Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663

https://www.czasopisma.uni.lodz.pl/foe/
1(370) 2025
DOI: https://doi.org/10.18778/0208-6018.370.01

Wojciech Kuryłek https://orcid.org/0000-0003-0692-3300

University of Warsaw, Faculty of Management, Warsaw, Poland, wkurylek@wz.uw.edu.pl

An Application of Autoregressive Distributed Lag-Models for Earnings per Share Forecasting in Poland

Abstract: This investigation delves into the significance of precise earnings forecasts for publicly traded companies in achieving investment success. It emphasises the importance of this aspect, particularly in markets with limited analyst coverage, such as emerging markets including Poland. The study assesses the accuracy of predictions generated using the Autoregressive Distributed Lag (ARDL) framework with the XGBoost type of modelling and various methods of combining forecasts compared to the seasonal random walk model. Positioned as an intermediate step between time series and multivariate forecasting, these models are applied to earnings per share (EPS) data of companies listed on the Warsaw Stock Exchange from 2009 to 2019, i.e. the last financial crisis and the pandemic shock. The seasonal random walk model attained the lowest error rates based on the Mean Arctangent Absolute Percentage Error (MAAPE) metric, a conclusion substantiated by rigorous statistical tests and robustness checks employing different periods and error metrics. The enhanced performance of the simpler seasonal random walk model may be ascribed to the relatively uncomplicated nature of the Polish stock market.

Keywords: earnings per share, random walk, Autoregressive Distributed Lags, XGBoost, financial forecasting, Warsaw Stock Exchange

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

1. Introduction

EPS forecasts provide foundational insights into a company's anticipated profitability and cash flow potential. They are critical for estimating future value, directly influencing a stock price. The reliability of EPS estimates matters greatly; any inaccuracies can result in a flawed valuation, potentially leading to mispricing and investment risk. In developed markets, extensive analyst coverage helps ensure these forecasts are meticulously researched, with frequent revisions that keep them aligned with changing financial conditions. However, in emerging markets, such as Poland, analyst coverage is limited, with fewer than 20% (according to EquityRT data) of stocks receiving consistent scrutiny. This gap in coverage means that EPS estimates for these companies may be sparse, creating a need for alternate forecasting methods. Given the limited analyst coverage in emerging markets, statistical and machine learning models are valuable for generating consistent and potentially more accurate EPS forecasts. These models analyse historical data, industry metrics, and market trends to produce projections, potentially bridging the gap created by scarce analyst insights. Furthermore, machine learning algorithms, which can handle large datasets and identify complex non-linear patterns, might provide better forecasts than traditional statistical models.

Univariate time series models focus solely on past values of a single series to predict future outcomes. This makes them highly tailored, as each model captures the specific patterns and seasonality of an individual company, preserving idiosyncratic trends and volatility. The advantage is simplicity; these models are less computationally intensive, more transparent, and capture direct company-specific patterns. However, they lack contextual insights, ignoring external influences that may have substantial impacts, such as industry-wide economic changes or competitive factors. On the other hand, multivariate models, adopt a 'one-size-fits-all' approach, allowing for simultaneous analysis across multiple companies by assuming interrelatedness or commonalities among them. This can be advantageous in sectors where companies follow similar economic cycles or regulatory patterns, as the models assume shared structures. Nevertheless, this generalisation may reduce precision when individual companies deviate from the broader trends, making it less adaptable to companies with unique patterns. Autoregressive Distributed Lag models (ARDL) offer a balanced, intermediary approach by enabling individual modelling of each time series while incorporating lagged values of external variables. This approach respects the idiosyncratic nature of each company while accounting for common external influences, such as sectoral trends or macroeconomic indicators, allowing the model to capture both internal and contextual dynamics. ARDL models have the flexibility to analyse short-term and long-term relationships across multiple lags, providing a more nuanced forecast than traditional univariate or multivariate methods.

Bansal, Nasseh, and Strauss (2015) found that individual ARDL models typically performed poorly in forecasting earnings. However, combining forecasts from the ARDL models and using different techniques consistently led to more accurate predictions than those of a standalone autoregressive (AR) benchmark. The ARDL models in this study integrated accounting-based predictor variables with key macroeconomic indicators, such as industrial production, inflation, and interest rate spreads. In a similar study, Ball and Ghysels (2017) evaluated ARDL-based combination forecasts against financial analysts' predictions, noting that ARDL combinations were often more accurate.

In the presented article, ARDL models are paired with eXtreme Gradient Boosting (XGB), a machine-learning technique shown to capture complex, non-linear relationships efficiently and perform well on structured data with fewer assumptions than traditional statistical models (Qian, 2017; Dou, Tan, Xie, 2023; Fourkiotis, Tsadiras, 2024). XGB's adaptability and lower reliance on assumptions make it well-suited for handling the intricate dynamics in earnings prediction. Hence, this research compares the forecasts given by this model to the seasonal random walk model, which is a benchmark for the Polish market (Kuryłek, 2023a; 2023b; 2024). The study aims to answer the research question of to what extent the abovementioned technique can be effectively adopted in the Polish market, which has never been analysed in this way in previous research. Therefore, the research hypothesis is that ARDL-type models using the XGBoost technique can outperform the benchmark SRW model. The article encompasses quarterly EPS data for 267 companies listed on the Polish stock exchange spanning the period from the 2008–2009 financial crisis through the 2020 pandemic.

Instead of depending exclusively on the traditional mean absolute percentage error (MAPE), which can be highly sensitive to very small actual values, this study adopts an alternative metric – the mean arctangent absolute percentage error (MAAPE) – as introduced by Kim and Kim (2016), and uses it for evaluation.

The research aims to: (1) test Autoregressive Distributed Lag (ARDL) models and forecast combination methods for EPS prediction, (2) validate results using multiple error metrics, timeframes, and statistical tests, (3) implement MAAPE for scenarios with near-zero profits, and (4) explain how these findings can inform Polish stock investment strategies.

2. Literature review

Scholarly investigation into Earnings per Share (EPS) prediction through computational methods commenced in the 1960s. Initial academic discourse centered predominantly on autoregressive integrated moving average (ARIMA) frameworks, constituting the earliest category of models subjected to rigorous examination (Ball, Watts, 1972; Watts, 1975; Foster, 1977; Griffin, 1977; Brown, Rozeff, 1979). Research findings exhibited considerable heterogeneity: certain analyses supported the efficacy of elemental random walk approaches, intimating that enhanced complexity did not consistently yield superior predictive capacity, while alternative studies produced contradictory evidence. The Polish market was the subject of comparable examination by Kurylek (2023a).

A scholarly consensus gradually crystallized around the typically superior accuracy of ARIMA-based methodologies (Lorek, 1979; Bathke Jr., Lorek, 1984). This academic position remained relatively unchallenged until approximately the late 1980s, when the prevailing academic perception shifted toward acknowledging financial analysts' predictions as superior to time series model outputs (Brown et al., 1987). It should be noted, however, that Conroy and Harris (1987) observed that analyst superiority was most pronounced in near-term forecasting horizons and progressively diminished over extended timeframes. This academic perspective predominated until contemporary scholarship once again interrogated the presumptive superiority of analyst forecasts relative to time series methodologies (Lacina, Lee, Xu, 2011; Bradshaw et al., 2012; Pagach, Warr, 2020; Gaio et al., 2021).

The application of exponential smoothing techniques for EPS forecasting has constituted a parallel research trajectory since the late 1960s (Ball, Watts, 1972; Elton, Gruber, 1972; Johnson, Schmitt, 1974; Brooks, Buckmaster, 1976; Ruland, 1980; Brandon, Jarrett, Khumawala, 1987; Jarrett, 2008). This methodological approach has generated inconsistent empirical outcomes across various studies.

Multivariate cross-sectional models demonstrated enhanced predictive capacity compared to both firm-specific and common-structure ARIMA frameworks, as established by Lorek and Willinger (1996). Theoretical foundations connecting earnings projections to accounting metrics and equity valuations employed in contemporary research were established by Pope and Wang (2005; 2014). Subsequent investigation by Harris and Wang (2019) determined that the Pope-Wang model (Pope, Wang, 2005) generally exhibited reduced bias and enhanced predictive accuracy relative to alternative forecasting frameworks.

Bansal, Nasseh, and Strauss (2015) found that the forecasting performance of individual ARDL models was generally poor in the case of earnings prediction. In contrast, combination forecast methods, based on individual ARDL model forecasts using several combination methods, including the simple mean, discount mean squared forecast error, clusters, and principal components, were able to consistently outperform an autoregressive (AR) benchmark model. In the ARDL model, the authors considered accounting-based predictor variables as well as the following macroeconomic variables: Industrial production, the CPI inflation index, the 3M money market rate, the Spread between the 10-year treasury and 3M rate, and the Default spread between BAA corporate bonds and AAA corporate bonds. Ball and Ghysels (2017) compared the errors of combined Autoregressive Distributed Lag (ARDL) models with financial analysts' forecasts in the prediction of quarterly earnings using a similar set of variables. They also found that constructed forecasts were more accurate and had forecast errors that were smaller than analysts' predictions when forecast dispersion was high and when the company size was smaller. They used different forecast combination schemes. In this article, ADRL-type models are used with the eXtreme Gradient Boosting (XGB) approach, since this machine-learning technique is far more efficient in making predictions and capturing complex non-linear patterns over traditional statistical methods for tabular data (Qian, 2017; Dou, Tan, Xie, 2023; Fourkiotis, Tsadiras, 2024). They also require much fewer assumptions to be satisfied than traditional statistical models since they learn from data.

3. Data and methodology

3.1. Data

The Warsaw Stock Exchange, integrated into the European Union post-2004, demonstrates considerable market robustness, evidenced by its $197 billion market capitalization and 774 listed entities as of year-end 2021. A notable disparity exists, however, in analyst coverage compared to Western European and American markets. Statistical data from 2019 reveals that merely one-fifth of the 711 companies listed received analytical scrutiny. This deficiency necessitates the implementation of statistical forecasting methodologies for crucial financial metrics utilizing analytical frameworks. The investigation concentrates primarily on earnings per share (EPS) chronological data and additional financial parameters extracted from the EquityRT financial analysis platform. Examination encompasses EPS developmental patterns of Warsaw Stock Exchange-listed corporations spanning from the first quarter of 2010 through the fourth quarter of 2019, a period characterized by substantial economic transformations: specifically, the financial crisis of 2008–2009 and the emergence of the COVID-19 pandemic in 2020. Beginning in 2020, the global landscape shifted into a phase of heightened turbulence, initiated by the COVID-19 pandemic, and further intensified by the onset of the energy crisis following the conflict in Ukraine in 2022. As a result, 2019 marked the last period of relative stability, offering a reliable environment for model testing and calibration. Models that fail to perform well during such stable and predictable periods are even less reliable during times of heightened market disruption, where extrapolating past trends into future outcomes becomes highly susceptible to error. Thus, such a period of stability was selected. Additionally, continuity in sampling was prioritised in this analysis to ensure direct comparability of results with those in other studies (Kuryłek 2023a; 2023b; 2024), thereby enhancing the consistency and robustness of the findings. The forecasting methodology employs quarterly data spanning from the first quarter of 2010 through the fourth quarter of 2018, constituting a 36-quarter estimation period. The subsequent four quarters of 2019 have been allocated for out-of-time validation assessment. The prognostic horizon extends from one to four quarters into the future, with supplementary analysis utilizing the 2017–2018 period as validation cohorts. Following a thorough examination of the dataset, which excluded instances of stock splits and reverse splits, a total of 267 corporate entities remained for analytical purposes.

3.2. The models

In this research, the following models were applied to EPS modelling: the seasonal random walk model and ARDL-type models estimated using XGBoost with three different forecast combination methods.

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

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

The forecast, denoted as is.

A benchmark forecasting approach utilizes the earnings per share value from four quarters prior, expressed as $\widehat{EPS}_t = EPS_{t-4}$, thereby eliminating the necessity for parameter estimation procedures. This methodological framework derives empirical validation from multiple studies by Kuryłek (2023a; 2023b; 2024), wherein superior predictive performance relative to conventional time series models was demonstrated within the Polish economic context.

Autoregressive Distributed Lag (ARDL) – type models

Autoregressive Distributed Lag models (ARDL) offer a balanced, intermediary approach by enabling individual modelling of each time series while incorporating lagged values of external variables to capitalise on the advantages of both time series and multivariate models. This approach respects the idiosyncratic nature of each company while accounting for common external influences, such as sectoral trends or macroeconomic indicators, allowing the model to capture both internal and contextual dynamics. Thus, ARDL models provide more nuanced forecasts than traditional univariate or multivariate methods. Some researchers argue (Bansal, Nasseh, Strauss, 2015; Ball, Ghysels, 2017) that this approach could yield superior results to EPS forecasting.

Pope and Wang (2005; 2014) explained the theoretical connection between earnings forecasts and three key accounting variables – Book Value (BV), Accruals (ACC), and non-accounting factors like Stock Price (P) – based on the principles of linear valuation, no-arbitrage, dividend irrelevance, and clean surplus accounting. These accounting variables are computed on a per-share basis to facilitate earnings per share (EPS) forecasting. Furthermore, the analysis incorporates four macroeconomic variables, including Industrial production (IP), the CPI inflation index (CPI), the 3M money market rate (3M), and the Spread between 10-years treasury and 3M rate (SPREAD). It is worth noting that data on the Default spread between BAA corporate bonds and AAA corporate bonds are scarce in Poland due to the limited popularity of bond credit ratings in the country. These macroeconomic variables are referenced in the studies of Bansal, Nasseh, and Strauss (2015) and Ball and Ghysels (2017). Each of these seven variables is separately applied in the ADRL model formulated as follows:

$$EPS_{t+4}^{i} = \alpha_0 + \alpha_1 EPS_t + \alpha_2 EPS_{t-4} + \beta_1 X_t^i + \beta_2 X_{t-4}^i + \varepsilon_t \tag{2}$$

and in a more general form, as used in this article, it can be rewritten as:

$$EPS_{t+4}^{i} = f\left(EPS_t,\, EPS_{t-4},\, X_t^i,\, X_{t-4}^i\right) + \varepsilon_t\,. \tag{3}$$

given X_t^i as the i-th independent variable in the quarter t, various forecasts $\widehat{EPS}_t^i$ can thus be generated employing these variables. These individual forecasts may subsequently be amalgamated into a singular forecast utilising diverse forecasts combination techniques. Positioned as an intermediate solution between univariate time series and multivariate regression models, the ARDL model leverages the identification of unique patterns for each company through a multivariate approach.

3.3. Mean Arctangent Absolute Percentage Error (MAAPE)

Denote $A_1^i, \ldots, A_4^i$ as the actual earnings per share (EPS) from the first to the fourth quarter of 2019 for a specific company i. Similarly, let $F_1^i, \ldots, F_4^i$ represent the forecasts of this variable for the corresponding periods (i.e., $\hat{Q}_t$, where $t = 37, \ldots, 40$ for the i-th company). For any firm i, during the j-th quarter of 2019, the absolute percentage error (APE) of such prediction can be mathematically expressed as:

$$APE_j^i = \left| \frac{A_j^i - F_j^i}{A_j^i} \right|. \tag{4}$$

A notable limitation of the absolute percentage error (APE) methodogy concerns its mathematical behavior when confronted with actual values approximating or equaling zero- a common occurrence in earnings prediction contexts. The APE metric potentially generates infinite or undefined outcomes in such scenarios. Furthermore, when actual values fall significantly below unity, the resultant percentage errors may manifest as statistical anomalies or outliers of considerable magnitude. The mathematical impossibility of calculating APE when actual values equal zero presents an additional methodological challenge. Addressing these technical constraints, the academic literature has been enhanced by Kim and Kim's (2016) contribution of the arctangent absolute percentage error, which represents an innovative methodological advancement in this analytical domain.

$$AAPE_j^i = \begin{cases} 0 & if\ A_j^i = F_j^i = 0 \\ arctan\left(\left| \frac{A_j^i - F_j^i}{A_j^i} \right| \right) & otherwise \end{cases} \tag{5}$$

The reasoning behind this stems from the characteristic of the arctan function, which maps values from the range of [−∞, +∞] to the interval [−π/2, π/2]. As a result, the Mean Arctangent Absolute Percentage Error (MAAPE) for the j-th quarter among all I companies in the dataset can be expressed as:

$$MAAPE_j = \frac{1}{I}\sum_{i=1}^{I} AAPE_j^i = \frac{1}{I}\sum_{i=1}^{I} \arctan\left(\left|\frac{A_j^i - F_j^i}{A_j^i}\right|\right). \tag{6}$$

The selection of Mean Arctangent Absolute Percentage Error (MAAPE) instead of Mean Absolute Percentage Error (MAPE) was deliberate, given the presence within the analyzed sample of corporations exhibiting profit figures approximating zero. In circumstances where a singular observation tends toward zero while remaining observations maintain substantial distance from this value, the MAPE calculation for this particular observation may approach infinity, thus potentially overwhelming the mean computation and rendering other observations statistically insignificant. Regarding individual models, MAAPE coefficients undergo averaging across the various financial instruments under consideration.

3.4. The statistical test

A nonparametric two-sided Wilcoxon test, as elucidated by Wilcoxon (1945), functions as the methodological approach for evaluating statistical significance among MAAPE variations across different models. The implementation of this paired difference test for matched samples necessitates minimal distributional assumptions, requiring only symmetry in score differences and independence between these differences. The application of the Wilcoxon test within validation contexts, particularly for determining statistical differences between errors generated by various EPS models, has been comprehensively examined by Ruland (1980). For analytical purposes, p-values are organized into separate tables corresponding to individual quarters (one through four) and an aggregate table encompassing all quarters collectively.

$$H_o : AAPEs \text{ of a pair of models are the same}. \tag{7}$$

The established significance threshold of 0.05 serves as the determinant for null hypothesis rejection. When test p-values registered below this predetermined threshold, the corresponding null hypothesis was deemed invalid – a widely accepted statistical principle substantiated by multiple sources, including the work of Ruland (1980).

3.5. Estimation of ARD-type models by XGBoost (XGB) – a gradient-boosting decision tree

XGBoost, an abbreviation for eXtreme Gradient Boosting, emerged as a significant upgrade to the original gradient boosting algorithm proposed by Chen and Guestrin (2016), renowned for its speed and effectiveness. This advanced machine learning technique, rooted in decision tree algorithms from the 1960s, is extensively used in regression and classification tasks. It constructs a prediction model consisting of an ensemble of weak prediction models, typically simple decision trees. Each successive iteration of trees aims to rectify errors in predictions made by its predecessors, with each new tree designed to correct the mistakes of the previous ones. The gradient descent algorithm is then employed to iteratively adjust the weights of the weak learners. This iterative process continues until the loss function is minimised or a predefined stopping criterion is met. XGBoost incorporates various techniques to improve the performance of the gradient boosting model, including regularisation to mitigate overfitting by penalising the loss function, tree pruning to eliminate redundant branches and enhance model stability, and parallelisation to expedite the training process. XGBoost effectively captures inherent non-linearities in the data. Applying machine learning models such as eXtreme Gradient Boosting (XGBoost) to time series regression with lags of both dependent and independent variables can help address issues of non-stationarity in several ways. Unlike classical statistical models, XGBoost does not rely on the assumption that variables are stationary. It captures patterns and relationships in the data directly, adapting to shifts in trends and seasonality without requiring strict stationarity. XGBoost can model these relationships with lagged variables, effectively 'learning' from the non-stationary patterns in the training phase without needing transformations such as differencing that traditional time series models require. Additionally, XGBoost is designed to minimise prediction errors by iteratively optimising weak learners, which lessens the impact of unrelated trends in predictor variables. It focuses on iterative improvement rather than fitting a specific model structure, allowing it to build meaningful representations even if time series components are non-stationary. This approach naturally reduces the risk of capturing spurious relationships. Also, a high collinearity among lagged predictors can compound the problem of spurious regression in traditional linear models. XGBoost includes regularisation, which reduces the risk of overfitting and handles multicollinearity effectively. This is useful when dealing with many lagged variables, especially if they display trends, further minimising spurious patterns. For deeper insights, Simon's (Simon, Pochetti, 2020) book provides valuable information. This approach was applied to the same set of variables described on page 5. Implementation of XGBoost was facilitated using the xgb library in Python.

To maintain model simplicity due to the limited sample size, a set of hyperparameters was chosen for the XGBoost model. The parameter max_depth, defining the maximum depth of each tree in the ensemble, was set to 2 to prevent overfitting. The parameter n_estimators, specifying the number of boosting rounds or trees to build and controlling the number of weak learners (trees) in the ensemble, was set to 100. The learning rate parameter, eta, determining the step size at each iteration in the gradient descent algorithm, was chosen as 0.3, following the default setting. Other model parameters were set to their default values. Hyperparameters were not optimised for each company due to the model's simplicity resulting from the very small sample size. Moreover, it is commonly understood in the machine learning community, as noted by Banerjee (2020), that while hyperparameter tuning can enhance model performance, such improvements are typically marginal for simple models.

3.6. Forecasts combinations

In an examination of the existing literature, Timmermann (2006) highlighted that estimating individual regressions for each predictor and subsequently employing forecast combination methods offers greater resilience to model misspecification and measurement error compared to estimating a single forecasting model encompassing all predictors. Forecast combinations have experienced significant growth within the forecasting community and have increasingly become integral to mainstream forecasting research and practices. The amalgamation of multiple forecasts generated for a target time series is now widely adopted to enhance accuracy by amalgamating insights from diverse sources, thereby eliminating the necessity to pinpoint a sole 'best' forecast. These combination strategies have advanced from rudimentary methods lacking estimation to sophisticated techniques incorporating time-varying weights, nonlinear combinations, inter-component correlations, and cross-learning. Hyndman et al. (2023) furnished an updated review of the extensive literature on forecast combinations and a reference to available open-source software implementations. The subsequent techniques for forecast combination are contemplated in this paper:

The simple average of forecast – 'average'

The collective forecast is determined by taking the straightforward average of the individual ARDL forecasts and is represented as:

$$\widehat{EPS}_t = \frac{1}{n}\sum_{i=1}^{n}\widehat{EPS}_t^i. \tag{8}$$

The weighted average of inverse errors – 'maape'

The collective forecast is determined by computing a weighted average of individual ARDL forecasts where the weights are inversely proportional to the errors incurred by these models. Thus,

$$\widehat{EPS}_t = \sum_{i=1}^{n} w^i\,\widehat{EPS}_t^i, \text{ where } w^i = \frac{1/MAAPE^i}{\sum_{j=1}^{n} 1/MAAPE^j} \tag{9}$$

and $MAAPE^i$ represents the Mean Arctangent Absolute Error calculated for the forecast of the *i*-th ADRL model, based on the training sample.

The 'optimal forecast combination' estimated by XGBoost – 'xgb'

The optimal forecast combination is constructed using the gradient-boosting decision tree method expressed as the following function:

$$\widehat{EPS}_t = g\left(\widehat{EPS}_t^1, \widehat{EPS}_t^3, \ldots, \widehat{EPS}_t^n\right). \tag{10}$$

4. Results

4.1. Empirical findings

Table 1 displays the MAAPE forecast errors for each quarter of 2019 across all examined models. The table is visualised in Figure 1. Notably, the seasonal random walk (SRW) model consistently demonstrates superior performance compared to other models. Furthermore, the outcomes of these two methods are statistically indistinguishable in these cases, as confirmed by the Wilcoxon nonparametric test. The p-values of these tests in the aforementioned cases are close to one, considerably higher than the significance level of 0.05. This outcome is inconsistent with the results obtained by Bansal, Nasseh, and Strauss (2015) and Ball and Ghysels (2017) which suggest the superiority of ARDL models.

The statistical significance analysis of error differentials between the foremost performing Seasonal Random Walk (SRW) methodology and the XGB estimation approach is documented in Table 2, presenting Wilcoxon test p-values across various forecast combination methodologies. The empirical evidence demonstrates consistent statistical significance in the superior performance of the SRW model, manifesting lower error metrics compared to alternative models throughout the examined temporal periods. An exception occurs in the second quarter (Q2), wherein statistical tests fail to establish significant differentiation between the error distributions of the SRW model and competing methodologies. The analytical findings permit the scholarly conclusion that the performance superiority of the SRW framework is not merely nominal but statistically verified when juxtaposed with all examined Autoregressive Distributed Lag (ARDL) models employing forecast combination techniques. Table 2 thus substantiates the empirical advantage of the SRW approach through rigorous statistical evaluation of comparative error metrics.

Table 1. Summary statistics on forecast errors for 2019 quarters

Forecast combination method	**Estimator**	**Q1 MAAPE**	**Q2 MAAPE**	**Q3 MAAPE**	**Q4 MAAPE**	**Total MAAPE**
	SRW	**0.658**	**0.702**	**0.653**	**0.736**	**0.687**
average	XGB	0.735	0.735	0.748	0.819	0.759
maape	XGB	0.740	0.732	0.755	0.807	0.758
xgb	XGB	0.778	0.749	0.776	0.814	0.779

Source: own calculations

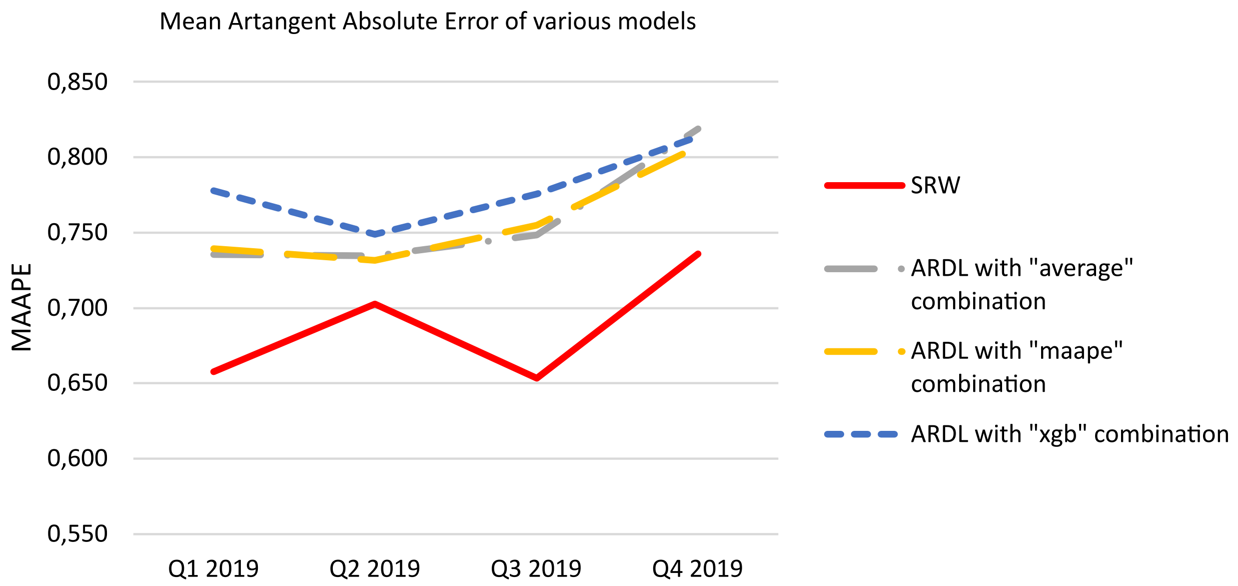

Figure 1. MAAPE of various models in respective quarters of 2019
Source: own calculations

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

	Model	Forecast combination method		
		average	**maape**	**xgb**
Q1	SRW	0.006	0.003	0.000
Q2	SRW	0.205	0.261	0.089
Q3	SRW	0.009	0.005	0.001
Q4	SRW	0.007	0.018	0.016
ALL	SRW	0.000	0.000	0.000

Source: own calculations

4.2. Robustness checks

Evaluations of methodological resilience were undertaken spanning diverse temporal periods and employing an array of conventional precision indicators.

The temporal analysis conducted across the 2017–2019 triennium revealed the systematic predominance of the SRW (seasonal random walk) framework relative to alternative ARDL configurations, irrespective of the forecast amalgamation techniques employed, as evidenced in Table 3. Furthermore, comparative statistical analysis utilizing the Wilcoxon procedure was implemented to juxtapose all modeling paradigms against the SRW benchmark, with resultant probability coefficients for respective annual periods delineated in Table 4. The SRW methodology demonstrated statistically significant performance advantages in relation to competing approaches throughout the entirety of the examined chronological sequence. Thus, the empirical evidence substantiates the consistent superiority of the seasonal random walk model throughout the investigated temporal interval.

Table 3. Summary statistics on forecast errors for SRW and XGB models in the years 2017–2019

		2017	2018	2019
		MAAPE	**MAAPE**	**MAAPE**
SRW		0.686	0.711	0.687
Forecast combination method	average	0.734	0.762	0.759
	maape	0.735	0.762	0.758
	xgb	0.758	0.785	0.779

Source: own calculations

Table 4. P-values of paired Wilcoxon test of forecast errors for SRW and XGB models in the years 2017–2019

Year	Model	Forecast combination method		
		average	**maape**	**xgb**
2017	SRW	0.001	0.000	0.000
2018	SRW	0.000	0.000	0.000
2019	SRW	0.000	0.000	0.000

Source: own calculations

An assessment of model efficacy employing Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for the consolidated quarters of 2019 is displayed in Table 5. These metrics underwent Consumer Price Index (CPI) inflation adjustments to ensure temporal equivalence between current nominal errors and the present value of future errors. The seasonal random walk model demonstrated superior performance across both RMSE and MAE indicators, corroborating earlier 2019 findings. Statistical significance in the performance differential between the SRW model and other models utilizing 'optimal' forecast combinations is substantiated by p-values extracted from Table 6 derived from the Wilcoxon test. Notably, less complex forecast aggregation methodologies, specifically the 'average' and 'maape' approaches applied to Autoregressive Distributed Lag models estimated via XGB technique, produced errors that were not statistically distinguishable. The analysis indicates that, when compared with all other models implemented through Autoregressive Distributed Lag frameworks estimated using gradient-boosting decision tree (XGB) methodologies, the seasonal random walk (SRW) model generates forecasts exhibiting both enhanced performance and statistical significance for the majority of forecast combination techniques with respect to both RMSE and MAE metrics.

Table 5. Summary statistics on forecast errors for RMSE and MAPE for SRW and XGB models in all quarters of 2019

Metric	SRW	Forecast combination method		
		average	maape	xgb
RMSE	0.937	1.082	1.093	1.061
MAE	0.705	0.869	0.874	0.801

Source: own calculations

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

Metric	Model	Forecast combination method		
		average	maape	xgb
RMSE	SRW	0.157	0.153	0.000
MAE	SRW	0.062	0.079	0.000

Source: own calculations

The conclusions drawn from this study are consistent with existing research. It is established in the literature that combining the weighted averages of two competing forecasts could potentially decrease mean squared prediction errors (MSE), but it may also introduce specific inefficiencies. Notably, Bentancor, Hardy, and Pincheira-Brown (2023) demonstrated that the act of averaging forecasts could lead to inefficiencies in the combination process, even when individual forecasts were efficient. Moreover, their findings suggest that the commonly termed 'optimal weighted average' as described in the literature might also be inefficient.

5. Discussion

The exceptional efficacy of rudimentary seasonal random walk methodologies within Polish market forecasting may be attributed to the distinctive characteristics of this economic environment, wherein intricate modeling approaches frequently introduce excessive parameterization when describing relatively uncomplicated economic trajectories. This analytical insight corresponds with scholarly contributions by Kurylek (2023a; 2023b), whose research established that conventional methodological frameworks – including ARIMA and exponential smoothing techniques that demonstrate considerable utility in American markets – were outperformed by elementary seasonal random walk approaches when applied to Polish economic data. Such empirical evidence substantiates the hypothesis that fundamental structural simplicity within Polish financial markets potentially explains the seasonal random walk (SRW) model's remarkable performance, alternatively indicating the requirement for supplementary calibration procedures when generating out-of-sample projections.

The implementation of sophisticated analytical frameworks beyond basic seasonal random walk models appears impractical for earnings per share (EPS) forecasting within Polish investment contexts. Considering that EPS patterns conform to seasonal random walk characteristics, and acknowledging that equity valuations result from EPS multiplication by price-to-earnings (P/E) ratios, logical deduction suggests stock prices manifest randomness at least equivalent to that observed in EPS behavior. Given that EPS patterns–characterized by random walk properties – present substantial forecasting difficulties, the accurate prediction of equity prices beyond quarterly horizons represents an exceptionally formidable analytical challenge.

For temporal intervals shorter than quarterly periods, when EPS remains static, equity price forecasting primarily reflects P/E multiple dynamics. Consequently, investigating methodologies for P/E multiple prediction within sub-quarterly periods – specifically between consecutive financial disclosure events – presents potentially valuable research opportunities from an investment analytics perspective. Forecasts generated through seasonal random walk (SRW) methodologies essentially replicate values from corresponding quarters in previous annual cycles. This observation indicates that P/E multiples may constitute more significant determinants of future price movements than projected annual corporate earnings (EPS), even across extended forecasting horizons.

This analytical framework aligns with established economic theory, which postulates that P/E multiples reflect anticipated earnings growth trajectories, future interest rate environments, and market risk premiums indicative of investor risk preferences (market sentiment), whereas EPS projections exclusively address near-term earnings expectations. The conclusive assessment remains consistent across both abbreviated and extended analytical timeframes: P/E multiple dynamics demonstrate greater relevance for investment analysis than EPS predictive modeling.

6. Conclusions

This study evaluates the predictive efficacy of five models: the seasonal random walk (SRW) and four models ('average,' 'maape,' and 'xgb') amalgamating forecasts generated by Autoregressive Distributed Lag (ARDL) models estimated via extreme gradient boosting. ARDL models serve as an intermediary between univariate time series and multivariate regression models, leveraging a multivariate approach to discern individual patterns for each company. The forecasting of EPS bears significance in emerging markets, where financial analysts' coverage of listed companies is scant, as exemplified by the case of Poland. Analysing quarterly EPS data from 267 Polish companies spanning the period from 2010 to 2019, it can be seen that the SRW model consistently manifests the lowest error rates, providing a more precise depiction of the Polish market than other models. Furthermore, the SRW model consistently outperforms alternative models across diverse periods and error metrics such as RMSE or MAE. This trend is substantiated by Wilcoxon tests and can be ascribed to the relatively simplistic nature of the Polish stock market. This is inconsistent with the results of research previously seen in the literature and obtained by Bansal, Nasseh, and Strauss (2015) and Ball and Ghysels (2017) which suggest the superiority of ARDL models.

This research indicates that utilizing methodologies of greater complexity than the conventional seasonal random walk for earnings per share (EPS) forecasting within the Polish market demonstrates limited practical value. It should be noted, however, that reliance on the seasonal random walk methodology for EPS modeling suggests potential significant randomness in projected equity valuations, thus creating forecasting difficulties. As a consequence, the prediction of price-to-earnings (P/E) ratios may hold greater forecasting relevance than EPS projections for future stock price determination, particularly during brief investment periods when EPS remains static.

Subsequent research opportunities include examining the relationship between forecasting efficiency and corporate magnitude, while sectoral analysis could inform the optimal model selection for EPS prediction. Significant insights might emerge from investigating time series transformations aimed at normalizing EPS statistical distributions. A promising research direction involves comparing the predictive accuracy of superior algorithmic models against market analyst forecasts. Moreover, the incorporation of textual analysis from corporate public disclosures represents an interesting research possibility. The assessment of various predictive methodologies and financial analyst projections during economic contractions, exemplified by the 2008–2009 financial crisis or the COVID-19 pandemic, could provide valuable knowledge. The identification of seasonal patterns through the seasonal random walk (SRW) model might contribute to investment strategy development, potentially questioning the validity of the 'weak form' of the Efficient Market Hypothesis (EMH).

References

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

Ball R., Watts R. (1972), *Some Time Series Properties of Accounting Income*, "The Journal of Finance", vol. 27(3), pp. 663–681, http://doi.org/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 [accessed: 14.02.2024].

Bansal N., Nasseh A., Strauss J. (2015), *Can we consistently forecast a firm's earnings? Using combination forecast methods to predict the EPS of Dow firms*, "Journal of Economics and Finance", vol. 39(1), pp. 1–22, https://doi.org/10.1007/s12197-012-9234-y

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.

Bentancor A., Hardy N., Pincheira-Brown P. (2023), *An Inconvenient Truth about Forecast Combinations*, "Mathematics", vol. 11(1)8, 3806, https://doi.org/10.3390/math11183806

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, http://doi.org/10.1007/s11142-012-9185-8

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

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

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, http://doi.org/10.2307/2490312

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, http://doi.org/10.1016/0165-4101(87)90017-6

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, https://doi.org/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, http://doi.org/10.1287/mnsc.33.6.725

Dou Y., Tan S., Xie D. (2023), *Comparison of machine learning and statistical methods in the field of renewable energy power generation forecasting: a mini review*, "Frontiers in Energy Research", vol. 11, 1218603, https://doi.org/10.3389/fenrg.2023.1218603

Elton E.J., Gruber M.J. (1972), *Earnings Estimates and the Accuracy of Expectational Data*, "Management Science", vol. 18(8), pp. 409–424, http://doi.org/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.

Fourkiotis K.P., Tsadiras A. (2024), *Applying Machine Learning and Statistical Forecasting Methods for Enhancing Pharmaceutical Sales Predictions*, "Forecasting", vol. 6(1), pp. 170–186, https://doi.org/10.3390/forecast6010010

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), eRAMF210164, https://doi.org/10.1590/1678-6971/eramf210164

Griffin P. (1977), *The Time-Series Behavior of Quarterly Earnings: Preliminary Evidence*, "Journal of Accounting Research", vol. 15(11), pp. 71–83, http://doi.org/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, https://doi.org/10.1016/j.bar.2018.10.002

Hyndman R., Kang Y., Li F., Wang X. (2023), *Forecast combinations: An over 50-year review*, "International Journal of Forecasting", vol. 39(4), pp. 1518–1547, https://doi.org/10.1016/j.ijforecast.2022.11.005

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

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

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, http://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*, "Journal of Banking and Financial Economics", vol. 1(19), pp. 26–43, http://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", vol. 54(6), pp. 651–672.

Kuryłek W. (2024), *Can we profit from BigTechs' time series models in predicting earnings per share? Evidence from Poland*, "Data Science in Finance and Economics", vol. 4(2), pp. 218–235, http://doi.org/10.3934/DSFE.2024008

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

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

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, http://doi.org/10.1016/j.adiac.2020.100497

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

Pope P.F., Wang P. (2014), *On the relevance of earnings components in valuation and forecasting links*, "Review of Quantitative Finance and Accounting", vol. 42, pp. 399–413, https://doi.org/10.1007/s11156-013-0347-y

Qian X.Y. (2017), *Financial Series Prediction: Comparison Between Precision of Time Series Models and Machine Learning Methods*, pp. 1–9, https://doi.org/10.48550/arXiv.1706.00948

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

Simon J., Pochetti F. (2020), *Learn Amazon SageMaker. A guide to building, training, and deploying machine learning models for developers and data scientists*, Packt, Birmingham–Mumbai.

Timmermann A. (2006), *Forecast combinations*, [in:] G. Elliott, C. Granger, A. Timmermann, *Handbook of Economic Forecasting*, vol. 1, Elsevier, Amsterdam, pp. 135–196.

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

Wilcoxon F. (1945), *Individual Comparisons by Ranking Methods*, "Biometrics", vol. 1, pp. 80–83, http://doi.org/10.2307/3001968

Czy modele autoregresyjne z rozkładem opóźnień mogą poprawić prognozowanie zysku na akcję w Polsce?

Streszczenie: Niniejszy artykuł analizuje znaczenie dokładnych prognoz zysków spółek notowanych na giełdzie dla osiągnięcia sukcesu inwestycyjnego. Podkreśla wagę tego aspektu, szczególnie na rynkach o ograniczonym pokryciu przez analityków, takich jak rynki wschodzące, do których zaliczana jest Polska. W badaniu dokonano oceny trafności prognoz generowanych przy użyciu metody autoregresyjnej z rozkładem opóźnień przy różnych metodach łączenia prognoz w porównaniu z sezonowym modelem błądzenia losowego. Modele te, stanowiące etap pośredni pomiędzy szeregami czasowymi a prognozowaniem uwzględniającym wiele zmiennych objaśniających, mają zastosowanie do danych dotyczących zysku na akcję spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie w latach 2008–2019, tj. między ostatnim kryzysem finansowym a szokiem spowodowanym pandemią. Model sezonowego błądzenia losowego osiągnął najniższe poziomy błędów na podstawie metryki średniego arcus tangensa bezwzględnego błędu procentowego. Wniosek ten jest poparty rygorystycznymi testami statystycznymi i kontrolami odporności z wykorzystaniem różnych okresów oraz wskaźników błędu. Lepszą wydajność prostszego modelu sezonowego błądzenia losowego można przypisać stosunkowo nieskomplikowanemu charakterowi polskiego rynku.

Słowa kluczowe: zysk na akcję, błądzenie losowe, model autoregresyjny z rozkładem opóźnień, XGBoost, prognozowanie finansowe, Giełda Papierów Wartościowych w Warszawie

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

Received: 2024-05-16; revised: 2024-11-03. Accepted: 2025-04-10