QUANTITATIVE METHODS IN ECONOMICS
Volume XXV, No. 3, 2024, pp. 107 – 123

IF MULTILAYER PERCEPTRON NETWORK MAY HELP IN MULTIVARIATE EPS FORECASTING. EVIDENCE FROM POLAND

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

Abstract: This study examines the role of accurate earnings forecasts for publicly traded firms in achieving investment success, focusing on markets with limited analyst coverage, such as Poland. It compares the accuracy of various models, including artificial neural networks, against a seasonal random walk model applied to EPS data from Warsaw Stock Exchange companies (2008-2019). The seasonal random walk model showed the lowest error based on MAAPE, with results confirmed by statistical tests. Simpler models may outperform complex ones due to overfitting and the relatively simple dynamics of Polish companies.

Keywords: earnings per share, seasonal random walk, artificial neural network, multilayer perceptron network, financial forecasting, Warsaw Stock Exchange

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

INTRODUCTION

The pricing of company stocks hinges on the multiplication of earnings per share (EPS) by the Price-to-Earnings multiple, a pivotal step in investment deliberations. Accurate forecasting of these components is crucial, with EPS predictions assuming particular importance. They furnish indispensable numerical insights into a company's future trajectory, furnishing valuable data on potential market valuation and guiding auditing expectations. The estimates of Earnings Per Share (EPS) by popular financial information services like Eikon (Refinitiv), Morningstar, Bloomberg, and others are consensus forecasts, which aggregate the forecasts of multiple equity analysts. These analysts, who may be part of banks,

https://doi.org/10.22630/MIBE.2024.25.3.10

brokerage firms, or independent financial research companies, provide their EPS predictions for the current year, next year, and sometimes over longer periods. While financial analysts extensively cover companies in developed markets like the US, according to the data from the EquityRT platform (which sources the data to Morningstar), only a fraction, approximately 20%, receive similar attention in emerging markets such as Poland. Consequently, there exists a compelling necessity to employ statistical or machine learning models for EPS forecasting.

This article undertakes a comparative assessment of various models, employing distinct sets of explanatory variables, utilizing the multilayer perceptron (MLP) artificial neural network, drawing insights from Li and Mohanram's [2014] research. It encompasses quarterly EPS data for 267 companies listed on the Polish stock exchange from the 2008-2009 financial crisis through the 2020 pandemic.

Rather than relying solely on the conventional mean absolute percentage error (MAPE) metric, which is prone to extreme values when the denominator is small, an alternative measure, the mean arctangent absolute percentage error (MAAPE) proposed by Kim and Kim [2016], is computed and employed in this study.

In summary, this article pursues several objectives. Firstly, it aims to assess the performance of the multilayer perceptron network (MLP) over different sets of explanatory variables in EPS prediction. Secondly, it seeks to apply diverse error metrics, varying timeframes, and a range of statistical tests to validate the outcomes of these experiments. Thirdly, it endeavors to adapt and utilize a relative performance error metric to address scenarios where actual profits approach zero, employing MAAPE as an error metric. Lastly, it strives to elucidate the practical implications of these findings for investment strategies in Polish stocks.

LITERATURE OVERVIEW

The algorithmic forecasting of Earnings per Share began in the 1960s, sparking scholarly exploration focused on autoregressive integrated moving average (ARIMA) models (see [Ball and Watts 1972, Watts 1975, Griffin 1977, Foster 1977, Brown and Rozeff 1979]). This marked the primary class of scrutinized models, with outcomes varying across investigations: while some studies supported the simplicity of the basic random walk model, suggesting that more intricate models did not consistently surpass it, others yielded divergent conclusions. Kurylek [2023a, 2023b] conducted a similar study regarding the Polish market.

Over time, however, a consensus emerged favoring ARIMA-type models for their typically precise forecasts (see [Lorek 1979, Bathke and Lorek 1984]). This consensus lasted until the late 1980s when a prevailing belief suggested that financial analysts' forecasts surpassed those generated by time series models (see [Brown et al. 1987]). Nevertheless, Conroy and Harris [1987] noted analysts' superiority in short forecast horizons, diminishing over longer periods. This perspective endured until recent years when the superiority of analysts over time series models was once again scrutinized (see [Lacina et al. 2011, Bradshaw et al. 2012, Pagach and Warr 2020, Gaio et al. 2021]).

Furthermore, since the late 1960s, researchers explored various approaches employing exponential smoothing for EPS prediction (see [Elton and Gruber 1972, Ball and Watts 1972, Johnson and Schmitt 1974, Brooks and Buckmaster 1976, Ruland 1980, Brandon et al. 1987, Jarrett, 2008]), resulting in mixed findings.

Lorek and Willinger [1996] demonstrated the superiority of multivariate cross-sectional models over firm-specific and common-structure ARIMA models. Lev and Thiagarajan [1993] identified 12 fundamental signals from financial ratios, subsequently utilized by Abarbanell and Bushee [1997] for EPS forecasting. Cao and Gan [2009], Cao and Parry [2009], Ahmadpour et al. [2015], and Ball and Ghysels [2017] employed similar fundamental variables for EPS multivariate forecasting using neural networks, affirming their effectiveness.

Ohlson [1995, 2001] formulated a residual earnings model, while Pope-Wang [2005, 2014] established theoretical frameworks linking earnings forecasts to accounting variables and stock prices. Li [2011] developed a model for forecasting earnings for loss-making firms, demonstrating its efficacy. Lev and Souginannis [2010] provided evidence of the usefulness of estimate-based accounting items for predicting next year's earnings, albeit with limited success in subsequent years. Hou et al. [2012] achieved substantial R2 coefficients in cross-sectional regression models for earnings forecasting. Li and Mohanram [2014] compared various models, revealing the superiority of some over others. Harris and Wang [2019] found Pope and Wang's [2005] model generally less biased and more accurate.

Recent research has placed significant emphasis on the utilization of artificial neural networks in EPS forecasting, yielding ambiguous results. A pioneering study by Atiya, Shaheen, and Talaat [1997] demonstrated the superiority of a neural network based on fundamental characteristics for stock price forecasting, establishing its consistent superiority. Cao et al. [2004] conducted a comparative analysis of neural feedforward networks (MLP) and found them to outperform other forecasting models, showcasing their enhanced accuracy. However, contrary findings were presented by Lai and Li [2006], who discovered that an ANN model exhibited the worst accuracy when predicting EPS. In another study, Cao and Parry [2009] consistently demonstrated the superiority of univariate neural network models over linear regression models. They further revealed that a genetic algorithm outperformed backpropagation in estimating natural network weights. Similarly, Cao and Gan [2009] confirmed the superior performance of neural network models, especially when optimized using a genetic algorithm, for predicting the EPS of Chinese listed companies. Gupta, Khirbat, and Singh [2013] identified an optimal architecture for stock market price forecasting using multiperceptron networks, highlighting the critical role of factors like EPS and public confidence in predictions. Ahmadpour, Etemadi, and Moshashaei [2015] utilized a standard multilayer perceptron (MLP) neural network with remarkable success, with extracted rules exhibiting significantly greater accuracy than pure MLP models. Chen et al. [2020]

explored various methods for EPS prediction, including decision trees and radial basis function networks, demonstrating the superiority of the ensemble method in terms of accuracy. Elend et al. [2020] compared LSTM networks to TCNs for predicting future EPS, with LSTM outperforming the naive persistent model, showcasing a significant improvement in prediction accuracy. Additionally, Suler, Vochozka, and Vrbka [2020] successfully employed an LSTM neural network model for bankruptcy prediction in the Czech Republic. Furthermore, Xiaoqiang's [2022] article provides a concise overview of deep learning and machine learning techniques, including convolutional neural networks and decision trees, applicable to EPS forecasting. In the latest research, Dreher et al.]2024] illustrated that considering accounting information on tax loss carryforwards did not enhance EPS forecasts and often deteriorated predictions in out-of-sample tests for German listed companies, utilizing tax footnotes information.

DATA & METHODOLOGY

Data

The Polish stock market, which became part of the European Union after 2004, is characterized by its substantial depth, boasting a market capitalization that soared to $197 billion and accommodating 774 listed companies by the conclusion of 2021. However, it's notable that these stocks do not receive the extensive analyst coverage observed in the United States or Western Europe. In 2019, merely around 20% of the 711 listed companies garnered attention from analysts. This highlights the imperative for statistical forecasting of crucial financial data utilizing analytical methodologies. This article primarily concentrates on the earnings per share (EPS) data series and other financial explanatory variables obtained from EquityRT, a financial analysis platform. The analysis probes into EPS patterns of companies listed on the Warsaw Stock Exchange from Q1 2010 to Q4 2019, i.e. between significant structural shifts: the 2008-2009 financial crisis and the onset of the COVID-19 pandemic in 2020. For forecasting purposes, data from Q1 2010 to Q4 2018 (36 quarters) are utilized for model estimations, while Q1 2019 to Q4 2019 data are set aside for out-of-time validation testing. Forecast horizons extend from 1 to 4 quarters ahead, with additional examination incorporating the years 2017 and 2018 as validation samples. Following thorough coverage, excluding splits and reverse splits, the dataset retains 267 companies.

Explanatory variables

Redundancy among financial ratios, frequently encountered when elucidating economic phenomena, demands approaches to refine and isolate a distinct, autonomous subset of vital financial variables applicable for EPS modeling. The ensuing models were utilized:

- The seasonal random walk model (SRW)

This process can be described as:

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

The forecast denoted as, $\widehat{EPS}_t = EPS_{t-4}$ relies on the value delayed by 4 quarters as the prediction, thereby obviating the necessity for parameter estimation. This model acts as a benchmark, based on research by Kuryłek [2023a, 2023b], demonstrating its superiority over time series models in the context of Poland.

- The Laurent model (L)

Laurent [1979] condensed a set of 45 financial ratios into 10 factors using principal component analysis, collectively explaining nearly 90% of the observed variance. Through a meticulous selection process, ten specific financial ratios were identified to represent these factors. These ratios offer a comprehensive insight into a firm's financial performance from various angles. Selection criteria included a strong correlation with the represented factor, independence from each other, and widespread acceptance in usage. The resulting set of ratios covers diverse aspects, such as 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 EPS. Consequently, the estimated equation can be formulated as follows:

$$EPS_{t+4} = f(R1_t, R2_t, .., R10_t) + \varepsilon_t \quad (2)$$

- The Lev and Thiagarajan model (LT)

The research conducted by Lev and Thiagarajan [1993] identified 12 fundamental signals extracted from various practical texts dedicated to utilizing financial ratios for delineating different facets of a firm’s performance. This approach stands as one of the most frequently referenced methodologies in the literature, as evidenced by Ahmadpour et al. [2015], Ball and Ghysels [2017], and others. Key variables that proved useful include 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). However, it's important to note that some variables that were mentioned in the initial research are missing in Poland. Polish listed companies do not disclose data on R&D expenditures and order backlog. Furthermore, the database lacks information on the chosen FIFO or LIFO accounting standard, and it is impossible to discern auditor qualifications. Therefore, the equation akin to Lev and Thiagarajan's, structured in quarterly terms for making one-year-ahead predictions, can be articulated as follows:

$$E_{t+4} = f(I_t, AR_t, CAPEX_t, GM_t, SAE_t, PROV_t, ET_t, LP_t) + \varepsilon_t \quad (3)$$

- The Residual Income model (RI)

Ohlson [1995, 2001], Olhson and Juettner-Nauroth [2005] developed a theoretical model of residual earnings grounded on the clean surplus assumption. This assumption posits that the market value of equity equals the book value plus the net present value of future abnormal returns. Residual income, or abnormal earnings, is quantified as the disparity between actual earnings and book value, multiplied by the cost of capital, and is presumed to follow an autoregressive stochastic process. As a result, future earnings are influenced by current earnings, book value, dividends paid, and the cost of capital, leading to the incorporation of the following accounting variables: Delayed earnings (E), a Dummy variable for firms with delayed negative earnings (NegE), a Polynomial term combining two preceding variables (NegE * E), Book value (BV), and Accruals (ACC). The specification of this model is provided by the research conducted by Li and Mohanram [2014]. With quarterly data and the application of this approach, the equation assumes the following structure:

$$E_{t+4} = f(E_t, NegE_t, BV_t, ACC_t) + \varepsilon_t \quad (4)$$

- The Pope and Wang model (PW)

Based on the assumptions of linear valuation, no-arbitrage, dividend irrelevance, and clean surplus accounting, Pope and Wang [2005, 2014] expounded upon the theoretical linkage between earnings forecasts and six observable accounting variables. These variables include Earnings (E), Book value (BV) lagged by one and two years, Accruals (ACC), and non-accounting variables such as Stock price (P) lagged by one and two years. The model's equation formulation was detailed by Harris and Wang (2019). For quarterly data aiming to predict earnings one year ahead, the expression is presented as follows:

$$E_{t+4} = f(E_t, BV_t, BV_{t-4}, ACC_t, P_t, P_{t-4}) + \varepsilon_t \quad (5)$$

- The Earnings Persistence model (EP)

Li [2011] formulated the earnings persistence model to account for earnings growth, integrating explanatory variables including delayed Earnings (E), the indicator for negative earnings (NegE), and the interaction term between NegE and earnings (NegE*E). This facilitates the discernment of profit and loss persistence. The equation specification for this model is outlined in the research by Li and Mohanram (2014). It can be represented in the following concise functional form:

$$E_{t+4} = f(E_t, NegE_t) + \varepsilon_t \quad (6)$$

- The Hou, van Dijk and Zhang model (HDZ)

Hou, van Dijk, and Zhang [2012] performed a cross-sectional regression analysis aimed at predicting earnings. They utilized data spanning the previous decade and included variables such as delayed Earnings (E), Total assets (A), Dividend payout (D), Accruals (ACC), a Dummy variable for dividend payers (DD), and a Dummy variable for firms with delayed negative earnings (NegE). Their efforts resulted in a substantial R2 coefficient of around 0.8. The equation introduced by Li and Mohanram (2014) is applied in the following manner:

$$E_{t+4} = f(E_t, NegE_t, ACC_t, A_t, D_t, DD_t) + \varepsilon_t \quad (7)$$

To forecast earnings per share (EPS), one divides future earnings by the constant number of shares. Because in the Pope Wang models 4 and 8 lags are used, 7 476 observations (28 quarters x 267 companies) observations are used for training the artificial network in 2019.

The multilayer perceptron network (MLP)

The artificial networks presented in this study were trained using the TensorFlow module in Python. These networks are of the feedforward type, meaning that data flows uni-directionally from the input layer to the output layer. Artificial neural networks (ANNs) are commonly employed for analyzing cause-effect relationships within complex systems, often in the context of big data frameworks. However, they can also be applied to small datasets, as demonstrated in fields like health sciences by Pasini [2015], as is the case in this article. Hyperparameters, such as the width and depth of networks (i.e., the number of neurons in each layer and the number of layers), were optimized using the hyperas library in Python. The models were trained using the backpropagation algorithm based on gradient descent, employing only 20 epochs (where one epoch constitutes one complete run-through of the training set). Backpropagation, popularized by Werbos [1988] in the late 80s, is a standard method for learning neural networks, involving the backward propagation of errors. It fine-tunes the weights of a neural network based on the error rate obtained in the previous epoch. The error, calculated as the difference between the predicted and actual output, is then fed back through the network. The weights are adjusted accordingly to minimize the error, with the learning rate determining the rate of adjustment. Proper tuning of the weights aims to reduce error rates and enhance the model's generalization ability. After a certain number of epochs, the algorithm converges to a state where there's minimal change in loss over subsequent epochs, typically reaching a local optimum of the defined loss function. Input parameters are usually standardized for ANNs when dealing with multivariate data. In all analyzed models, the hyperbolic tangent (tanh) activation function, a popular choice, was used in all layers. Additionally, the weights between layers were initialized using the glorot_uniform initializer, proposed by Bengio and Glorot [2010], which generates initial weight values from a uniform distribution. Further insights into different network architectures and parameters can be found in the book by Bengio et al. [2017].

A multilayer perceptron (MLP) represents a form of artificial neural network structured with multiple layers of interconnected nodes. Each layer's nodes establish connections with those in the subsequent layer, with the connection weights learned during training. Typically, an MLP comprises three layers or more: an input layer, one or more hidden layers, and an output layer. Within each hidden layer, the output of each node results from a weighted sum of the preceding layer's node outputs, augmented by a bias term. The inception of MLP neural networks dates back to 1958 when Rosenblatt [1958] introduced a layered network of perceptrons. It featured an input layer, a hidden layer with fixed weights that didn't adapt, and an output layer with learning connections. Rosenblatt drew inspiration from the brain's functionality. The number of layers and neurons constitutes the network's hyperparameters, subject to fine-tuning. While deeper neural networks excel in data processing, excessively deep layers can engender challenges like vanishing gradients and overfitting. Empirical rules of thumb guide the determination of the optimal number and size of hidden layers, as detailed in Heaton's [2008] book. According to this source, a single hidden layer suffices to approximate any function. Consequently, the network in this study was designed with one hidden layer. Additionally, a widely endorsed guideline suggests that the hidden layer's optimal size should lie approximately between that of the input and output layers. In this instance, the hidden layer's size equates to the mean of the sizes of the input and output layers.

Mean Arctangent Absolute Percentage Error (MAAPE)

Denoting $A_1^i, \dots, A_4^i$, as the actual earnings per share (EPS) from the first to the fourth quarter of 2019 for a specific firm I, and $F_1^i, \dots, F_4^i$ as the corresponding forecasts (i.e. $\hat{Q}_t$, where t=37,..,40 for i-th company), the absolute percentage error (APE) of such prediction during the j-th quarter of 2019, for any firm i, can be expressed 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: it may result in infinite or undefined values when the actual figures approach or reach zero, a situation frequently encountered in earnings forecasts. Moreover, extremely low actual figures, typically below one, can result in substantial percentage errors (outliers). Furthermore, when actual values are zero, APE becomes infinite. To address this issue, Kim and Kim (2016) proposed the arctangent absolute percentage error as a novel solution in the domain.

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

This reasoning 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

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

The decision to opt for MAAPE over MAPE (Mean Absolute Percentage Error) was intentional because of the inclusion of companies with actual profits very close to zero in the analyzed sample. When only one observation approaches zero while others are substantially distant from it, the MAPE of this specific observation can become extremely large (almost infinite), potentially dominating the mean calculation and making the rest of observations negligible.

The statistical test

To assess the statistical significance of MAAPE variations among multiple models, a nonparametric two-sided Wilcoxon test, as detailed by Wilcoxon (1945), is employed. This test serves as a paired difference test for two matched samples. It's important to highlight that this test does not require specific assumptions regarding a probability distribution, except for the symmetry of the difference in scores and the independence of these differences. Ruland (1980) extensively explained the application of the Wilcoxon test in validation, particularly in determining whether errors generated by different EPS models display statistical differences. Separate tables containing p-values are generated for each quarter, ranging from one to four, as well as for all quarters collectively.

$$H_o: medians\ of\ AAPEs\ of\ a\ pair\ of\ models\ are\ the\ same \quad (11)$$

If the p-values of each test fell below the predetermined significance threshold of 0.05, the null hypothesis for each test was considered invalid. This principle, widely utilized, draws from various sources, including Ruland (1980).

RESULTS

Empirical findings

The seasonal random walk (SRW) model, as outlined in Table 1, consistently outperforms all other models estimated within the multilayer perceptron (MLP) framework across every quarter, demonstrating superior overall performance. Conversely, the Residual Income model (RI) displays the poorest performance, while the Laurent model (L) achieves the second-best performance. The MAAPEs of all other models fall in between, at comparable levels.

To assess whether the errors of the top-performing model differ significantly from those of the other models, the Wilcoxon nonparametric test was utilized to compare the AAPE medians between the SRW model and all other models. As depicted in Table 2, the findings reveal that the seasonal random walk (SRW) model consistently exhibits statistically lower errors compared to the other models across all analyzed periods. However, in the fourth quarter of 2019, the significance levels of the Laurent model (L), the Lev and Thiagarajan model (LT), and the Pope Wang model (PW) are in proximity to the 0.05 threshold.

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
L	0.785	0.782	0.785	0.785	0.785
LT	0.785	0.786	0.788	0.791	0.787
RI	1.016	0.965	0.986	0.917	0.971
PW	0.788	0.788	0.791	0.790	0.789
EP	0.930	0.880	0.898	0.859	0.892
HDZ	0.923	0.871	0.891	0.874	0.890

Source: own calculations

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

quarter	model	L	LT	RI	PW	EP	HDZ
1	SRW	0.000	0.000	0.000	0.000	0.000	0.000
2	SRW	0.004	0.005	0.000	0.004	0.000	0.000
3	SRW	0.000	0.000	0.000	0.000	0.000	0.000
4	SRW	0.046	0.046	0.000	0.049	0.000	0.000
ALL	SRW	0.000	0.000	0.000	0.000	0.000	0.000

Source: own calculations

Robustness checks

Robustness assessments were undertaken across different years and various prevalent error metrics. Notably, across all scrutinized years—2017, 2018, and 2019—the seasonal random walk model (SRW) consistently produced superior forecasts compared to alternative models, as depicted in Table 3. In both 2017 and 2018, the least effective model was the simplest among all multivariate models, the Earnings Persistence model (EP). However, in 2019, as previously noted, the Residual Income model (RI) exhibited the highest forecast errors. Furthermore, the Wilcoxon test was employed to compare all model pairs alongside the seasonal random walk model, and the corresponding p-values for each year are delineated in Table 4. Across each of these years, the seasonal random walk model (SRW) demonstrated statistically superior outcomes compared to alternative methods. Hence, the consistent dominance of the seasonal random walk model becomes apparent over time.

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
	L	0.785	0.790	0.785
	LT	0.790	0.784	0.787
	RI	0.784	0.785	0.971
	PW	0.785	0.785	0.789
	EP	0.852	0.953	0.892
	HDZ	0.809	0.829	0.890

Source: own calculations

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

year	model	L	LT	RI	PW	EP	HDZ
2017	SRW	0.000	0.000	0.000	0.000	0.000	0.000
2018	SRW	0.000	0.001	0.001	0.000	0.000	0.000
2019	SRW	0.000	0.000	0.000	0.000	0.000	0.000

Source: own calculations

Table 5 illustrates an assessment of the performance of the examined models using alternative error metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). This assessment encompasses all quarters aggregated for the year 2019. To facilitate a fair comparison, these metrics underwent adjustment for Consumer Price Index (CPI) inflation. This adjustment ensures parity in the present value of future errors in nominal terms with current errors. Consistent with previous observations in 2019, the seasonal random walk model demonstrated the lowest errors across all metrics, encompassing both RMSE and MAE.

The p-values derived from Table 6, per the Wilcoxon test, highlight significant disparities between the outcomes of the SRW model and other model pairings. This indicates that the forecasts produced by the seasonal random walk (SRW) model, in terms of both RMSE and MAE, exhibit superior performance and statistical distinctiveness compared to all other models implemented via the multilayer perceptron (MLP) methodology.

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

	SRW	L	LT	RI	PW	EP	HDZ
RMSE	0.937	1.334	1.327	1.501	1.346	1.352	1.362
MAE	0.705	1.105	1.097	1.247	1.116	1.117	1.125

Source: own calculations

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

measure	model	L	LT	RI	PW	EP	HDZ
RMSE	SRW	0.000	0.000	0.000	0.000	0.000	0.000
MAE	SRW	0.000	0.000	0.000	0.000	0.000	0.000

Source: own calculations

Discussion

The relatively inferior performance of more intricate models employing artificial neural networks can be attributed to overfitting, which leads to unstable relationships among variables contingent on the pertinent test dataset. The utilization of such relationships in making predictions is reasonable only if the statistical relationship is sufficiently robust (see [Lev and Souginannis 2010]). This assertion is consistent with the findings of Dreher et al. [2024], who also demonstrated for German listed companies that complex deep learning approaches, which optimize explanatory power within the sample, do not fare well for out-of-sample prediction. These sophisticated models risk overparameterizing the market's straightforward behavior, resulting in larger forecast errors.

The rationale behind the superior performance of simpler models may align with the Polish scenario, as advanced models often tend to be overly intricate, possessing an excess of parameters to describe relatively straightforward economic phenomena. This observation further corroborates the research conducted by Kurylek [2023a, 2023b], which showed that even basic models like ARIMA and exponential smoothing, effective for the US market, were outperformed by the simple seasonal random walk model in Poland. This reinforces the hypothesis that the inherent simplicity of the Polish stock market likely underpins the effectiveness of the seasonal random walk (SRW) model, or alternatively, additional calibration for out-of-sample predictions might be necessary.

Hence, straightforwardly applying any of these sophisticated techniques beyond the conventional seasonal random walk in Poland for EPS forecasting in investment contexts appears impractical. Furthermore, considering that EPS behavior follows a seasonal random walk and acknowledging that stock prices are derived from the multiplication of the P/E multiple by EPS, one might infer that stock prices exhibit at least as much randomness as EPS. Since EPS behavior, characterized by a random walk, is inherently challenging, accurately predicting stock prices for a period extending at least one quarter ahead becomes even more daunting.

In shorter timeframes, when EPS remains constant, stock price forecasting behaves similarly to P/E multiples. Consequently, exploring methods to forecast P/E multiples for periods shorter than one quarter, occurring between the publication of quarterly financial reports, could be of significant interest from an investment perspective. The forecast generated by the seasonal random walk (SRW) essentially represents a value from the corresponding quarter of the previous year. This implies that for predicting future prices, even over extended horizons, the P/E multiple might carry more significance than next year's earnings of companies (EPS).

This aligns with economic theory, which suggests that the P/E multiple is influenced by expected future earnings growth, future interest rates, and market sentiment or premium reflecting investor risk appetite (i.e., market sentiment), whereas EPS forecasts pertain only to near-future earnings. In both short-term and long-term contexts, the conclusion is clear: for investment, the P/E multiple holds greater importance than EPS prediction.

Conclusions

This study assesses the predictive performance of seven models: the seasonal random walk (SRW), the Laurent model (L), the Lev and Thiagarajan model (LT), the Residual Income model (RI), the Pope and Wang model (PW), the Earnings Persistence model (EP), and the Hou, van Dijk, and Zhang model (HDZ). The forecasting of EPS holds significant value in emerging markets, where coverage of listed firms by financial analysts is sparse, as evidenced by Poland's case. When applied to quarterly EPS data from 267 Polish companies spanning 2010 to 2019, the SRW model consistently demonstrated the lowest error, offering a more accurate portrayal of the Polish market compared to other models. Furthermore, the SRW model consistently surpassed other models across different periods and error metrics like RMSE or MAE. This trend is supported by Wilcoxon tests and can be attributed to the over-parameterization of complex models, their tendency to overfit, and the relatively simple nature of the Polish stock market.

The practical implication of this research suggests that utilizing techniques more sophisticated than the standard seasonal random walk for EPS forecasting in Poland lacks practical merit. However, relying on the seasonal random walk for EPS modeling implies that forecasted stock prices may exhibit significant randomness, posing challenges for prediction. Hence, forecasting the P/E multiple might be more critical than predicting EPS for future stock price forecasts, especially in shorter investment horizons when EPS remains constant.

Future research could explore the relationship between forecasting accuracy and firm size, with industry sector analysis potentially influencing the choice of the most suitable model for EPS forecasting. Investigating time series transformations to normalize EPS distributions could offer valuable insights. Additionally, a broader set of explanatory variables warrants exploration. Comparing the predictive accuracy of the best algorithmic model with forecasts from market analysts presents an intriguing avenue. Furthermore, evaluating the performance and accuracy of various predictive models and financial analysts' projections during economic downturns, such as the 2008-2009 financial crisis or the COVID-19 pandemic, could yield valuable insights. Identifying seasonal patterns through the SRW model may offer insights into investment strategies, potentially challenging the "weak form" of the Efficient Market Hypothesis (EMH).

REFERENCES

Abarbanell J., Bushee B. (1997) Fundamental Analysis, Future EPS, and Stock Prices. Journal of Accounting Research, 35(1), 1-24.

Ahmadpour A., Etemadi H., Moshashaei S. (2015) Earnings per Share Forecast using Extracted Rules from Trained Neural Network by Genetic Algorithm. Computational Economics, 46(1), 55-63.

Atiya A., Shaheen S., Talaat N. (1997) An Efficient Stock Market Forecasting Model using Neural Networks. IEEE International Conference on Neural Networks - Conference Proceedings.

Ball R., Ghysels E. (2017) Automated Earnings Forecasts: Beat Analysts or Combine and Conquer? Management Science, 64(10), 4936-4952.

Ball R. Watts R. (1972) Some Time Series Properties of Accounting Income. The Journal of Finance, 27(3), 663-681.

Banerjee P. (2020) A Guide on XGBoost Hyperparameters Tuning, Accessed February 14, 2024. https://www.kaggle.com/code/prashant111/a-guide-on-xgboost-hyperparameters-tuning.

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

Bradshaw M., Drake M., Myers J., Myers L. (2012) A Re-Examination of Analysts' Superiority over Time-Series Forecasts of Annual Earnings. Review of Accounting Studies, 17(4), 944-968.

Bengio Y., Courville A., Goodfellow I. (2017) Deep Learning. Cambridge, Massachusetts: The MIT Press.

Bengio Y., Glorot X. (2010) Understanding the Difficulty of Training Deep Feedforward Neural Networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 9, 249-256.

Brandon Ch., Jarrett J. E., Khumawala S. B. (1987) A Comparative Study of the Forecasting Accuracy of Holt-Winters and Economic Indicator Models of Earnings Per Share For Financial Decision Making. Managerial Finance, 13(2), 10-15.

Brooks L. D., Buckmaster D. A. (1976) Further Evidence of the Time Series Properties of Accounting Income. The Journal of Finance, 31(5), 1359-1373.

Brown L. D., Griffin P. A., Hagerman R. L., Zmijewski M. E. (1987) Security Analyst Superiority Relative to Univariate Time-Series Models in Forecasting Quarterly Earnings. Journal of Accounting and Economics, 9(1), 61-87.

Brown L. D., Rozeff M. S. (1979) Univariate Time-Series Models of Quarterly Accounting Earnings per Share: A Proposed Model. Journal of Accounting Research, 17(1), 179-189.

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

Cao Q., Parry M. (2009) Neural Network Earnings per Share Forecasting Models: A Comparison of Backward Propagation and the Genetic Algorithm. Decision Support Systems, 47 (1), 32-41.

Cao Q., Schniederjans M. J., Zhang W. (2004) Neural Network Earnings per Share Forecasting Models: A Comparative Analysis of Alternative Methods. Decision Sciences, 35 (2), 205-237.

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

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, 785-794.

Conroy R., Harris R. (1987) Consensus Forecasts of Corporate Earnings: Analysts' Forecasts and Time Series Methods. Management Science, 33(6), 725-738.

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, 94(1), 1-39.

Elend L., Kramer O., Lopatta K., Tideman S. (2020) Earnings prediction with deep learning. German Conference on Artificial Intelligence (Künstliche Intelligenz), KI 2020: Advances in Artificial Intelligence, 267-274.

Elton E. J., Gruber M. J. (1972) Earnings Estimates and the Accuracy of Expectational Data, Management Science, 18(8), B409-B424.

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

Gaio L., Gatsios R, Lima F., Piamenta Jr. T. (2021) Re-Examining Analyst Superiority in Forecasting Results of Publicly-Traded Brazilian Companies. Revista de Administracao Mackenzie, 22(1), eRAMF210164.

Griffin P. (1977) The Time-Series Behavior of Quarterly Earnings: Preliminary Evidence. Journal of Accounting Research, 15(1), 71-83.

Gupta R., Khirbat G., Singh S. (2013) Optimal Neural Network Architecture for Stock Market Forecasting. Proceedings - 2013 International Conference on Communication Systems and Network Technologies, CSNT 2013, 557-561.

Harris R. D. F., Wang P. (2019) Model-Based Earnings Forecasts vs. Financial Analysts' Earnings Forecasts. British Accounting Review, 51(4), 424-437.

Heaton J. (2008) Introduction to Neural Networks for Java, 2nd Edition. Heaton Research Inc.

Hou K., van Dijk M., Zhang Y. (2012) The Implied Cost of Capital: A New Approach. Journal of Accounting and Economics, 53(3), 504-526.

Jarrett J. E. (2008) Evaluating Methods for Forecasting Earnings per Share. Managerial Finance, 16, 30-35.

Johnson T. E., Schmitt T. G. (1974) Effectiveness of Earnings per Share Forecasts. Financial Management, 3(2), 64-72.

Kim S., Kim H. (2016) A New Metric of Absolute Percentage Error for Intermittent Demand Forecasts. International Journal of Forecasting, 32(3), 669-679.

Kuryłek W. (2023a) The Modeling of Earnings per Share of Polish Companies for the Post-Financial Crisis Period using Random Walk and ARIMA Models. Journal of Banking and Financial Economics, 1(19), 26-43.

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

Lacina M., Lee B., Xu R. (2011) An Evaluation of Financial Analysts and Naïve Methods in Forecasting Long-Term Earnings. [In:] K. D Lawrence, R. K. Klimberg (Eds.), Advances in business and management forecasting (pp. 77-101), Bingley, UK, Emerald.

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

Laurent C. (1979) Improving the Efficiency and Effectiveness of Financial Ratio Analysis. Journal of Business Finance & Accounting, 6(3), 401-413.

Lev B., Souginannis T. (2010) The Usefulness of Accounting Estimates for Predicting Cash Flows and Earnings. Review of Accounting Studies, 15(4), 779-807.

Lev B., Thiagarajan S. (1993) Fundamental Information Analysis. Journal of Accounting Research, 31(2), 190-215.

Li K. K. (2011) How Well Do Investors Understand Loss Persistence? Review of Accounting Studies, 16(3), 630-667.

Li K. K., Mohanram P. (2014) Evaluating Cross-Sectional Forecasting Models for the Implied Cost of Capital. Review of Accounting Studies, 19(3), 1152-1185.

Lorek K. S. (1979) Predicting Annual Net Earnings with Quarterly Earnings Time-Series Models, Journal of Accounting Research, 17(1), 190-204.

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

Ohlson J. A. (1995) Earnings, Book Values, and Dividends in Equity Valuation. Contemporary Accounting Research, 11(2), 661-687.

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

Ohlson J. A., Juettner-Nauroth B. E. (2005) Expected EPS and EPS Growth as Determinants of Value. Review of Accounting Studies, 10(2-3), 349-365.

Pagach D. P., Warr R. S. (2020) Analysts Versus Time-Series Forecasts of Quarterly Earnings: A Maintained Hypothesis Revisited. Advances in Accounting, 51, 1-15.

Pasini A. (2015) Artificial Neural Networks for Small Dataset Analysis. Journal of Thoracic Disease, 7(5), 953-960.

Pope P. F., Wang P. (2005) Earnings Components, Accounting Bias and Equity Valuation. Review of Accounting Studies. 10(4), 387-407.

Pope P., Wang P. (2014) On the Relevance of Earnings Components: Valuation and Forecasting Links. Review of Quantitative Finance and Accounting. 42, 399-413.

Rosenblatt F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review, 65(6), 386-408.

Ruland W. (1980) On the Choice of Simple Extrapolative Model Forecasts of Annual Earnings. Financial Management, 9(2), 30-37.

Suler P., Vochozka M., Vrbka J. (2020) Bankruptcy or Success? The Effective Prediction of a Company's Financial Development using LSTM. Sustainability, 12(18), 2299-2314. https://doi.org/10.3390/su12187529.

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

Werbos P. (1988) Backpropagation: Past and Future. IEEE International Conference on Neural Networks, 343-353.

Wilcoxon F. (1945) Individual Comparisons by Ranking Methods. Biometrics, 1, 80-83.

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). Accessed February 14, 2024. https://doi.org/10.1109/icetci55101.2022.9832404.