Wojciech Kuryłek
University of Warsaw
e-mail: wkurylek@wz.uw.edu.pl
ORCID: 0000-0003-0692-3300

DOI: 10.15290/oes.2025.03.121.19

HOW XGBOOST MAY HELP IN MULTIVARIATE EPS FORECASTING FOR COMPANIES LISTED ON THE WARSAW STOCK EXCHANGE

Summary

Purpose | The present analysis evaluates the forecasting capabilities of seven distinct methodologies: seasonal random walk (SRW), Laurent framework (L), Lev and Thiagarajan approach (LT), Residual Income methodology (RI), Pope and Wang framework (PW), Earnings Persistence approach (EP), and Hou, van Dijk, and Zhang methodology (HDZ) in predicting earnings per share.

Research method | It uses the eXtreme Gradient Boosting (XGBoost) approach, which can handle non-stationary data. This approach is one of the most efficient to apply for tabular data. To examine accuracy of predictions various metrics are utilised and the statistical significance of difference between them is tested using the Wilcoxon and Brunner-Munzel tests.

Results | Both the seasonal random walk framework and the Pope Wang methodology, when implemented through XGBoost, exhibited minimal prediction errors and generated superior representations of the Polish market dynamics relative to alternative approaches.

Originality/value/implications/recommendations | Such comprehensive application of multiple models for earnings per share prediction within the Polish market utilising XGBoost methodology represents an unprecedented approach. The notable effectiveness of the relatively simple seasonal random walk framework potentially reflects the less complex characteristics of the Polish equity market. In contrast, the robust performance of the Pope Wang framework indicates the importance of specific financial indicators.

Keywords: earnings per share, random walk, multivariate approach, XGBoost, financial forecasting, Warsaw Stock Exchange

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

1. Introduction

Stock valuation procedures necessitate the multiplication of earnings per share (EPS) by Price-to-Earnings multiples, constituting a fundamental element in investment decision-making processes. The forecasting precision of these components remains critical, with EPS projections assuming exceptional importance. Such projections deliver quantitative indicators regarding corporate future performance, supplying crucial information concerning prospective market valuations and establishing benchmarks for auditing processes. Professional analytical coverage extends to numerous enterprises within US markets, whereas Polish markets experience coverage for merely 20% of the listed entities, as documented by the EquityRT financial platform. This analytical coverage disparity originates from the limited proportion of the Warsaw Stock Exchange-listed entities that generate interest among institutional investors capable of producing adequate trading volumes, thereby creating commission revenues sufficient for brokerage firms to maintain costly professional analytical departments. The Polish equity market, which achieved the European Union integration following 2004, demonstrates considerable depth characterised by substantial market capitalisation. Consequently, the implementation of statistical or machine learning methodologies for EPS forecasting becomes imperative.

The present investigation examines predictive accuracy across multiple multivariate models incorporating numerous explanatory variables. These models employ differentiated explanatory variable configurations, incorporating methodological insights from Li and Mohanram [2014], utilising the eXtreme Gradient Boosting (XGBoost) methodology as the estimation framework. The analysis incorporates quarterly EPS data spanning 267 entities listed on the Polish stock exchange, covering the period from the 2008–2009 financial crisis through the 2020 pandemic.

The conventional mean absolute percentage error (MAPE) metric demonstrates vulnerability to extreme values when denominators approach minimal values; therefore, the mean arctangent absolute percentage error (MAAPE), introduced by Kim and Kim [2016], serves as the computational basis and the analytical metric within this investigation.

The research encompasses multiple objectives. The primary objective evaluates whether multivariate models implementing XGBoost estimation methodology enhance EPS prediction capabilities and demonstrate superior performance relative to basic seasonal random walk models, representing an unexplored area within the existing literature. The secondary objective applies varied error metrics, multiple temporal frameworks, and comprehensive statistical testing procedures for experimental outcome validation. The tertiary objective involves the adaptation and implementation of relative performance error metrics addressing situations where actual profit values approach zero, utilising MAAPE as the error measurement framework. The final objective clarifies the practical ramifications of research findings for the investment strategy formulation within the Polish equity markets.

2. Literature review

Automated Earnings per Share prediction emerged during the 1960s, marking the beginning of academic investigation centered on the autoregressive integrated moving average (ARIMA) methodologies [Ball and Watts, 1972; Watts, 1975; Griffin, 1977; Foster, 1977; Brown, Rozeff, 1979], producing diverse findings among studies. The Polish market witnessed analogous research through Kuryłek [2023a], emphasizing these modelling approaches.

ARIMA-type models eventually gained recognition for delivering characteristically accurate predictions [Lorek, 1979; Bathke, Lorek, 1984], establishing widespread agreement. This unanimity persisted through the late 1980s, after which evidence suggested financial analysts' predictions exceeded time series model performance [Brown et al., 1987]. Brokerage house analysts' EPS estimates demonstrate capacity for rapid incorporation of emerging data from sources including labour disruptions, unexpected incidents, and corporate announcements. Conroy and Harris [1987] documented the analysts' predictive advantage during brief forecasting periods, which decreased across extended timeframes. This understanding remained dominant before recent investigations reopened examination of analyst performance relative to time series approaches [Lacina et al., 2011; Bradshaw et al., 2012; Pagach and Warr, 2020; Gaio et al., 2021].

Beginning in the late 1960s, exponential smoothing techniques attracted research attention for EPS prediction purposes. Elton and Gruber [1972] assessed forecast precision across nine automated models incorporating exponential smoothing procedures. Additional investigations by Ball and Watts [1972], Johnson and Schmitt [1974], Brooks and Buckmaster [1976], Ruland [1980], [1986, 1987], and Jarrett [2008] implemented exponential smoothing as a fundamental approach for EPS prediction, yielding diverse outcomes.

Multivariate cross-sectional models demonstrated enhanced performance compared to firm-specific and common-structure ARIMA approaches according to Lorek and Willinger [1996]. Research identified varying combinations of predictive factors. Lev and Thiagarajan [1993] distinguished 12 essential indicators derived from financial ratios, later applied by Abarbanell and Bushee [1997] in EPS prediction. Cao and Gan [2009], Cao and Parry [2009], Ahmadpour et al. [2015], and Ball and Ghysels [2017] incorporated comparable fundamental factors for multivariate EPS prediction through the neural network applications, confirming their utility.

The residual earnings framework emerged through Ohlson [1995, 2001], while theoretical structures linking earnings predictions with accounting metrics and equity valuations appeared in Pope and Wang [2005, 2014]. Li [2011] developed specialised methodology for loss-generating entities' earnings prediction, establishing practical applicability. Empirical validation for estimate-based accounting elements in the subsequent-year earnings prediction came from Lev and Sougiannis [2010]. The HDZ model, constructed by Hou et al. [2012] utilising 1963–2000 US-listed company data, demonstrated substantial R^2 values. The comparative analysis by Li and Mohanram [2014] examined the random walk approach, HDZ model, earnings persistence (EP) framework, and residual income (RI) methodology. The EP and RI frameworks generated superior prediction precision despite HDZ's elevated R^2, especially for smaller enterprises and entities without analyst attention. Harris and Wang [2019] determined Pope and Wang's [2005] framework exhibited reduced bias and enhanced precision overall.

Decision tree methodologies (CHAID, C5.0, QUEST, and C&RT) served in Delenz et al. [2013] research examining financial ratio influence on organisational outcomes, with CHAID and C5.0 achieving optimal predictive precision. Gramacy and Gerakos [2013] investigated the advanced prediction techniques encompassing classification trees, regression trees, and random forests, determining basic random walk approaches matched complex methodology performance. Middle Eastern and North African banking EPS prediction models developed by Elamir [2020] incorporated simple regression, regression tree variants, conditional inference tree, and cubist regression tree methods, identifying cubist regression tree superiority. Machine learning and deep learning model comparisons for multivariate EPS forecasting by Chen et al. [2020] revealed decision tree learning's superior precision. Emerging machine learning and deep learning capabilities facilitated novel experimental approaches. Xiaoqiang [2022] presented the condensed examination of machine learning methods suited to the financial ratio prediction, encompassing EPS. The application of eXtreme Gradient Boosting (XGBoost) methodology specifically for EPS prediction remains unexplored in the existing literature.

Accounting data regarding the tax loss carryforwards failed to improve EPS prediction accuracy and frequently compromised out-of-sample forecast quality for German listed entities, as demonstrated by Dreher et al. [2024].

3. Data

The earnings per share (EPS) data series and the additional financial variables obtained from EquityRT, a financial analysis platform, constitute the primary subject of the examination. The investigation into EPS trends encompasses companies listed on the Warsaw Stock Exchange throughout the complete timeframe spanning Q1 2010 through Q4 2019, a period marked by major structural transformations: the financial crisis of 2008–2009 and the emergence of the COVID-19 pandemic in 2020. Beginning in 2020, a phase of heightened global volatility took shape, at first sparked by the COVID-19 pandemic and later exacerbated by the energy crisis linked to the 2022 conflict in Ukraine. This shift underscored 2019 as the final year of comparative stability, providing a suitable baseline for model validation and calibration. Periods of stability are characterised by low noise levels and consistent relationships among variables, offering optimal conditions for model learning. Failure in such an environment implies that the model is unlikely to generalise effectively under volatile scenarios, where abrupt shocks, structural breaks, and nonlinear dynamics challenge assumptions based on historical continuity. A model that lacks strong performance during such a predictable period is less likely to yield reliable outcomes in disrupted markets, where projecting past trends forward is prone to substantial inaccuracies. Consequently, the selection of this stable timeframe was intentional. Additionally, consistency in sample selection was upheld across analyses to enable direct comparison with prior studies [Kuryłek 2023a, 2023b], strengthening both the coherence and reliability of findings. The model estimation process utilises quarterly data spanning the period from the first quarter of 2010 through the fourth quarter of 2018, encompassing 36 quarters in total. The four quarters of 2019 constitute the holdout sample designated for out-of-time validation procedures. Forecast horizons extend from one quarter to four quarters into the future. An additional validation exercise incorporates data from the calendar years 2017 and 2018 as alternative test samples. Following the application of comprehensive data coverage criteria and the exclusion of corporate actions involving stock splits and reverse stock splits, the final dataset retains 267 companies.

4. Explanatory variables

Redundancy in financial ratios, often observed in attempts to explain economic phenomena, necessitates methods to streamline and identify a precise, independent subset of essential financial ratios suitable for EPS modelling. The subsequent models were employed:

The seasonal random walk model (SRW)

can be described as:

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

The forecasting equation takes the form $\widehat{EPS}_t = EPS_{t-4}$, whereby the projected value corresponds to the observation from four quarters prior, thus obviating parameter estimation requirements. This specification functions as a benchmark model, with its adoption motivated by findings from Kuryłek [2023a, 2023b] demonstrating its outperformance relative to the time series methodologies in the Polish context. The referenced studies established the benchmark's dominance through comparative analysis against ARIMA-class models and multiple exponential smoothing specifications, with results confirming the seasonal random walk's superior predictive accuracy.

The Laurent model (L)

Laurent [1979] achieved condensation of 45 financial ratios into 10 factors, with these factors collectively accounting for approximately 90% of the observed variance. A rigorous selection methodology subsequently yielded ten distinct financial ratios representative of these factors. These ratios offer multifaceted insights into the organisational financial performance. The selection criteria encompassed high correlation with the corresponding factor, mutual independence, and the established prevalence in practical application. The final ratio compilation addressed diverse financial dimensions: Profit before interest and tax to total assets (R1), Long-term debt to total assets (R2), Revenue to working capital (R3), Revenue to fixed assets (R4), Revenue to shareholders' funds (R5), Revenue to inventory (R6), Revenue to debt (R7), Quick liquidity ratio (R8), Profit before interest and taxes to interest (R9), and Reserves to net income (R10). These ratios serve as explanatory variables for earnings per share. The resulting estimated equation takes the following form:

$$EPS_{t+4} = \alpha_0 + R1_t + R2_t + \ldots + R10_t + \varepsilon_t \quad (2)$$

The Lev and Thiagarajan model (LT)

Lev and Thiagarajan [1993] conducted an investigation that yielded 12 fundamental indicators through analysing multiple applied publications concerning financial ratio employment in characterising various dimensions of corporate performance. This methodology represents among the most commonly cited frameworks within academic literature, as evidenced by Ahmadpour et al. [2015] and Ball and Ghysels [2017]. Significant indicators demonstrating utility encompass Inventory (I), Accounts receivable (AR), Capital expenditure (CAPEX), Gross margin (GM), Sales and administrative expenses (SAE), Provision for doubtful receivables (PROV), Effective tax (ET), and Labor intensity of sales (LP). Nevertheless, several indicators from the original study remain unavailable within the Polish context. The Polish publicly traded entities do not report R&D expenditure figures or order backlog information. Additionally, the database contains no details regarding the selected inventory valuation method (FIFO versus LIFO) and fails to enable differentiation among auditor opinion types. The unavailability of R&D expenditure and order backlog information from the Polish entities constrains the LT framework's capacity for incorporating critical prospective measures of innovative capabilities and anticipated revenue generation, thereby possibly diminishing its effectiveness in earnings projection. The absence of inventory valuation methodology disclosure (FIFO compared to LIFO) eliminates the framework's capability to compensate for divergent profitability effects these approaches generate in inflationary or deflationary environments, potentially resulting in misrepresented inter-company earnings analyses. The unavailability of auditor opinion information prevents incorporation of indicators regarding the possible financial statement quality concerns, omitting an essential risk measure that Lev and Thiagarajan's original investigation determined significant for fundamental evaluation. Consequently, the formulation corresponding to Lev and Thiagarajan's approach, configured using quarterly data for generating annual forward projections, takes the following form:

$$\begin{aligned} E_{t+4} &= \alpha_0 + \alpha_1 I_t + \alpha_2 AR_t + \alpha_3 CAPEX_t + \alpha_4 GM_t + \alpha_5 SAE_t + \\ &+ \alpha_6 PROV_t + \alpha_7 ET_t + \alpha_8 LP_t + \varepsilon_t \end{aligned} \tag{3}$$

The Residual Income model (RI)

The theoretical model of residual earnings developed by Ohlson [1995, 2001] and Ohlson and Juettner-Nauroth [2005] rests upon the clean surplus assumption, whereby equity market value corresponds to the sum of book value and the net present value of the prospective abnormal returns. The concept of residual income, alternatively termed abnormal earnings, represents the differential between the realised earnings and the product of book value and capital cost, exhibiting characteristics of an autoregressive stochastic process. This framework establishes that prospective earnings depend on prevailing earnings levels, book value, dividend distributions, and capital cost, yielding the subsequent accounting variables: Delayed earnings (E), a Dummy variable designating entities with delayed negative earnings (NegE), a Polynomial expression combining the two preceding variables (NegE * E), Book value (BV), and Accruals (ACC). The model's configuration draws from the investigation by Li and Mohanram [2014]. Implementation with quarterly data transforms the equation into the subsequent formulation:

$$E_{t+4} = \alpha_0 + \alpha_1 E_t + \alpha_2 NegE_t + \alpha_3 NegE_t * E_t + \\ + \alpha_4 BV_t + \alpha_5 ACC_t + \varepsilon_t \quad (4)$$

The Pope and Wang model (PW)

The theoretical linkage connecting earnings forecasts with six observable accounting variables emerges from Pope and Wang [2005, 2014] through a framework grounded in linear valuation, no-arbitrage conditions, dividend irrelevance, and clean surplus accounting principles. These variables comprise Earnings (E), Book value (BV) with one- and two-year lags, Accruals (ACC), alongside non-accounting measures including Stock price (P) incorporating one- and two-year lags. Harris and Wang [2019] provided the mathematical specification rendering this model in equation format. The formulation enabling quarterly earnings prediction with a one-year forecast horizon takes the following configuration:

$$E_{t+4} = \alpha_1 E_t + \alpha_2 BV_t + \alpha_3 BV_{t-4} + \alpha_4 ACC_t + \\ + \alpha_5 P_t + \alpha_6 P_{t-4} + \varepsilon_t \quad (5)$$

The Earnings Persistence model (EP)

The earnings persistence model accommodating growth dynamics emerged from Li [2011], featuring three key explanatory components: lagged Earnings (E), the binary indicator marking negative earnings (NegE), and their multiplicative interaction (NegE*E), thereby enabling distinction between persistence characteristics of profits versus losses. Li and Mohanram [2014] present the mathematical formulation governing this model's equation structure.

$$E_{t+4} = \alpha_0 + \alpha_1 E_t + \alpha_2 NegE_t + \alpha_3 NegE_t * E_t + \varepsilon_t \quad (6)$$

The Hou, van Dijk and Zhang model (HDZ)

A cross-sectional regression methodology for earnings forecasting was implemented by Hou, van Dijk, and Zhang [2012], utilising a ten-year historical data period and integrating multiple predictive variables: lagged Earnings (E), Total assets (A), Dividend payout (D), Accruals (ACC), a binary indicator for dividend-distributing entities (DD), and a binary indicator for entities reporting lagged negative earnings (NegE). This analytical approach produced an R^2 value of approximately 0.8. The mathematical formulation developed by Li and Mohanram [2014] serves as the operational framework in the following manner:

$$E_{t+4} = \alpha_0 + \alpha_1 E_t + \alpha_2 NegE_t + \alpha_3 ACC_t + \\ + \alpha_4 A_t + \alpha_5 D_t + \alpha_6 DD_t + \varepsilon_t \tag{7}$$

In order to predict earnings per share (EPS), the future earnings are divided by the number of shares, assuming them constant.

4. Estimation technique – the XGBoost (XGB)

The eXtreme Gradient Boosting (XGBoost) algorithm emerged through the work of Chen and Guestrin [2016] as a substantial enhancement to conventional gradient-boosting methods, recognised for exceptional speed and effectiveness. The foundation of gradient boosting, a sophisticated machine learning approach extensively applied in regression and classification applications, originates from decision tree algorithms established during the 1960s. The technique generates predictive models through assemblages of weak predictors, predominantly basic decision trees. Sequential tree iterations target the correction of preceding prediction inaccuracies, with each subsequent tree specifically designed to address the antecedent model deficiencies. Gradient descent methodology facilitates iterative weight adjustments across weak learners. Iterations persist until loss function minimisation occurs or predetermined termination criteria activate. Selection of XGBoost from the available Gradient Boosting algorithms reflects comparable performance characteristics documented by Bentéjac et al. [2020], Boldini et al. [2023], and Türkmen and Sezen [2024], with XGBoost maintaining predominant usage. Multiple optimisation strategies characterise XG-

Boost implementation: regularisation mechanisms mitigate overfitting through loss function penalties, tree pruning procedures eliminate superfluous branches while strengthening model robustness, and parallelisation protocols expedite training operations. Additionally, the algorithm demonstrates proficiency in detecting data-inherent non-linear relationships. Application of machine learning methodologies such as eXtreme Gradient Boosting (XGBoost) to time series regression scenarios incorporating lagged dependent and independent variables establishes effective frameworks for non-stationary data pattern management. XGBoost's approach differs from the traditional statistical models by avoiding the prerequisite of stationarity among variables. It can flexibly capture shifts in trends and seasonal patterns within the data itself, adapting to these variations without requiring pre-stabilisation techniques, such as differencing, commonly employed in the conventional time series analysis. This model is designed to reduce prediction errors through iterative refinement, focusing on error minimisation in a way that reduces the influence of unrelated patterns or noise in the predictors. This iterative refinement develops a strong model representation without the rigid structural constraints typical of linear models, reducing the chance of spurious relationships. Moreover, XGBoost integrates regularisation, which is beneficial for managing issues such as multicollinearity that are commonly observed with many lagged variables. This regularisation further minimises overfitting risks and is especially useful when handling trends among variables, which can otherwise lead to erroneous interpretations in traditional regression setups. For a deeper understanding, Simon's [2020] book provides valuable insights. This methodology has been applied XGBoost as an estimation technique for the regression models for different sets of explanatory variables. Implementation of XGBoost, with hyperparameters fine-tuned [Banerjee, 2020] to overcome overfitting problem and to enhance the performance of forecasts, was facilitated using the xgb (version: 2.1.0) and hyperopt (version 0.2.7) libraries in Python (version 3.10.12).

5. Mean Arctangent Absolute Percentage Error (MAAPE)

The symbols $A_1^i,\ldots,A_4^i$ represent the actual earnings per share (EPS) for the first through fourth quarters of 2019 for firm *i*. Similarly, $F_1^i,\ldots,F_4^i$ denote the forecasted values of this variable for the corresponding periods (specifically

$\hat{Q}_t$, where $t = 37,\ldots,40$ for company i). For firm i in quarter j of 2019, the absolute percentage error (APE) of the prediction is defined as:

$$APE_j^i = \left| \frac{A_j^i - F_j^i}{A_j^i} \right| \tag{8}$$

However, the absolute percentage error (APE) exhibits a significant limitation: the metric generates infinite or undefined values when actual values approach or equal zero, a frequent situation in earnings forecasts. Moreover, actual values substantially below unity can produce extreme percentage errors (outliers). Additionally, zero actual values result in infinite APEs. To address this limitation, Kim and Kim [2016] proposed the arctangent absolute percentage error as an alternative metric within the field.

$$AAPE_j^i = \begin{cases} 0 & if\ A_j^i = F_j^i = 0 \\ arctan\left(\left|\frac{A_j^i - F_j^i}{A_j^i}\right|\right) & otherwise \end{cases} \tag{9}$$

This approach exploits the mathematical property of the arctangent function, which maps values from the interval [–∞,+∞] to the bounded interval [-π/2,π/2]. Therefore, the Mean Arctangent Absolute Percentage Error (MAAPE) for quarter j across all I companies in the dataset is expressed as:

$$MAAPE_j = \frac{1}{I}\sum_{i=1}^{I} AAPE_j^i = \frac{1}{I}\sum_{i=1}^{I} arctan\left(\left|\frac{A_j^i - F_j^i}{A_j^i}\right|\right) \tag{10}$$

The Mean Arctangent Absolute Percentage Error (MAAPE) across all four quarters and all I companies in the sample is calculated using:

$$MAAPE = \frac{1}{4}\sum_{j=1}^{4} MAAPE_j \tag{11}$$

The selection of MAAPE over MAPE (Mean Absolute Percentage Error) was motivated by the presence of companies with actual profits approaching zero in the analysed sample. When a single observation approaches zero while others remain substantially different from zero, the MAPE becomes extremely large (approaching infinity), thereby dominating the mean calculation and effectively nullifying the contribution of other observations.

Furthermore, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) – widely employed error metrics – are computed as robustness checks. The primary distinction between these metrics is that RMSE applies greater penalties to larger errors compared to MAE through the squaring of errors prior to averaging, resulting in increased sensitivity to outliers.

6. The statistical tests

Statistical assessment of MAAPE fluctuations among various models relies on the nonparametric two-sided Wilcoxon test, as originally described by Wilcoxon [1945]. This methodology operates through the paired difference analysis of the corresponding sample sets. The procedure requires minimal distributional assumptions, demanding solely symmetrical score differences and inter-difference independence. Ruland [1980] thoroughly examined the implementation of the Wilcoxon test within the verification contexts, specifically addressing the detection of statistically significant discrepancies among errors generated through diverse EPS modelling approaches. Tabular compilations of p-values undergo construction on both quarterly intervals (spanning quarters one through four) and aggregate temporal frameworks encompassing the complete quarterly dataset. Another test employed is the Brunner-Munzel [Brunner and Munzel, 2000] test. This test is favoured due to its robustness and applicability under a broader range of conditions, as it requires fewer assumptions than the Wilcoxon test. Specifically, it does not necessitate the symmetry of differences between the paired samples; it only assumes that the median of the differences is zero under the null hypothesis (indicating no difference between medians). The tests were executed using the statsmodels library (version 0.14.1) in Python.

$$\begin{aligned} &H_o : AAPEs\left(or\ MAAPEs\ for\ full\ years\right) \\ &of\ a\ pair\ of\ models\ are\ the\ same \end{aligned} \tag{12}$$

The rejection of the null hypothesis occurred when p-values fell below the conventional alpha level of 0.05, following the criterion widely employed in the statistical practice as documented by Ruland [1980] and numerous other researchers.

7. Results

Table 1 juxtaposes forecast errors for the respective quarters of 2019, gauged by MAAPE across all examined models. Both the seasonal random walk (SRW) model and the Pope Wang (PW) model achieve optimal predictive efficacy. The SRW model prevails in Q1, Q3, and throughout all quarters of 2019, while the EP model dominates in Q2 and Q4. Nevertheless, it merits emphasis that the hypothesis regarding the significant statistical distinction of their outcomes cannot be refuted in Q2 and Q4 utilising the Wilcoxon and Brunner-Munzel tests at a 0.05 significance level, as demonstrated in Table 2. Both tests present consistent results. Consequently, it may be posited that either the seasonal random walk model (SRW) or the Pope Wang model (PW) offers the most accurate predictions of earnings per share in 2019. It is consistent with the result of the research obtained by Harris and Wang [2019].

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
XGB_L	0.984	0.950	0.984	0.968	0.972
XGB_LT	0.790	0.766	0.812	0.825	0.798
XGB_RI	0.778	0.758	0.791	0.762	0.772
XGB_PW	0.699	0.676	0.714	0.709	0.699
XGB_EP	0.730	0.697	0.727	0.753	0.727
XGB_HDZ	0.809	0.776	0.840	0.811	0.809

Source: EquityRT, own calculations.

TABLE 2
P-values of statistical tests of forecast errors for SRW and respective models in 2019

Wilcoxon test p-values							
quarter	model	XGB_L	XGB_LT	XGB_RI	XGB_PW	XGB_EP	XGB_HDZ
1	SRW	0.000	0.000	0.000	0.028	0.001	0.000
2	SRW	0.000	0.002	0.072	0.609	0.503	0.016
3	SRW	0.000	0.000	0.000	0.009	0.007	0.000
4	SRW	0.000	0.000	0.487	0.334	0.182	0.032
ALL	SRW	0.000	0.000	0.000	0.880	0.027	0.000
Brunner-Munzel test p-values							
quarter	model	XGB_L	XGB_LT	XGB_RI	XGB_PW	XGB_EP	XGB_HDZ
1	SRW	0.000	0.006	0.004	0.032	0.005	0.002
2	SRW	0.000	0.009	0.092	0.719	0.601	0.072
3	SRW	0.000	0.002	0.005	0.012	0.011	0.003
4	SRW	0.000	0.005	0.567	0.391	0.225	0.039
ALL	SRW	0.000	0.006	0.007	0.980	0.033	0.002

Source: EquityRT, own calculations

The exemplary performance exhibited by the Pope Wang model implies the critical significance of variables utilised within this model, such as lagged Earnings (E), Book value (BV), Accruals (ACC), and Stock price (P), in forecasting the earnings of the Polish listed companies.

Robustness Checks

Robustness checks were conducted concerning both different years and various popular error metrics.

Throughout 2018's entirety, XGBoost-implemented models demonstrate forecast superiority over the seasonal random walk (SRW) benchmark, excluding Laurent's (L) specification, according to Table 3's evidence. Statistical equivalence to SRW predictions characterises exclusively the Lev and Thiagarajan

[1993] (LT) model's outputs. Wilcoxon and Brunner-Munzel test statistics for SRW-paired model comparisons appear in Table 4, with both methodologies producing concordant results. Among all specifications, the Pope-Wang [2005] (PW) model achieves minimal MAAPE-measured forecast errors (Table 3), with these predictions demonstrating statistical differentiation from SRW benchmarks (Table 4).

The 2017 period exhibits patterns analogous to 2019 observations. Superior forecasting performance characterizes the seasonal random walk (SRW) specification relative to alternative models (Table 3). Statistical differentiation marks these predictions when contrasted with other specifications, including the Pope-Wang [2005] (PW) model, as Table 4's first row indicates.

Temporal analysis reveals alternating dominance between two specifications: the seasonal random walk (SRW) and the XGBoost-estimated Pope-Wang [2005] (PW) model across varying periods.

TABLE 3

Summary statistics on forecast errors for all quarters 2017–2019

		2017	2018	2019
		MAAPE	MAAPE	MAAPE
model	SRW	0.686	0.711	0.687
	XGB_L	0.981	0.966	0.972
	XGB_LT	0.825	0.696	0.798
	XGB_RI	0.762	0.644	0.772
	XGB_PW	0.727	0.601	0.699
	XGB_EP	0.725	0.524	0.727
	XGB_HDZ	0.762	0.652	0.809

Source: EquityRT, own calculations.

The alternative error metrics, specifically the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE), constitute the evaluation framework displayed in Table 5 for model performance assessment across the entirety of quarters during [2019]. The seasonal random walk model demonstrated minimal error values for both RMSE and MAE metrics, maintaining consistency with the documented patterns from [2019].

TABLE 4

P-values statistical tests of forecast errors for all quarters 2017–2019 and SRW model

Wilcoxon test p-values							
year	model	XGB_L	XGB_LT	XGB_RI	XGB_PW	XGB_EP	XGB_HDZ
2017	SRW	0.000	0.000	0.000	0.029	0.023	0.000
2018	SRW	0.000	0.282	0.000	0.000	0.000	0.001
2019	SRW	0.000	0.000	0.000	0.880	0.027	0.000
Brunner-Munzel test p-values							
year	model	XGB_L	XGB_LT	XGB_RI	XGB_PW	XGB_EP	XGB_HDZ
2017	SRW	0.000	0.003	0.005	0.032	0.028	0.002
2018	SRW	0.000	0.312	0.005	0.002	0.000	0.005
2019	SRW	0.000	0.003	0.004	0.960	0.030	0.002

Source: EquityRT, own calculations.

Statistical significance testing through Wilcoxon and Brunner-Munzel methodologies, with corresponding p-values documented in Table 6, indicates substantial differences when comparing the SRW model against alternative model configurations, excluding the PW model pairing. The seasonal random walk (SRW) model's predictive outputs demonstrate statistically significant superiority in RMSE and MAE measurements relative to competing models, with the sole exception being the Pope-Wang (PW) model operationalised via XGBoost methodology. The temporal specificity of these findings remains confined to [2019], as indicated in the analysis, with divergent patterns emerging in alternative temporal periods, particularly [2018], during which competing models achieved superior MAAPE performance relative to the seasonal random walk (SRW) model.

TABLE 5

Summary statistics on forecast errors for RMSE and MAPE in all quarters 2019

	SRW	XGB_L	XGB_LT	XGB_RI	XGB_PW	XGB_EP	XGB_HDZ
RMSE	0.951	1.957	1.345	1.163	1.053	1.195	1.250
MAE	0.717	1.452	0.997	0.870	0.820	0.908	0.942

Source: EquityRT, own calculations.

TABLE 6

P-values of statistical tests of forecast errors for RMSE and MAE in 2019

Wilcoxon test p-values							
measure	model	XGB_L	XGB_LT	XGB_RI	XGB_PW	XGB_EP	XGB_HDZ
RMSE	SRW	0.000	0.000	0.016	0.343	0.047	0.001
MAE	SRW	0.000	0.000	0.001	0.158	0.071	0.000
Brunner-Munzel test p-values							
measure	model	XGB_L	XGB_LT	XGB_RI	XGB_PW	XGB_EP	XGB_HDZ
RMSE	SRW	0.000	0.002	0.006	0.394	0,.034	0.004
MAE	SRW	0.000	0.003	0.004	0.210	0.000	0.011

Source: EquityRT, own calculations.

Companies were grouped based on EquityRT information into two categories: Covered by analysts (43 firms) and Not-Covered by analysts (224 firms) in 2019. The forecast errors for each group were assessed for 2019, with MAAPE values presented in Table VII and the results of the Wilcoxon and Brunner-Munzel tests shown in Table VIII. The findings demonstrate a similar pattern as in the full sample with errors only slightly lower for Covered companies. Also in this case the Pope-Wang (PW) model is the only one that makes errors higher, but statistically indistinguishable from the seasonal random walk (SRW) model.

TABLE 7

MAAPE forecast errors by analyst coverage in 2019

	SRW	XGB_L	XGB_LT	XGB_RI	XGB_PW	XGB_EP	XGB_HDZ
Covered	0.656	0.954	0.776	0.753	0.667	0.701	0.789
Not- Covered	0.692	0.986	0.801	0.788	0.707	0.740	0.818

Source: EquityRT, own calculations.

TABLE 8
P-values of statistical tests of forecast errors by analyst coverage in 2019

Wilcoxon test p-values							
	model	XGB_L	XGB_LT	XGB_RI	XGB_PW	XGB_EP	XGB_HDZ
Covered	SRW	0.000	0.000	0.021	0.285	0.034	0.001
Not-Covered	SRW	0.000	0.000	0.001	0.178	0.040	0.000
Brunner-Munzel test p-values							
	model	XGB_L	XGB_LT	XGB_RI	XGB_PW	XGB_EP	XGB_HDZ
Covered	SRW	0.000	0.001	0.003	0.444	0.033	0.006
Not-Covered	SRW	0.000	0.000	0.001	0.228	0.000	0.013

Source: EquityRT, own calculations.

8. Discussion

Superior earnings predictions enable investors to identify undervalued stocks before the market fully incorporates the information, potentially generating alpha through better portfolio allocation decisions and position sizing based on the confidence intervals around forecasts. For portfolio decisions, more accurate forecasts lead to improved asset allocation, better risk-adjusted returns, and the ability to capitalise on market inefficiencies by overweighting undervalued securities and avoiding overvalued ones. The reduction in valuation errors leads to securities pricing closer to their intrinsic values, minimising the probability of large losses from fundamental misjudgments while maximising the capture of upside potential when their forecasts prove correct.

The seasonal random walk (SRW) and Pope Wang (PW) model demonstrate superior performance in EPS forecasting within the Polish market context. Such outcomes correspond with the conclusions drawn by Kuryłek [2023a, 2023b], who documented that the simple seasonal random walk model outperforms traditional forecasting techniques including ARIMA and exponential smoothing – methods that prove effective in the US market – when applied to the Polish data. The present investigation employs a comprehensive multivariate framework, contrasting with the univariate methodology utilised in

Kuryłek's [2023a, 2023b] prior studies. The structural simplicity characterising the Polish stock market provides a plausible explanation for the SRW model's effectiveness, particularly given that both SRW and PW represent less complex alternatives to the sophisticated models developed by Lev and Thiagarajan [1993], which maintain prominence in academic literature.

The Pope Wang (PW) model suggests that variables such as current or delayed Earnings (E), Book value (BV), Accruals (ACC), and Stock price (P) may significantly contribute to predicting future earnings. Therefore, the above mentioned accounting variables should be watched carefully, when making investment decisions. The inclusion of stock price, which aligns with the strongest form of the Efficient Market Hypothesis (EMH), implying that stock prices provide insight into future earnings prospects, is particularny intriguing.

Furthermore, robustness checks confirm the dominance of the seasonal random walk model (SRW) in 2017 and 2019, while in 2018, the Pope Wang model emerged as the top performer. The statistically significant differences in the forecast errors between these two models in 2017 and 2018 underscore their distinct predictive capabilities. Given the inherent uncertainty in predicting the best-performing model for a given year, it is prudent to utilise both models and to integrate their forecasts.

The simplistic seasonal approach suggests that employing sophisticated techniques beyond the ordinary seasonal random walk for EPS forecasting in the Polish investment contexts may not be practical. However, the strong performance of the more advanced Pope Wang model implies the relevance of financial variables in forecasting future earnings, which sometimes might outperform predictions solely based on the naïve seasonal random walk model.

9. Conclusions

Seven predictive frameworks undergo evaluation for earnings per share projection capabilities: seasonal random walk (SRW), Laurent methodology (L), Lev-Thiagarajan framework (LT), Residual Income approach (RI), Pope-Wang methodology (PW), Earnings Persistence framework (EP), and Hou-van Dijk-Zhang approach (HDZ). Automated earnings per share projection carries heightened significance within the developing economies characterised by sparse financial analyst coverage of publicly traded entities, with Poland serving as a representative case. Quarterly observations encompassing 267 Polish corporations during the 2010–2019 period constitute the dataset, with Mean Arctangent Absolute Error (MAAPE) serving as the forecast accuracy metric.

Superior performance emerged from both the SRW framework and the Pope-Wang methodology implemented through XGBoost, yielding minimal prediction errors and generating enhanced representations of the Polish market dynamics relative to alternative approaches. Such outcomes persist across corporate subsets regardless of the analyst coverage status. Additionally, the Pope-Wang framework identifies key predictive indicators including equity valuations, accounting book values, and accrual components that augment earnings forecasting capability beyond current period earnings alone. These findings align with empirical evidence documented by Harris and Wang [2019].

While XGBoost often achieves superior predictive accuracy compared to the traditional linear models, this accuracy gain comes at the cost of interpretability, forcing investors to choose between maximising returns through complex models or maintaining the ability to understand and communicate the economic rationale behind their investment decisions. The method's high sensitivity to hyperparameter tuning creates risks of overfitting to historical data, potentially leading to poor out-of-sample performance and significant portfolio losses when market regimes shift or unprecedented events occur.

This study is the first to apply the XGBoost algorithm to a broad set of multivariate EPS forecasting models on the Polish stock market, demonstrating that even simple models like the seasonal random walk can outperform more complex alternatives in this context. By highlighting the predictive strength of the Pope-Wang model and the relevance of specific financial variables, the research bridges machine learning techniques with the traditional accounting-based forecasting, offering new insights for emerging markets with limited analyst coverage.

The correlation between forecasting accuracy and firm size constitutes a potential area for the subsequent investigation. The industry sector represents a critical factor in the determination of optimal models for EPS forecasting. Time series transformations for standardising EPS distributions warrant examination for their potential analytical value. Multivariate models incorporate sophisticated techniques including artificial neural networks. The comparative analysis of predictive accuracy between optimal mechanised models and market analyst forecasts constitutes a compelling research direction. The performance and accuracy assessment of diverse predictive models and financial analyst projections during the economic turbulence periods, including the [2008–2009] financial crisis, the COVID-19 pandemic, and the onset of the Ukraine conflict, represents an additional significant research domain. The SRW model identifies seasonal patterns with implications for investment strategy formulation. This strategic approach potentially contradicts the weak form of the Efficient Market Hypothesis (EMH). Conversely, the Pope-Wang model's incorporation of stock price data, consistent with the weak form of the Efficient Market Hypothesis (EMH), suggests stock prices contain information regarding the future earnings potential. Consequently, this domain requires additional investigation.

References

Abarbanell J., Bushee B., 1997, *Fundamental Analysis, Future EPS, and Stock Prices*, "Journal of Accounting Research", vol. 35(1), pp.1–24, DOI: 10.2307/2491464.

Ahmadpour A., Etemadi H., Moshashaei S., 2015, *Earnings Per Share Forecast Using Extracted Rules from Trained Neural Network by Genetic Algorithm*, "Computational Economics", vol. 46(1), pp. 55–63, DOI: 10.1007/s10614-014-9455-6.

Ball R., Ghysels E., 2017, *Automated Earnings Forecasts: Beat Analysts or Combine and Conquer?*, "Management Science", vol. 64(10), pp. 4936–4952, DOI: 10.1287/mnsc.2017.2864.

Ball R., Watts R., 1972, *Some Time Series Properties of Accounting Income*, "The Journal of Finance", vol. 27(3), pp. 663–681, DOI: 10.1111/j.1540–6261.1972.tb00991.x.

Banerjee P., 2020, *A Guide on XGBoost Hyperparameters Tuning*, https://www.kaggle.com/code/prashant111/a-guide-on-xgboost-hyperparameters-tuning [date of access: 14.07.2024].

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", vol. 59(2), pp. 163–176.

Bentéjac C., Csörgő A., Martínez-Muñoz G., 2020, *A Comparative Analysis of Gradient Boosting Algorithms*, "Artificial Intelligence Review", vol. 54(3), pp. 1937–1967, DOI: 10.1007/s10462-020-09896-5.

Boldini D., Grisoni F., Kuhn D., Friedrich L., Sieber S.A., 2023, *Practical Guidelines for the Use of Gradient Boosting for Molecular Property Prediction*, "Journal of Cheminformatics", vol. 15(1), pp. 1–13, DOI: 10.1186/s13321-023-00743-7.

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", vol. 17(4), pp. 944–968, DOI: 10.1007/s11142-012-9185-8.

Brandon Ch., Jarrett J.E., Khumawala S.B., 1986, *Comparing Forecast Accuracy for Exponential Smoothing Models of Earnings-per-share Data for Financial Decision Making*, "Decision Sciences", vol. 17(2), pp. 186–194, DOI: 10.1111/j.1540-5915.1986.tb00220.x.

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", vol. 13(2), pp. 10–15, DOI: 10.1108/eb013581.

Brooks L.D., Buckmaster D.A., 1976, *Further Evidence Of The Time Series Properties Of Accounting Income*, "The Journal of Finance", vol. 31(5), pp. 1359–1373, DOI: 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", vol. 9(1), pp. 61–87, DOI: 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", vol. 17(1), pp. 179–189, DOI: 10.2307/2490312.

Brunner E., Munzel U., 2000, *The Nonparametric Benhrens-Fisher Problem: Asymptotic Theory and a Amall-sample Approximation*, "Biometrical Journal", vol. 42, pp. 17–25, DOI: 10.1002/(sici)1521-4036(200001)42:1<17::aid-bimj17>3.0.co;2-u.

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, pp. 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", vol. 47(1), pp. 32–41, DOI: 10.1016/j.dss.2008.12.011.

Chen Y., Chen S., Huang H., Sangaiah A., 2020, *Applied Identification of Industry Data Science Using an Advanced Multi-componential Discretization Model*, "Symmetry", vol. 12(10), pp. 1–28, DOI: 10.3390/sym12101620.

Chen T., Guestrin C., 2016, *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, DOI: 10.1145/2939672.2939785.

Conroy R., Harris R., 1987, *Consensus Forecasts of Corporate Earnings: Analysts' Forecasts and Time Series Methods*, "Management Science", vol. 33(6), pp. 725–738, DOI: 10.1287/mnsc.33.6.725.

Delen D., Kuzey C., Uyar A., 2013, *Measuring Firm Performance Using Financial Ratios: A Decision Tree Approach*, "Expert Systems with Applications", vol. 40(10), pp. 3970–3983, DOI: 10.1016/j.eswa.2013.01.012.

Dreher S., Eichfelder S., Noth F., 2024, *Does IFRS Information on Tax Loss Carryforwards and Negative Performance Improve Predictions of Earnings and Cash Flows?*, "Journal of Business Economics", vol. 94(1), pp. 1–39, DOI: 10.1007/s11573-023-01147-7.

Elamir E., 2020, *Modeling and Predicting Earnings per Share via Regression Tree Approaches in Banking Sector: Middle East and North African Countries Case*, "Investment Management and Financial Innovations", vol. 17(2), pp. 51–68. DOI: 10.21511/imfi.17(2).2020.05.

Elton E.J., Gruber M.J., 1972, *Earnings Estimates and the Accuracy of Expectational Data*, "Management Science", vol. 18(8), pp. B409-B424, DOI: 10.1287/mnsc.18.8.B409.

Foster G., 1977, *Quarterly Accounting Data: Time-Series Properties and Predictive-Ability Results*, "The Accounting Review", vol. 52(1), pp. 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", vol. 22(1), pp. eRAMF210164, DOI: 10.1590/1678-6971/eramf210164.

Gramacy R.B., Gerakos J., 2013, *Regression-Based Earnings Forecasts*, Chicago Booth Research Paper, no. 12–26, DOI: 10.2139/ssrn.2112137.

Griffin P., 1977, *The Time-Series Behavior of Quarterly Earnings: Preliminary Evidence*, "Journal of Accounting Research", vol. 15(1), pp. 71–83, DOI: 10.2307/2490556.

Harris R.D. F., Wang P., 2019, *Model-based Earnings Forecasts vs. Financial Analysts' Earnings Forecasts*, "British Accounting Review", vol. 51(4), pp. 424–437, DOI: 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", vol. 53(3), pp. 504–526, DOI: 10.1016/j.jacceco.2011.12.001.

Jarrett J.E., 2008, *Evaluating Methods for Forecasting Earnings Per Share*, "Managerial Finance", vol. 16, pp. 30–35, DOI: 10.1108/eb013647.

Johnson T.E., Schmitt T.G., 1974, *Effectiveness of Earnings Per Share Forecasts*, "Financial Management", vol. 3(2), pp. 64–72, DOI: 10.2307/3665292.

Kim S., Kim H., 2016, *A New Metric of Absolute Percentage Error for Intermittent Demand Forecasts*, "International Journal of Forecasting", vol. 32(3), pp. 669–679, DOI: 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*, "Journal of Banking and Financial Economics", vol. 1(19), pp. 26–43, DOI: 10.7172/2353-6845.jbfe.2023.1.2.

Kuryłek W., 2023b, *Can Exponential Smoothing Do Better thanSeasonal Random Walk for Earnings per Share Forecasting in Poland?*, "Bank & Credit", vol. 54(6), pp. 651–672.

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

Laurent C., 1979, *Improving the Efficiency and Effectiveness of Financial Ratio Analysis*, "Journal of Business Finance & Accounting", vol. 6(3), pp. 401–413, DOI: 10.1111/j.1468-5957.1979.tb01099.x.

Lev B., Souginannis T., 2010, *The Usefulness of Accounting Estimates for Predicting Cash Flows and Earnings*, "Review of Accounting Studies", vol. 15(4), pp. 779–807, DOI: 10.1007/s11142-009-9107-6.

Lev B., Thiagarajan S., 1993, *Fundamental Information Analysis*, "Journal of Accounting Research", vol. 31(2), pp. 190–215, DOI: 10.2307/2491270.

Li K.K., 2011, *How Well Do Investors Understand Loss Persistence?*, "Review of Accounting Studies", vol. 16(3), pp. 630–667, DOI: 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", vol. 19(3), pp. 1152–1185, DOI: 10.1007/s11142-014-9282-y.

Lorek K.S., 1979, *Predicting Annual Net Earnings with Quarterly Earnings Time-Series Models*, "Journal of Accounting Research", vol. 17(1), pp. 190–204, DOI: 10.2307/2490313.

Lorek K. S, Willinger G.L., 1996, *A Multivariate Time-series Model for Cash-flow Data*, "Accounting Review", vol. 71, pp. 81–101.

Ohlson J.A., 1995, *Earnings*, *Book Values*, *and Dividends in Equity Valuation*, "Contemporary Accounting Research", vol. 11(2), pp. 661–687, DOI: 10.1092/7tpj-rxqn-tqc7-ffae.

Ohlson J.A., 2001, *Earnings*, *Book Values*, *and Dividends in Equity Valuation: An Empirical Perspective*, "Contemporary Accounting Research", vol. 18(1), pp. 107–120, DOI; 10.1092/7tpj-rxqn-tqc7-ffae.

Ohlson J.A., Juettner-Nauroth B.E., 2005, *Expected EPS and EPS Growth as Determinants of Value*, "Review of Accounting Studies", vol. 10(2–3), pp. 349–365, DOI: 10.1007/s11142-005-1535-3.

Pagach D.P., Warr R.S., 2020, *Analysts versus Time-series Forecasts of Quarterly Earnings: A Maintained Hypothesis Revisited*, "Advances in Accounting", vol. 51, pp. 1–15, DOI: 10.1016/j.adiac.2020.100497.

Pope P.F., Wang P., 2005, *Earnings Components*, *Accounting Bias and Equity Valuation*, "Review of Accounting Studies", pp. 10(4), pp. 387–407, DOI: 10.1007/s11142-005-4207-4.

Pope P.F., Wang P., 2014, *On the Relevance of Earnings Components: Valuation and Forecasting Links*, "Review of Quantitative Finance and Accounting", vol. 42, pp. 399–413, DOI: 10.1007/s11156-013-0347-y.

Ruland W., 1980, *On the Choice of Simple Extrapolative Model Forecasts of Annual Earnings*, "Financial Management", vol. 9(2), pp. 30–37, DOI: 10.2307/3665165.

Simon J., 2020, *Learn Amazon SageMaker: A Guide to Building*, *Training*, *and Deploying Machine Learning Models for Developers and Data Scientists*, Packt> Birmingham – Mumbai.

Türkmen G., Sezen A., 2024, *A Comparative Analysis of XGBoost and LightGBM Approaches for Human Activity Recognition: Speed and Accuracy Evaluation.* "International Journal of Computational and Experimental Science and Engineering", vol. 10(2), pp. 262–270, DOI: 10.22399/ijcesen.329.

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

Wilcoxon F., 1945, *Individual Comparisons by Ranking Methods*, "Biometrics", vol. 1, pp. 80–83, DOI: 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),* DOI: /10.1109/icetci55101.2022.9832404.