponential smoothing without a trend component yielded among the best-performing models. Subsequently, several authors, including Ball and Watts (1972), Johnson and Schmitt (1974), Brooks and Buckmaster (1976), Ruland (1980), and Brandon et al. (1983), discovered that EPS time series tended to follow random walks, with exponential smoothing delivering similar results in terms of forecast accuracy. Brandon et al. (1986) further highlighted the effectiveness of the Holt-Winter (see Holt, 1957, 2004) exponential smoothing model for EPS prediction, particularly for short-term forecasts. This cost-effective model consistently produced accurate forecasts in comparison to other methods. These findings were corroborated by subsequent studies by Brandon et al. (1987) and Jarrett (2008), which reaffirmed the superior performance of the Holt-Winters model, as measured by the Mean Absolute Percentage Error (MAPE) metric.

Turning the focus to the Polish market, Kuryłek (2023a, 2023b) conducted analogous studies that compared various univariate time series models, including multiple naive random walk models, ARIMA-type models, and exponential smoothing models. These models were employed for EPS data related to Polish companies spanning from the aftermath of the 2008–2009 financial crisis to the onset of the 2020 pandemic shock. Notably, the seasonal random walk (SRW) model emerged as the standout performer across all quarters, providing a relatively accurate representation of the Polish market's behavior in contrast to the other models examined.

A consensus regarding ARIMA models' effectiveness persisted until the late 1980s when a prevailing belief surfaced, suggesting that forecasts generated by financial analysts surpassed those produced by time series models (Brown et al., 1987). Nevertheless, Conroy and Harris (1987) observed that analysts tended to excel in short forecast horizons, with their advantage diminishing over longer

timeframes. This perspective persisted until recent years when the superiority of analysts over time series models was once again questioned. Lacina et al. (2011) noted that analysts' forecasts were no more accurate than naive random walk (RW) forecasts. Lev et al. (2010) provided evidence that estimate-based accounting items are less useful for the prediction of cash flows, however, they improve the prediction of next year's earnings, though not of subsequent years' earnings. A notable study by Bradshaw et al. (2012) revisited the widely accepted notion that analysts' EPS forecasts outperformed random walk (RW) time-series forecasts. To their surprise, basic RW forecasts exhibited higher accuracy, particularly for longer time horizons, smaller or newer firms, and situations where analysts predicted negative or substantial changes in EPS. In a similar vein, research by Pagach and Warr (2020) validated that ARIMA time-series forecasts of quarterly EPS were as accurate as, or even more precise than, the consensus analysts' forecasts in approximately 40% of cases. Moreover, this time-series superiority became more pronounced with longer forecast horizons, decreased firm size, and particularly in the case of high-technology firms. Similarly, Gaio et al. (2021) suggested that the random walk model outperformed market analysts' forecasts in Brazil.

Recent research has placed a significant emphasis on the utilization of artificial neural networks for EPS forecasting. Applications of big data approach and artificial intelligence to construction and infrastructure problems can be found in the following articles: Aidan et al. (2020), Al-Somaydai et al. (2022), Al-Zwainy et al. (2016, 2018, 2020), Al-Zwainy and Raheem (2020). However, the outcomes of this research have been inconclusive. Cao et al. (2004) conducted a comparative analysis of forecasting accuracy using three-layer neural feedforward networks in both univariate and multivariate settings, employing the logistic (sigmoid) activation function. This study demonstrated that the application of neural networks led to more accurate forecasts compared to other forecasting models. In contrast to these findings, Lai and Li (2006) examined the performance of various models for EPS prediction, including ARIMA models and Artificial Neural Networks (ANN), and identified the ANN model as having the poorest accuracy. Cao and Parry (2009) showcased that the univariate neural network model consistently outperformed both univariate models and linear regression models. In a related study from the same year, Cao and Gan (2009) applied neural network models to predict the EPS of Chinese listed companies, confirming that the neural network model, with weights optimized using a genetic algorithm, outperformed a similar model using backpropagation for weight estimation, regardless of whether a univariate or multivariate approach was employed. Ahmadpour et al. (2015) delved into EPS forecasting utilizing a standard multilayer perceptron (MLP) neural network with three layers. Additionally, they used a genetic algorithm to extract rules from the neural network. Interestingly, these extracted rules exhibited significantly greater accuracy than a pure MLP model. In a more recent investigation by Elend et al. (2020), who compared long short-term memory (LSTM) networks to temporal convolution networks (TCNs) for predicting future EPS, focusing on a diverse sample of US firms. The results demonstrated that LSTM outperformed the naive persistent model, achieving a significant enhancement in prediction accuracy. TCNs also displayed promising results. Notably, the predictive accuracy of these neural networks was at least equivalent to, if not superior to, that of analysts, particularly for non-financial companies. The article by Xiaoqiang (2022) provides a concise overview of deep learning and machine learning techniques applicable to financial ratios forecasting, including EPS. However, the newest achievements in time series methods developed by BigTech companies haven't been explored in this context yet.

As the newest one, it is also worth mentioning the work by Dreher et al. (2024). The authors showed that considering accounting information on tax loss carryforwards does not enhance

performance forecasts and typically worsens the predictions of earnings in out-of-sample tests, for German listed companies, using the information for tax footnotes. Their article was based on the findings of Flagmeier (2017), who found a negative and significant association between unrecognized tax losses of German firms and future pre-tax cash flows and earnings.

### 3. Data and methods

#### 3.1. Data

The Polish stock market, which integrated into the European Union post-2004, distinguishes itself with its depth. It boasted a market capitalization of \$197 billion and had 774 listed companies as of the close of 2021. However, it is important to highlight that its stocks lack the extensive coverage by financial analysts seen in the United States or Western Europe. In 2019, only about 20% of the 711 listed companies received analyst coverage. This underscores the pressing need for employing statistical forecasting techniques to predict key financial data for these companies. I primarily focus on analyzing the earnings per share (EPS) data series, which is sourced from EquityRT, an analytical platform. The analysis explores the EPS figures of firms listed on the Warsaw Stock Exchange, spanning from Q1 2010 to Q4 2019. This period extends between two significant structural shifts: the first being the financial crisis of 2008–2009, and the second being the onset of the COVID-19 pandemic in 2020. For the purpose of forecasting, data from Q1 2010 to Q4 2018 (36 quarters) are used for model estimations. Furthermore, data from Q1 2019 to Q4 2019 are reserved as a validation sample to assess the accuracy of forecasts. Alternatively, sliding window approaches were explored by utilizing the years 2017 and 2018 as validation samples. After comprehensive coverage of a full-time window and the exclusion of the effects of splits and reverse splits by removing them from the sample, 267 companies remained in the dataset.

#### 3.2. The models

Denote  $Q_t$  as the realization of *EPS* in quarter  $t$ . Individually, time series models were estimated for each company.

The seasonal random walk model (SRW) can be described as:

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

This model is proposed because of the research by Kuryłek (2023a, 2023b), which proves its superiority over either ARIMA or exponential smoothing model types for Poland. The forecast is  $\hat{Q}_t = Q_{t-4}$ , so the model uses the value delayed by 4 quarters as the forecast, removing the necessity for parameter estimation.

BigTech companies such as Facebook, Amazon, Google, and LinkedIn have played a pivotal role in advancing the field of time series modeling. They have not only developed cutting-edge research but have also made their proprietary time series models available to the public, thereby influencing the industry landscape and promoting specific approaches and technologies.

##### 3.2.1. Additive approach

An additive model can be defined through the following decomposition of a time series:

$$Q_t = T_t + S_t + H_t + \varepsilon_t \quad (2)$$

where  $T_t$  is a trend,  $S_t$  is a seasonality component,  $H_t$  a holiday component and  $\varepsilon_t$  a random error. The trend component is responsible for encapsulating the overarching direction exhibited by the time series, while the seasonality component effectively captures the cyclic patterns inherent in the time series data. Additionally, the holiday component is dedicated to encompassing the influence of holidays on the time series. This approach finds application in the algorithms developed by both Facebook and LinkedIn. Here the length of time series used for training the models for each of 267 companies is 36 observations.

#### *The Prophet model (PROPH) by Facebook*

The Prophet model, created by Facebook employees Letham and Taylor (2018), is a versatile tool. It can capture both linear and non-linear trends, automatically detecting changepoints, accommodating shifts in the trend, handling multiple seasonal patterns, and effectively managing outliers. It employs piecewise linear or logistic trend functions. This model because of its additive structure is simple, interpretable, and flexible because it may handle a wide variety of time series. The model can be however sensitive to parameter tuning and it can be difficult to find the optimal parameter values for a given time series. The piecewise linear trend is a straightforward adaptation of the linear model, where different segments of the independent variable exhibit distinct linear relationships. This setting is used for modeling. On the other hand, the logistic trend handles non-linear growth patterns that saturate, indicating that the growth rate diminishes over time until it levels off. To model the seasonality component of time series data, the Prophet model utilizes a Fourier series, a mathematical representation of periodic signals that decomposes them into their constituent frequencies. For forecasting, the algorithm adopts a Bayesian approach, which considers the uncertainty in estimated parameters when making predictions. The fbprophet library in the Python programming language is used to implement this model.

#### *The SilverKite model (SILV) by LinkedIn*

The SilverKite model, a prominent algorithm within the GreyKite library, was detailed in a research paper authored by LinkedIn employees Al Orjany et al. (2022). This model's additive nature makes it a straightforward, easy-to-understand tool that can handle diverse time series patterns. However, its parameter sensitivity can make it challenging to achieve optimal performance for specific time series. The SilverKite is a relatively new model, and its performance compared to established methods is being evaluated. It operates by separately modeling the conditional mean of a time series and the volatility or uncertainty associated with the error term. In the initial phase, the model extracts raw features from timestamps, event data, and historical records. These features are then transformed into suitable basis functions, such as the Fourier transformation for capturing seasonality. A changepoints detection algorithm is applied to identify shifts in the trend, and an appropriate machine learning algorithm is trained to accommodate potential covariates, including techniques like ridge or quantile regression. In the second phase, the model focuses on modeling the conditional variance of residuals using an auto-regression process, which can effectively address any remaining correlations within the series. To perform these calculations, the greykite library in Python is employed.

### 3.2.2. Artificial neural network approach

The behavior of EPS can be described by the following formula:

$$Q_{t+1} = f_{model}(Q_t, Q_{t-1}, \dots, Q_{t-n}) + \varepsilon_t \quad (3)$$

where  $f_{model}()$  is a function described by an artificial recurrent neural network (RRNs) and  $\varepsilon_t$  is a residual term.

These concepts are outlined in the book authored by Bengio et al. (2017). An artificial neural network can be conceptualized as a black box capable of prediction but not explanation of modeled behavior. The parameter  $n$  denotes the length of the lookback period and is set at 8 to encompass at least one year and an equivalent period of full years. This choice aligns with the relevance of the seasonal random walk model, as demonstrated by Kurylek (2023a, 2023b), which is rooted in this one-year delay. Additionally, given the relatively short time series utilized for model training, it is essential to keep the number of delayed years in check so as not to over-restrict the volume of observations for model training. As a result, a duration of two years, equivalent to 8 quarters, was selected to explore dependencies longer than one year while maintaining sufficient length. Both Amazon and Google have adopted the artificial neural network approach. Number of observations used for training the models for each of 267 companies is 28.

#### *The Autoregressive Recurrent Network model (DEEPAR) by Amazon*

The model's details are outlined in the paper authored by Flunkert et al. (2020) and form a component of the Amazon SageMaker service. DeepAR is highly scalable, making it well-suited for handling large volumes of time series data, but it is a complex neural network architecture and as such it can make it almost impossible to understand and interpret compared to simpler forecasting models. Also due to its neural network nature training processes can be computationally expensive. DeepAR utilizes a recurrent neural network to parameterize a distribution function that captures the level of uncertainty in forecasts. Recurrent neural networks excel in identifying and learning nonlinear long-term dependencies within data and are adept at processing sequential data. In the case of a Gaussian distribution, the model predicts both the mean and standard deviation. The algorithm follows an autoregressive approach, where predictions are generated one step at a time. At each time step within the forecast horizon, the model conditions its prediction on values generated in prior steps. DeepAR also autonomously generates feature time series based on the frequency of the target time series. To deploy the model in Python, the gluonts library is utilized.

#### *The Temporal Fusion Transformer model (TFT) by Google*

The TFT model, crafted and published by Google researchers in a paper authored by Arik et al. (2021), centers around the transformer architecture and is tailored for probabilistic distribution forecasting. Transformers, a type of neural network, excel in modeling extended dependencies within data. They employ an encoder-decoder architecture, where the encoder handles historical time series data, and the decoder generates forecasts for upcoming time steps. Both the encoder and decoder employ a self-attention mechanism to grasp enduring relationships across various time steps. This mechanism allows the model to selectively focus on different segments of the input sequence, capturing dependencies over time. The TFT architecture incorporates Gated Residual Networks (GRNs) at different levels. These GRNs introduce residual connections that feed the output of a specific layer

to upper layers in the network that aren't directly adjacent. The model's automated feature detection mechanism automatically identifies and incorporates relevant features of any time series into the forecasting process, alleviating the need for manual feature engineering. Its neural network architecture makes the model hard to interpret. As a relatively nascent model, TFT's performance is still under evaluation against established forecasting methods. The model's implementation in Python relies on the `pytorch_forecasting` library.

### 3.3. Mean arctangent absolute percentage error (MAAPE)

It can be denoted  $A_1^i, \dots, A_4^i$  as the earnings per share (EPS) values for the first through fourth quarters of 2019 for a specific firm  $i$ . Similarly,  $F_1^i, \dots, F_4^i$  represent the predicted values of EPS for the corresponding quarters, denoted as  $\hat{Q}_t$ , where  $t=37, \dots, 40$  for  $i$ -th company. To calculate the absolute percentage error (APE) for any firm  $i$  during the  $j$ -th quarter of 2019, the following formula can be used:

$$APE_j^i = \left| \frac{A_j^i - F_j^i}{A_j^i} \right| \quad (4)$$

Nevertheless, the absolute percentage error (APE) has a notable limitation: It can produce infinite or undefined results when the actual values are close to zero, which is a common occurrence in earnings forecasts. Additionally, when the actual values are extremely small, often below one, it can result in exceptionally high percentage errors, essentially outliers. Furthermore, when actual values are zero, it leads to infinite APEs. To tackle this challenge, Kim and Kim (2016) introduced a novel method in the literature known as the arctangent absolute percentage error.

$$AAPE_j^i = \arctan \left( \left| \frac{A_j^i - F_j^i}{A_j^i} \right| \right) \quad (5)$$

This is due to the property of the arctan function, which transforms values in the range of  $[-\infty, +\infty]$  to a value within the interval  $[-\pi/2, \pi/2]$ . Consequently, the mean arctangent absolute percentage error (MAAPE) for the  $i$ -th firm can be formulated as follows:

$$MAAPE^i = \frac{1}{4} \sum_{j=1}^4 AAPE_j^i = \frac{1}{4} \sum_{j=1}^4 \arctan \left( \left| \frac{A_j^i - F_j^i}{A_j^i} \right| \right) \quad (6)$$

Moreover, the Mean Arctangent Absolute Percentage Error (MAAPE) for the  $j$ -th quarter across all  $I$  companies in the dataset can be formulated as:

$$MAAPE_j = \frac{1}{I} \sum_{i=1}^I AAPE_j^i = \frac{1}{I} \sum_{i=1}^I \arctan \left( \left| \frac{A_j^i - F_j^i}{A_j^i} \right| \right) \quad (7)$$

Thus, the following formula concisely summarizes the Mean Arctangent Absolute Percentage Error (MAAPE) across all four quarters and for all  $I$  companies in the sample:

$$MAAPE = \frac{1}{I} \sum_{i=1}^I MAAPE^i = \frac{1}{4} \sum_{j=1}^4 MAAPE_j \quad (8)$$

Predictions are generated using the mentioned models, and for each model represented as  $m$ , and the values  $MAAPE(m)_1, \dots, MAAPE(m)_4$ , along with  $MAAPE(m)$ , are subsequently computed.



### 3.4. The equality of means tests

To evaluate the statistical significance of differences in mean arctangent absolute percentage errors (MAAPEs) across multiple models, three statistical tests have been utilized: the one-way ANOVA test, the Alexander-Govern test, and the Kruskal-Wallis test. Descriptions of these tests are presented below.

#### *The one-way ANOVA test*

The one-way ANOVA test, as outlined by Lowry (2014), is employed to ascertain whether there is a statistically significant difference in the mean of errors, denoted as MAAPEs. This test is frequently used to examine whether means are equal. However, for the resulting p-value to be meaningful, specific critical assumptions must be satisfied. These assumptions encompass the independence of variables that yield observations in the sample, the consistency of variances among distinct groups, and the normal distribution of their datasets. The sample for all tests mentioned below consists of 267 observations.

$$H_0: \text{Means of AAPEs of all 4 models are the same} \quad (9)$$

#### *The Alexander-Govern test*

Unlike the one-way ANOVA test, this specific test does not require the assumption of homoscedasticity. Instead, it relaxes the requirement of equal variances, as pointed out by Alexander and Govern (1994). The other assumptions, such as the normality of distribution, remain relevant.

$$H_0: \text{Means of AAPEs of all 4 models are the same} \quad (10)$$

#### *The Kruskal-Wallis test*

Next, a Kruskal-Wallis one-way H-test is conducted, as detailed by Corder and Foreman (2009). This nonparametric test avoids the complexities associated with the potential normality of errors. The close proximity of the average ranks of the four models indicates that I cannot reject the null hypothesis of median AAPE equality. These computations for each quarter and for all forecast quarters are performed, resulting in Kruskal-Wallis H statistics and their respective p-values.

$$H_0: \text{Medians of AAPEs of all 4 models are the same} \quad (11)$$

It is noteworthy that in prior research, both the Alexander-Govern test and the Kruskal-Wallis test were not considered.

#### *The Wilcoxon test*

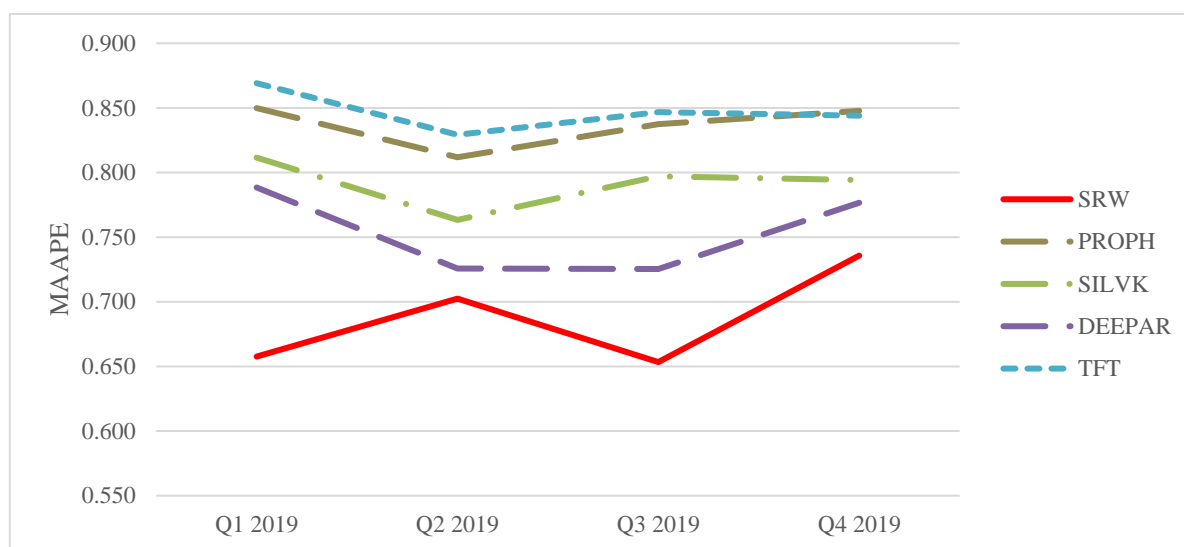
Last, a nonparametric two-sided Wilcoxon test is employed, as originally proposed by Wilcoxon (1945). This test facilitates a paired comparison of forecast errors, allowing us to assess the similarity in median errors across various models. Notably, this test does not rely on specific assumptions about probability distributions, except for the symmetry of the difference in scores and the independence of the random variables generating observations. Ruland (1980) provided a comprehensive explanation of the Wilcoxon test's utility in the context of verification, particularly for determining whether errors from different EPS models exhibit statistically significant differences.

$$H_0: \text{Medians of AAPEs of a pair of models are the same} \quad (12)$$

The null hypothesis for each test was rejected only if their respective p-values fell below the commonly accepted significance level of 0.05, a principle widely employed, as explained by Ruland (1980), among others. If the tests indicated that the mean (median) errors of a specific type of model were not only lower but also statistically discernible from those of other models, it would suggest the superiority of that particular model class over its alternatives. The calculation of the aforementioned test statistics and their corresponding p-values was performed using the Scipy library within Python.

#### 4. Results

The seasonal random walk (SRW) model, as detailed in Table 1, consistently demonstrates superior performance over all other models in every quarter, excelling in overall performance. In contrast, the Amazon-developed DeepAR model (DEEPAR) exhibits the second-best performance. The remaining three BigTech models (PROPH, SILVK, TFT) consistently occupy lower positions across all periods. Among these, LinkedIn's SilverKite model (SILVK) delivers the most favorable results, with Facebook's Prophet model (PROPH) following closely behind. The Google-developed TFT model (TFT) records the highest MAAPE.



**Figure 1.** Mean arctangent absolute error of various models.

Table 1 and Figure 1, similar to the graphical presentation made by Dreher et al. (2024), present the outcome of various models as well as the results of several equality of means tests, including the one-way ANOVA test (F statistic), the Alexander-Govern test (AG statistic), and the Kruskal-Wallis test (H statistics). Across each of the respective 1st, 2nd, 3rd, and 4th quarters, as well as when considering all quarters collectively, these tests consistently reject the null hypothesis claiming that the means (medians, in the case of the Kruskal-Wallis test) of arctangent absolute percentage errors (AAPEs) for all four models are statistically equivalent.

**Table 1.** Summary statistics on forecast errors and mean equality tests for 2019 quarters.

Model	Quarters				All Quarters MAAPE
	Q1 MAAPE	Q2 MAAPE	Q3 MAAPE	Q4 MAAPE	
SRW	0,658	0,702	0,653	0,736	0,687
PROPH	0,850	0,812	0,837	0,848	0,837
SILVK	0,812	0,763	0,797	0,794	0,792
DEEPAR	0,788	0,726	0,725	0,777	0,754
TFT	0,869	0,829	0,847	0,844	0,847
F statistics	8,067	3,488	7,760	2,990	9,212
F pvalue	0,000	0,008	0,000	0,018	0,000
H statistics	31,401	13,390	28,892	12,146	26,983
H pvalue	0,000	0,010	0,000	0,016	0,000
AG statistics	33,414	13,770	31,694	11,515	38,879
AG pvalue	0,000	0,008	0,000	0,021	0,000

To evaluate whether the errors of the best model exhibit statistical distinctions from those of the other models, the Wilcoxon nonparametric test was employed for comparing the medians of AAPEs between the SRW model and all other models. As outlined in Table 2, the results indicate that the seasonal random walk (SRW) model yields statistically lower median errors compared to the other models, except the DeepAR model (DEEPAR) only for the 2nd and 4th quarter.

**Table 2.** P-values of paired Wilcoxon test of forecast errors in respective quarters of 2019.

Quarter	Model	PROPH	SILVK	DEEPAR	TFT
Q1	SRW	0,000	0,000	0,000	0,000
Q2	SRW	0,000	0,018	0,452	0,000
Q3	SRW	0,000	0,000	0,039	0,000
Q4	SRW	0,002	0,022	0,125	0,000
ALL	SRW	0,000	0,000	0,001	0,000

This finding leads to the conclusion that the seasonal random walk model (SRW) offers the best performance among all the tested models in every analyzed quarter. The evidence presented supports the supremacy of the seasonal random walk model (SRW) for EPS forecasting in Poland.

#### 4.1. Robustness checks

Tests for robustness were performed, taking into account both time and various widely used error metrics. As shown in Table 3, the seasonal random walk model (SRW) had the smallest MAAPE error metric and achieved the best results in 2019, 2018, and 2017. The statistical significance of differences among the various models is confirmed by the remarkably low p-values obtained from all statistical tests conducted, including the one-way ANOVA test, the Alexander-Govern test, and the Kruskal-Wallis test. Additionally, the Wilcoxon test was applied to all model pairs in conjunction with the seasonal random walk model, and the corresponding p-values for each year are outlined in Table 4. In each of these years, the seasonal random walk model (SRW) achieved statistically superior results compared to other methods. Therefore, the consistent superiority of the seasonal random walk model is evident over time.

**Table 3.** Summary statistics on forecast errors and mean equality tests for all quarters 2017–2019.

Model	2017 MAAPE	2018 MAAPE	2019 MAAPE
SRW	0,686	0,711	0,687
PROPH	0,797	0,832	0,837
SILVK	0,791	0,815	0,792
DEEPAR	0,765	0,793	0,754
TFT	0,836	0,870	0,847
F statistics	6,569	7,030	9,212
F pvalue	0,000	0,000	0,000
AG statistics	22,978	24,079	26,983
AG pvalue	0,000	0,000	0,000
H statistics	28,352	29,901	38,879
H pvalue	0,000	0,000	0,000

**Table 4.** P-values of paired Wilcoxon test of forecast errors for all quarters 2017–2019 and SRW model.

Year	Model	PROPH	SILVK	DEEPAR	TFT
2017	SRW	0,000	0,000	0,000	0,000
2018	SRW	0,000	0,000	0,000	0,000
2019	SRW	0,000	0,000	0,001	0,000

Table 5 presents my evaluation of the performance of the examined models using alternative error metrics, specifically the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE), across all quarters in 2019. These metrics were adjusted for CPI inflation to ensure comparability, with the aim of making future errors have the same present value in nominal terms as current errors. Once again, the seasonal random walk model consistently demonstrated the lowest errors, whether assessed through RMSE or MAE.

**Table 5.** RMSE and MAE in all joint quarters 2019.

	SRW	PROPH	SILVK	DEEPAR	TFT
RMSE	0,937	1,679	1,259	1,007	1,616
MAE	0,705	1,397	1,079	0,828	1,388

**Table 6.** Summary statistics for RMSE and MAE in all joint quarters 2019.

	F statistics	F pvalue	AG statistics	AG pvalue	H statistics	H pvalue
RMSE	1,300	0,273	3,196	0,362	7,536	0,057
MAE	1,379	0,248	4,275	0,233	9,265	0,026

However, most of the statistical tests presented in Table 6 did not confirm any statistically significant differences among the outcomes generated by all these techniques. This lack of distinction can be attributed to the statistical proximity of errors between the SRW and DeepAR models, as indicated in Table 7 and Table 8 by the Wilcoxon tests. This suggests that the forecasts produced by the seasonal random walk (SRW) model, whether in terms of RMSE or MAE, outperform those of the three sophisticated models (PROPH, SILVK, TFT), with only the DeepAR model by Amazon (DEEPAR) achieving a statistically similar range of errors.

**Table 7.** P-values of paired Wilcoxon test of forecast errors for RMSE in 2019.

RMSE				
Model	PROPH	SILVK	DEEPAR	TFT
SRW	0,000	0,002	0,373	0,000

**Table 8.** P-values of paired Wilcoxon test of forecast errors for MAE in 2019.

MAE				
Model	PROPH	SILVK	DEEPAR	TFT
SRW	0,000	0,000	0,207	0,000

## 5. Discussion

The relative underperformance of more complex models developed by BigTechs, such as PROPH, SILVK, DEEPAR, and TFT, compared to the SRW model, fit well with Dreher et al. (2024) who also found that complex machine learning approaches that optimize explanatory power within the sample do not perform well for out-of-sample prediction. The main reason for that is model overfitting, which results in unstable relations among variables that depend on the relevant test data set. Using this relation in making predictions makes sense only if the statistical relation is sufficiently valid (Lev et al. 2010). More complex models may overparameterize the market's straightforward behavior, leading to larger forecast errors. These findings are also consistent with the work of Kuryłek (2023a, 2023b), which showed that even much simpler models like ARIMA and exponential smoothing models, that were relevant for the US market, cannot beat the naïve seasonal random walk for Poland. Moreover, the SRW model consistently outperformed other models across time and different error metrics like RMSE or MAE. This suggests that utilizing techniques more sophisticated than the SRW for EPS forecasting in Poland may lack practical merit or a further calibration for out of development sample predictions could be needed. Interestingly, even in the United States, research results are not unanimous. Some research, like Cao et al. (2004), argues that advanced neural network methods are superior, while others, such as Lai and Li (2006), provide evidence supporting simpler random walk models.

In conjunction with the results obtained by Kuryłek (2023a, 2023b), this strengthens the argument that even the most advanced BigTech time series models cannot outperform the simplicity of the random walk model in EPS forecasting in univariate time series settings. As a result, the use of these advanced techniques beyond the ordinary seasonal random walk in Poland for EPS forecasting in investment contexts seems impractical.

## 6. Conclusions

I explore the predictive capabilities of five univariate time-series models for forecasting Earnings per Share (EPS): The seasonal random walk (SRW), the Prophet model by Facebook (PROPH) model, the SilverKite model by LinkedIn (SILVK), the DeepAR model by Amazon (DEEPAR), and the TFT model by Google (TFT). These models represent the state of the art in time-series forecasting techniques, developed by leading BigTech companies. Mechanical forecasting of EPS holds particular significance in emerging markets, where financial analysts' coverage of listed companies is limited. This is exemplified by the case of Poland. When applied to quarterly EPS data for 267 Polish companies spanning from 2010 to 2019, the SRW model consistently outperformed the other models,

providing a more accurate representation of the Polish market. This assertion is supported by rigorous statistical tests, including the one-way ANOVA, Alexander-Govern, Kruskal-Wallis, and Wilcoxon tests.

I align with Dreher et al.'s (2024) findings that complex machine learning models maximizing in-sample explanatory power often underperform in out-of-sample prediction. This is primarily due to overfitting, where unstable relationships emerge based on the specific sample used for training. Applying these relationships to new data only holds validity if the underlying statistical basis is robust (Lev et al., 2010). In this case, the intricate models likely overparameterize the underlying, simpler market behavior, resulting in inflated forecast errors.

Moreover, the SRW model consistently outperformed other models across time and different error metrics like RMSE or MAE. This suggests that utilizing techniques more sophisticated than the SRW for EPS forecasting in Poland might lack practical merit or a further calibration for out of development sample predictions could be needed.

Future research could explore the relationship between forecasting accuracy and firm size. Additionally, the industry sector in which a company operates may be a crucial factor in determining the most accurate model for EPS forecasting. Investigating time series transformations to normalize EPS distributions could also be valuable. Furthermore, multivariate models incorporating fundamental variables can be explored too. A compelling area of interest lies in comparing the prediction accuracy of the best mechanical statistical/machine learning/neural network-based model with forecasts provided by market analysts. It could be also interesting to investigate how various model predictions, as well as analysts' forecasts, behaved during crisis times like the great financial crisis of 2008–2009 or COVID-19. From an investment perspective, further research is warranted on P/E multiple prediction techniques. Additionally, there may be identifiable seasonal patterns uncovered by the SRW model, offering potential insights into investment strategies. Such strategies may challenge the “weak form” of the Efficient Market Hypothesis (EMH).

### **Use of AI tools declaration**

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### **Funding**

The author has received no funding from any source in the preparation of this work.

### **Conflict of interest**

The author declares no conflicts of interest in this paper.

### **References**

Ahmadpour A, Etemadi H, Moshashaei S, et al. (2015) Earnings Per Share Forecast Using Extracted Rules from Trained Neural Network by Genetic Algorithm. *Comput Econ* 46: 55–63. <https://doi.org/10.1007/s10614-014-9455-6>

- Aidan IA, Al-Jeznawi D, Al-Zwainy FM, et al. (2020) Predicting earned value indexes in residential complexes' construction projects using artificial neural network model. *Int J Intell Eng Syst* 13: 248–259. <http://dx.doi.org/10.22266/ijies2020.0831.22>
- Alexander RA, Govern DM (1994) A New and Simpler Approximation for ANOVA under Variance Heterogeneity. *J Educ Stat* 19: 91–101. <http://dx.doi.org/10.2307/1165140>
- Ahammad P, Al Orjany SE, Chen A, et al. (2022) Greykite: deploying flexible forecasting at scale at LinkedIn. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3007–3017. <https://doi.org/10.1145/3534678.3539165>
- Al-Somaydai JA, Al-Zwainy FM, Hammoody O, et al. (2022) Forecasting and determining of cost performance index of tunnels projects using artificial neural networks. *Int J Comput Civil Struct Eng* 18: 51–60. <https://doi.org/10.22337/2587-9618-2022-18-1-51-60>
- Al-Zwainy FM, Amer R, Khaleel T, et al. (2016) Reviewing of the simulation models in cost management of the construction projects. *Civil Eng J* 2: 607–622. <https://doi.org/10.28991/cej-2016-00000063>
- AL-Zwainy FM, Huda F, Ibrahim A, et al. (2018) Development of an Analytical Software for Cost Estimation for Highway Project. *Int J Appl Eng Res* 13: 6944–6951.
- Al-Zwainy FM, Jaber FK, Jasim NA, et al. (2020) Forecasting techniques in construction industry: earned value indicators and performance models. *Sci Rev Eng Environ Sci* 29: 234–243. <https://doi.org/10.22630/PNIKS.2020.29.2.20>
- Al-Zwainy FM, Raheem SH (2020) Innovation of Analytical Software for Financing Construction Projects: Infrastructure Projects in Iraq as a Case Study, In: *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 978: 012015. <https://doi.org/10.1088/1757-899X/978/1/012015>
- Arik SÖ, Lim B, Loeff N, et al. (2021) Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *Int J Forecast* 37: 1748–1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- Ball R, Watts R (1972) Some Time Series Properties of Accounting Income. *J Financ* 27: 663–681. <http://dx.doi.org/10.1111/j.1540-6261.1972.tb00991.x>
- Bathke Jr AW, Lorek KS (1984) The Relationship between Time-Series Models and the Security Market's Expectation of Quarterly Earnings. *Account Rev* 59: 163–176.
- Bengio Y, Courville A, Goodfellow I, et al. (2017) *Deep Learning*. Cambridge, Massachusetts: The MIT Press.
- Bradshaw M, Drake M, Myers J, et al. (2012) A re-examination of analysts' superiority over time-series forecasts of annual earnings. *Rev Account Stud* 17: 944–968. <http://dx.doi.org/10.1007/s11142-012-9185-8>
- Brandon Ch, Jarrett JE, Khumawala SB, et al. (1983) On the predictability of corporate earnings per share. *J Bus Financ Account* 10: 373–387.
- Brandon Ch, Jarrett JE, Khumawala SB, et al. (1986) Comparing forecast accuracy for exponential smoothing models of earnings-per-share data for financial decision making. *Decision Sci* 17: 186–194. <http://dx.doi.org/10.1111/j.1540-5915.1986.tb00220.x>
- Brandon Ch, Jarrett JE, Khumawala SB, et al. (1987) A Comparative Study of the Forecasting Accuracy of Holt-Winters and Economic Indicator Models of Earnings Per Share For Financial Decision Making. *Manag Financ* 13: 10–15. <http://dx.doi.org/10.1108/eb013581>

- Brooks LD, Buckmaster DA (1976) Further Evidence of The Time Series Properties Of Accounting Income. *J Financ* 31: 1359–1373. <http://dx.doi.org/10.1111/j.1540-6261.1976.tb03218.x>
- Brown LD, Griffin PA, Hagerman RL, et al. (1987) Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. *J Account Econ* 9: 61–87. [http://dx.doi.org/10.1016/0165-4101\(87\)90017-6](http://dx.doi.org/10.1016/0165-4101(87)90017-6)
- Brown LD, Rozeff MS (1979) Univariate Time-Series Models of Quarterly Accounting Earnings per Share: A Proposed Model. *J Account Res* 17: 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. *Decis Support Syst* 47: 32–41. <http://dx.doi.org/10.1016/j.dss.2008.12.011>
- Cao Q, Schniederjans MJ, Zhang W, et al. (2004) Neural network earnings per share forecasting models: A comparative analysis of alternative methods. *Decision Sci* 35: 205–237. <https://doi.org/10.1111/j.00117315.2004.02674.x>
- Conroy R, Harris T (1987) Consensus Forecasts of Corporate Earnings: Analysts' Forecasts and Time Series Methods. *Manag Sci* 33: 725–738. <http://dx.doi.org/10.1287/mnsc.33.6.725>
- Corder GW, Foreman DI (2009) Comparing More than Two Unrelated Samples: The Kruskal–Wallis H-Test. In: *Nonparametric Statistics for Non-Statisticians*, John Wiley & Sons, Hoboken, New Jersey, 99–121. <http://dx.doi.org/10.1002/9781118165881>
- Dreher S, Eichfelder S, Noth F, et al. (2024) Does IFRS information on tax loss carryforwards and negative performance improve predictions of earnings and cash flows? *J Bus Econ* 94: 1–39. <http://dx.doi.org/10.1007/s11573-023-01147-7>
- Elend L, Kramer O, Lopatta J, et al. (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](http://dx.doi.org/10.1007/978-3-030-58285-2_22)
- Elton EJ, Gruber MJ (1972) Earnings Estimates and the Accuracy of Expectational Data. *Manag Sci* 18: B409–B424. <http://dx.doi.org/10.1287/mnsc.18.8.B409>
- Finger CA (1994) The ability of earnings to predict future earnings and cash flow. *J Account Res* 32: 210–223. <http://dx.doi.org/10.2307/2491282>
- Flagmeier V (2022) The information content of deferred taxes under IFRS. *Eu Account Rev* 31: 495–518. <http://dx.doi.org/10.1080/09638180.2020.1826338>
- Foster G (1977) Quarterly Accounting Data: Time-Series Properties and Predictive-Ability Results. *Account Rev* 52: 1–21.
- Flunkert V, Gasthaus J, Januschowski T, et al. (2020) DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *Int J Forecast* 36: 1181–1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>
- Gaio L, Gatsios R, Lima F, et al. (2021) Re-examining analyst superiority in forecasting results of publicly-traded Brazilian companies. *Revista de Administracao Mackenzie* 22: eRAMF210164. <https://doi.org/10.1590/1678-6971/eramf210164>
- Gerakos J, Gramacy RB (2013) Regression-Based Earnings Forecasts. *Chicago Booth Research Paper*, 12–26. <https://doi.org/10.2139/ssrn.2112137>
- Griffin P (1977) The Time-Series Behavior of Quarterly Earnings: Preliminary Evidence. *J Account Res* 15: 71–83. <http://dx.doi.org/10.2307/2490556>



- Grigaliūnienė Ž (2013) Time-series models forecasting performance in the Baltic stock market. *Organ Market Emerg E* 4: 104–120.
- Holt Ch C (1957) Forecasting seasonals and trends by exponentially weighted moving averages. *Working Paper, Carnegie Institute of Technology*.
- Holt Ch C (2004) Forecasting seasonals and trends by exponentially weighted moving averages. *J Econ Soc Meas* 29: 123–125. <https://doi.org/10.1016/j.ijforecast.2003.09.015>
- Jarrett JE (2008) Evaluating Methods for Forecasting Earnings Per Share. *Manag Financ* 16: 30–35. <http://dx.doi.org/10.1108/eb013647>
- Johnson TE, Schmitt TG (1974) Effectiveness of Earnings Per Share Forecasts. *Financ Manag* 3: 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. *Int J Forecast* 32: 669–679. <http://dx.doi.org/10.1016/j.ijforecast.2015.12.003>
- Kuryłek W (2023a) The modeling of earnings per share of Polish companies for the post-financial crisis period using random walk and ARIMA models. *J Bank Financ Econ* 1: 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: 651–672.
- Lacina M, Lee B, Xu R, et al. (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*, Bingley, UK: Emerald, 77–101. [http://dx.doi.org/10.1108/S1477-4070\(2011\)0000008009](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. *Appl Financ Econ* 16: 1375–1388. <http://dx.doi.org/10.1080/09603100600592752>
- Letham B, Taylor SJ (2018) Forecasting at scale. *Am Stat* 72: 37–45. <https://doi.org/10.7287/peerj.preprints.3190v1>
- Lev B, Souginannis T (2010) The usefulness of accounting estimates for predicting cash flows and earnings. *Rev Account Stud* 15: 779–807. <http://dx.doi.org/10.1007/s11142-009-9107-6>
- Lorek KS (1979) Predicting Annual Net Earnings with Quarterly Earnings Time-Series Models. *J Account Res* 17: 190–204. <http://dx.doi.org/10.2307/2490313>
- Lorek KS, Willinger GL (1996) A multivariate time-series model for cash-flow data. *Account Rev* 71: 81–101.
- Lorek KS, Willinger GL (2007) The contextual nature of the predictive power of statistically-based quarterly earnings models. *Rev Q Financ Account* 28: 1–22. <http://dx.doi.org/10.1007/s11156-006-0001-z>
- Lowry R (2014) Concepts and Applications of Inferential Statistics, Chapter 14. Available from: <https://onlinebooks.lib.rary.upenn.edu/webbin/book/lookupid?key=olbp66608>.
- Pagach DP, Warr RS (2020) Analysts versus time-series forecasts of quarterly earnings: A maintained hypothesis revisited. *Adv Account* 51: 1–15. <http://dx.doi.org/10.1016/j.adiac.2020.100497>
- Ruland W (1980) On the Choice of Simple Extrapolative Model Forecasts of Annual Earnings. *Financ Manag* 9: 30–37. <http://dx.doi.org/10.2307/3665165>
- Watts RL (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 Bulletin* 1: 80–83.  
<http://dx.doi.org/10.2307/3001968>
- Xiaoqiang W (2022) Research on enterprise financial performance evaluation method based on data mining. In: *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*. <https://doi.org/10.1109/icetci55101.2022.9832404>



AIMS Press

© 2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)