Bank i Kredyt 54(6), 2023, 651-672

BANK I KREDYT
BANK & CREDIT

Can exponential smoothing do better than seasonal random walk for earnings per share forecasting in Poland?

Wojciech Kuryłek*

Submitted: 7 July 2023. Accepted: 12 September 2023.

Abstract

The accurate prediction of listed companies' earnings plays a critical role in successful investing. This piece of research contrasts estimation errors of the seasonal random walk model and exponential smoothing models employed in the earnings per share (EPS) data for Polish listed businesses from the timespan between 2008–2009. The models are compared using the mean arctangent absolute percentage error (MAAPE) metric. The best model across all quarters and years is the seasonal random walk (SRW) model, when contrasted with the other models studied regardless of the analysed time spans and error metrics. Contrary to the results obtained from the US market, the more intricate exponential smoothing model, comprising a seasonal and a trend component, does not suitably explain the behaviour of Polish companies. This could be attributable to the simpler demeanour of the Polish market and the absence of a trend in the EPS data.

Keywords: earnings per share, random walk, exponential smoothing, financial forecasting, Warsaw Stock Exchange

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

* University of Warsaw, Faculty of Management; e-mail: wkurylek@wz.uw.edu.pl.

1. Introduction

EPS forecasts are vital because they provide useful numerical information about a firm's prospects. They disclose information about the future market value of securities and set expectations in auditing. The motivation for this study was an article by Brandon, Jarrett and Khumawala (2007) in which the authors showed the usefulness of the quite old-fashioned Holt-Winter (HW) model in predicting EPS for a sample of US firms over a 20-year horizon. For short-term forecasting, the HW model provided relatively accurate forecasts compared to other methods. According to the authors, the model was to be a cost-effective alternative to more time-consuming and expensive techniques. The Holt-Winters exponentially weighted average forecast methodology was first described in the paper by Holt (1957, 2004) and Winters (1960), and since then it has been widely used. Moreover, these findings were confirmed in another paper by Brandon, Jarrett and Khumawala (2008). In this paper, the authors tested the random walk, Holt-Winters, and regression based on economic indicators models. The results indicated that the HW model outperformed other models. Similar conclusions were made by Jarret (2008). It was found that no one procedure was considered superior to the others in all aspects, but that exponential smoothing models proved to be the most accurate and better than individual autoregressive integrated moving average (ARIMA) models. Interestingly, the above findings contradict well-documented earlier economic literature (Ball, Watts 1972; Johnson, Schmitt 1974; Buckmaster 1976; Ruland 1980; Brandon, Jarrett, Khumawala 1983). According to these many articles, the naïve random walk process outperformed more sophisticated exponential smoothing techniques for annual EPS data. There is a lack of literature referring to quarterly data in this respect. Hence, this article fills this gap. The studies that use quarterly data focus mainly on the methodology developed by Box and Jenkins (1976). The autoregressive integrated moving average models (ARIMA) invented by them were many times examined and compared with naïve random walk class of models (Ball, Watts 1972; Watts 1975; Griffin 1977; Foster 1977; Brown, Rozeff 1977, 1979). In some articles, it was argued that the naïve model provided the best results and more advanced mechanical models were not able to beat the naïve ones, whereas in others conclusions were different. But in the late 1970s, a consensus arose among researchers that ARIMA-type models were the best (Lorek 1979; Bathke, Lorek 1984). This changed in the late 1980s along with the opinion that forecasts provided by financial analysts were better than those made by time series models (Brown et al. 1987). This approach prevailed till the most recent years when the superiority of analysts over time series was questioned (Pagach, Warr 2020). For the Polish market, a study was made by Kuryłek (2023). He compared forecast errors of different univariate time series models, including various naïve random walk models as well as ARIMA-type models, and found that the best model was the seasonal random walk (SRW) model across all analysed quarters. It can be easily noted that all the above-mentioned papers except one are focused on the US stock market. Furthermore, the above research is restricted only to annual data. Hence, there is a need to also test the relevance of the exponential smoothing model for other markets using quarterly data. In the US the application of exponential smoothing models to EPS forecasting has been explored since the early 1970s. The usefulness of the exponential smoothing model with a trend and a seasonal component (the Holt-Winter model) was pointed out by Brandon in 2007. This model provided accurate forecasts in comparison to other methods used and was likely to be a cost-effective alternative to more time-consuming and expensive techniques. These findings were confirmed in other papers by Brandon, Jarrett and Khumawala (2008) and by Jarret (2008).

In this study, the relative accuracy of the three exponential smoothing methods and the seasonal random walk method are compared. These models are applied to 267 Polish-listed companies using quarterly EPS numbers. The examined period lasts from the financial crisis of 2008–2009 till the pandemic shock of 2020. The years 2017–2019 are used for forecast testing purposes. The sample is the same as in the study by Kuryłek (2023).

Rather than using the standard mean absolute percentage error (MAPE) metric, which exhibits explosive values when the denominator is small (i.e. when actual earnings are close to zero), the mean arctangent absolute percentage error (MAAPE) was calculated and utilized (Kim, Kim 2016).

Summarizing, the objective of this article is three-fold. The first is to find out whether the last made conclusions of exponential smoothing superiority are also valid for the Polish market, for the most recent period of stable market behaviour, i.e. for the years 2010–2019. The second is to examine the above relevance using quarterly data, because all existing studies were based on annual numbers. The third is to modify in analysis the widely used MAPE metric to deal with situations where actual profits are close to zero, i.e. use the mean arctangent absolute percentage error (MAAPE) as an error metric. Additionally, a wide range of analysed time spans, error metrics as well as different statistical tests were applied to strengthen the conclusions.

2. Literature review

The application of exponential smoothing in economics was first suggested in the statistical literature by Robert Goodell Brown in 1956, where he focused on using this technique in the prediction of demand. Then, his approach was extended to the Holt-Winters exponentially weighted average forecast methodology, which was first described in the papers by Holt (1957, 2004) and Winters (1960), both in terms of its concepts and its first computer implementation. The original motivation for the development of the methodology by Holt was the widespread need for a feasible technique that could be applied to the forecasting of sales on a product-by-product basis. Once introduced, the method has come to be widely applied as a practical technique. An extensive theoretical approach was then presented in the paper by Brown, D'Esopo and Meyer (1961). Higher orders of smoothing were defined and the procedures of initial values estimation as well as smoothing constants were described. More studies were conducted on the use of exponential smoothing to forecast demand. Groff (1973) compared the short-range forecasting effectiveness of exponentially smoothed and selected Box-Jenkins models for sixty-three monthly sales series. Among verified exponential smoothing models were Winters' three-parameter model and a single-parameter model. The forecasting errors of the best of the Box-Jenkins models were found approximately equal to or greater than the errors of the corresponding exponentially smoothed models. The general usefulness of exponential smoothing methods was suggested in the book by Granger and Newbold (1977). The authors considered exponential smoothing models to be the best predictors for short-term time series. They also described a procedure for initial values and smoothing parameters' estimation. This was confirmed in later research. For instance, in the study by Makridakis, Hibon and Moser (1979), the authors used 111 time series to examine the accuracy of various forecasting methods, particularly time-series methods. The study showed that simpler methods like Holt-Winters performed well in comparison to the more complex and statistically sophisticated ARMA models concerning the mean absolute percentage error (MAPE) metric.

The issue of Earnings Per Share (EPS) forecasting had been investigated in the literature since the late 1960s, mostly for US companies. Various models were examined, including naïve random walk class of models as well as autoregressive integrated moving average (ARIMA) type models (Ball, Watts 1972; Watts 1975; Griffin 1977; Foster 1977; Brown, Rozeff 1977, 1979). The results of these studies were ambiguous – in some works, it was argued that the naïve model yielded the best results and more advanced mechanical models were not able to beat the naïve ones, whereas in others different conclusions were drawn. However, a consensus formed among researchers that ARIMA-type models performed the best (Lorek 1979; Bathke, Lorek 1984). This trend in opinions lasted until the late 1980s, when the widespread consensus that forecasts provided by financial analysts were better than those made by time series models was formed (Brown et al. 1987). This opinion prevailed till the most recent years when the superiority of analysts over time series was questioned again. In their study, Bradshaw et al. (2012) undertook a fresh examination of the commonly held belief that analysts' EPS forecasts outperform random walk (RW) time-series forecasts. Surprisingly, their findings revealed that basic RW forecasts exhibit greater accuracy compared to analysts' forecasts when considering longer time horizons, smaller or younger firms, and situations where analysts predict negative or substantial changes in EPS. Pagach and Warr's (2020) findings validated that in a significant proportion of cases (around 40%), ARIMA time-series forecasts of quarterly EPS were either on par with or more precise than the consensus analysts' forecasts. Additionally, the degree of time-series superiority grew for longer forecast horizons, intensified as firm size decreased, as in the previous research, and became more pronounced for high-technology firms. Similarly, Gaio et al. (2021) suggested the forecasting superiority of the random walk model when compared to the market analysts' forecasts in Brazil. Moreover, Bansal, Nasseh and Strauss (2015) demonstrated that numerous financial and economic variables, including the price-earnings ratio, dividend yield, and Treasury bill rate, did not effectively predict EPS in out-of-sample scenarios when compared to a basic autoregressive model. Conversely, the authors employed a combination forecast method that integrated both firm-specific and macroeconomic variables and observed significant improvement in predictive accuracy compared to the autoregressive benchmark.

The most recent part of the research is focused on the application of artificial neural networks to earnings per share forecasting. It was found by Cao and Parry (2009) that the univariate neural network model significantly outperformed four alternative univariate models examined in prior research. In another article from the same year Cao and Gan (2009) used neural network models to forecast earnings per share (EPS) of Chinese listed companies. They showed that the neural network model with weights estimated with the genetic algorithm outperformed the neural network with weights estimated with the backpropagation. Ahmadpour, Etemadi and Moshashaei (2015) examined EPS forecasting using a multi-layer perceptron (MLP) neural network and rule extraction from neural network by genetic algorithm technique and showed that this rule was significantly more accurate than the MLP model. Elend et al. (2020) compared long-term short-term memory (LSTM) networks to temporal convolution networks (TCNs) in the prediction of future EPS. For a broad sample of US firms, they found that both LSTMs outperformed the naïve persistent model with up to 30.0% more accurate predictions, while TCNs achieved an improvement of 30.8%. Both types of networks were at least as accurate as analysts and exceed them by up to 12.2% (LSTM) and 13.2% respectively.

For the Polish market, a study was made by Kuryłek (2023). The author compared forecast errors of different univariate time series models including various naïve random walk models as well as ARIMA-type models. He applied them to the EPS data for Polish companies from the period between the last financial crisis of 2008–2009 and the pandemic shock of 2020. The best model was the seasonal random walk (SRW) model across all quarters, which described quite well the behaviour of the Polish market compared to other analysed models.

The exponential smoothing methods were also applied for EPS forecasting. The research by Elton and Gruber (1972) examined the accuracy of the EPS forecasts produced by nine mechanical models, including exponential smoothing ones. They found that the additive exponential smoothing with no trend dominated other models. Additionally, it turned out in their work that the differences in forecast accuracy of mechanical models and security analysts' forecasts were not significant. Using exponential smoothing techniques, other scientists examined whether the underlying EPS-generating process might have a martingale or sub-martingale nature. In the paper by Ball and Watts (1972), the authors used exponential smoothing as a primary method for assessing whether an underlying EPS process is a martingale or sub-martingale. They looked at a smoothing constant and found it to be either equal to one or slightly lower than one. Hence they concluded that the EPS time series must be a martingale type. Johnson and Schmitt (1974) tested various mechanical models for EPS forecasting, including naïve random walk, moving average model, linear projection model, single double and triple exponential smoothing models, and the accuracy of their forecasts was calculated. The naïve model gave the best results and even more complex mechanical models were not able to beat the naïve one. This work was extended by Salamon and Smith (1977) using a similar framework based on exponential smoothing. They showed that selection bias in the study conducted by Ball and Watts caused them to overestimate the instability in the EPS time series. They also claimed that there was a diversity in time series characteristics of the EPS sequence of individual firms. In the following research Brooks and Buckmaster (1976) applied single, double, and triple exponential smoothing models for different strata of earnings time series. The best-smoothing constant for each of those three models was then determined for each stratum. It occurred that for most of the strata, the best smoothing model was of order one with a smoothing constant equal to one, which indicated that income time series normally followed a martingale or similar process. However, for the outer stratum, the best model was the model of order two with a smoothing constant of less than one which indicated that these time series didn't follow a martingale process. The authors also suggested that it indicated that income tended to revert to previous levels in the period subsequent to a substantial deviation from an operationally defined norm. Brandon and Jarret (1979) tested seven extrapolative forecasting models including the random walk model, ARIMA models with the autocorrelation part and exponentially smoothing models with optimal linear correction technique developed by Theil as well as the Bayesian revision procedure. The results indicated that the optimal correction technique was superior to not correcting for past errors. In the study of Chant (1980) there were employed forecasting models utilizing economic leading indicators to examine the predictability of annual EPS behaviour. The lead indicators were the growth of money supply, stock market index, and bank loan growth. These models were compared to three univariate time series models, i.e. an average growth model, an exponential smoothing model, and a random walk model (RW). The author found that the errors for the models relying on economic leading variables were smaller than pure time series models. The results indicated the existence of a predictable relationship between some economic leading indicators and EPS numbers. Ruland (1980) evaluated the relative ability of various extrapolative models to predict future annual earnings, one of which was an exponential smoothing model. A comparison of alternative models revealed that the simple martingale dominated the other models tested. The results showed that the simple martingale model outperformed the other models except in the case of large prior period earnings changes and in one of the ten specific industries selected for examination. Brandon, Jarrett and Khumawala (1983) tested various models, including the random walk model and three models of simple exponential smoothing. The results of this study indicated that some accounting time series were random walks, which was consistent with many previous studies. However, exponential smoothing yielded similar results in terms of the relative mean absolute error metric. Conroy and Harris (1987) examined the primary forecasting advantages of analysts over time series methods, including random walk, single, double, and triple exponential smoothing, and the simple average of the last 5 years of EPS numbers. However, the authors concluded that, on average, the primary forecasting advantages of analysts over time series methods appeared to occur over short forecast horizons (less than one year). Moreover, this superiority declines steadily as the forecast horizon increases. Neither analysts nor other time series methods substantially outperformed a random walk prediction when the forecasts were made at the beginning of the year. It is worth mentioning that all the above studies were based on annual data.

In 2007 Brandon, Jarrett and Khumawala pointed out the usefulness of the Holt-Winter (HW) model in predicting EPS for a random sample of firms in the US over a 20-year horizon. They used the mean absolute percentage error (MAPE) measure. For short-term forecasting, the HW model provided relatively accurate forecasts in comparison to other methods used. The HW model was likely to be a cost-effective alternative to more time-consuming and expensive techniques. These findings were confirmed in another paper by Brandon, Jarrett and Khumawala (2008), where the authors tested the random walk, Holt-Winters, and regression based on economic indicators models. As an error metric, they used the MAPE error metric. The results indicated that the HW model outperformed other models. Similar conclusions were made by Jarret (2008). The four tested models were: (1) the Holt-Winters multiplicative exponential smoothing model, (2) the univariate Box-Jenkins model, (3) linear autoregression of data seasonally adjusted by the Census II–XII method, and (4) linear autoregression of the data seasonally adjusted by the X11-ARIMA method. No one procedure was considered superior to the others in all aspects, but according to the MAPE measure, exponential smoothing models proved to be the most accurate and better than individual Box-Jenkins models.

3. Data and methodology

3.1. Data

The Polish stock market is one of the deepest among those that joined the European Union after 2004, with a capitalization of USD 197 billion and 774 listed companies at the end of 2021. Its stocks are not as widely covered by financial analysts as those of the US or Western Europe, with only around 20% of the 711 listed companies being covered in 2019. I focus on the earnings per share (EPS) data series. Data sourcing was conducted through EquityRT, a financial analysis platform.[1] Consequently, I analysed EPS firms listed on the Warsaw Stock Exchange, spanning from Q1 2010 to Q4 2019, i.e. between two structural shifts, the first of which being the financial crisis of 2008–2009, and the second being the onset of COVID-19 in 2020. The respective time series were analysed on a level scale.

[1] Provided by a Turkish company RASYONET.

For forecasting purposes, the data for Q1 2010–Q4 2018 (36 quarters) were used for model estimations, with data from Q1 2019–Q4 2019 acting as a validation sample for accuracy testing of 1 quarter-ahead, 2 quarters-ahead, 3 quarters-ahead, and 4 quarters-ahead forecasts. This led to a survivorship bias – due to the need for a long enough series of data – though comparability of results was assured. But this problem is inevitable in the selection of all sufficiently long time series. Alternatively, expanding window approaches were tested using the years 2017 and 2018 as validation samples. After applying a full-time window coverage and excluding the impact of splits and reverse splits and removing them from the sample, 267 companies remained in the sample. This included those affected by government regulations, such as utilities and financial sectors, as it was hard to determine the extent to which government regulations shaped businesses' income.

3.2. Methodology and research

The models

Denote as Q_t realization of *EPS* in quarter *t*. The following four time series models, estimated for each individual company separately, are analysed in this paper:

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

$$Q_t = Q_{t-4} + \varepsilon_t \tag{1}$$

where ε_t are IID and $\varepsilon_t \sim N\left(0, \sigma^2\right)$.

The forecast is $\hat{Q}_t = Q_{t-4}$, so the model doesn't need any estimation of parameters to make the forecasts because the value delayed by 4 quarters is the forecast. To estimate the variance of the disturbance term: $\varepsilon_t = Q_t - Q_{t-4}$ the following calculations have to be made:

$$\hat{\sigma}^2 = \sum_{t=1}^{T} \frac{\left(e_t - \bar{e}_t\right)^2}{T-1}$$

where $\bar{e}_t = \sum_{t=1}^{T} \frac{e_t}{T}$ and e_t are realizations of ε_t variable.

2. The Brown model – Simple Exponential Smoothing (SES)

$$l_t = \alpha \cdot Q_t + \left(1-\alpha\right) \cdot l_{t-1} \tag{2}$$

where l_t represents a level component.

Hence the forecast can be expressed as $\hat{Q}_{t+h} = l_{t-1}$. The model was initially proposed by Brown (1956). The unknown parameters – initial value and smoothing constant l_0, α are estimated by minimizing the sum of squared errors.

3. The Holt model – Double Exponential Smoothing (DES)

The additive version of the model is the following (a similar multiplicative version of the model can be easily derived):

$$\begin{cases} l_t = \alpha \cdot Q_t + (1-\alpha) \cdot (l_{t-1} + b_{t-1}) \\ b_t = \beta \cdot (l_t - l_{t-1}) + (1-\beta) \cdot b_{t-1} \end{cases} \tag{3}$$

where l_t, b_t are level and trend components respectively.

Thus the forecast can be denoted as $\hat{Q}_{t+h} = l_{t-1} + \beta \cdot h$. This model was developed by Holt (1957). The parameters l_0, b_0, α, β are estimated by minimizing the sum of squared errors.

4. The Holt-Winters model – Triple Exponential Smoothing (TES)

The additive version of the model is the following (a similar multiplicative version of the model can be easily derived):

$$\begin{cases} l_t = \alpha \cdot (Q_t - s_{t-T}) + (1-\alpha) \cdot (l_{t-1} + b_{t-1}) \\ b_t = \beta \cdot (l_t - l_{t-1}) + (1-\beta) \cdot b_{t-1} \\ s_t = \gamma \cdot (Q_t - l_t - b_t) + (1-\gamma) \cdot s_{t-T} \end{cases} \tag{4}$$

where $l_t, b_t, s_{t \bmod T}$ are level, trend, and seasonal components (with a period T) respectively.

The frequency of the seasonality, i.e. the number of seasons in a year, is the T parameter, which is equal to 4, since quarterly data are used. Thus the forecast can be formulated as $\hat{Q}_{t+h} = l_{t-1} + \beta \cdot h + s_{t+h \bmod T}$. This approach was published by Winters (1957). The parameters $l_0, b_0, \alpha, \beta, \gamma, s_0, \dots, s_{T-1}$ are estimated by minimizing the sum of squared errors.

The forecasts were calculated with the use of the Statsmodels library in Python computer programming language.

Mean arctangent absolute percentage error (MAAPE)

Let us denote $A_1^i, \dots, A_4^i$ as the actual earnings per share (EPS) for the 1-st,…, 4-th quarter of 2019 for a given firm i. $F_1^i, \dots, F_4^i$ are the forecasts of this variable in the corresponding periods (i.e. $\hat{Q}_t$, where t = 37,…, 40 for i-th company). For any firm i, during the j-th quarter of 2019, the absolute percentage error (APE) of such forecasts can be expressed as follows:

$$APE_j^i = \left| \frac{A_j^i - F_j^i}{A_j^i} \right| \tag{5}$$

Nevertheless, APE is accompanied by a major setback: when actual values are close to or equal to zero, then it produces infinite or undefined results – a common problem during earnings forecasts. Additionally, if the actual values happen to be very small (usually lower than one), this would then generate extreme percentage errors (outliers). Finally, zero actual values will lead to infinite APEs. To fix this dilemma, Kim and Kim (2016) proposed the arctangent absolute percentage error as a new approach in the literature.

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

This is due to *arctan* being a function that maps any value between the range $[-\infty, +\infty]$ into one from the interval $[-\pi/2, \pi/2]$.

Consequently, the mean arctangent absolute percentage error (MAAPE) for the *i*-th firm can be formulated as follows:

$$MAAPE^i = \frac{1}{4}\sum_{j=1}^{4} AAPE_j^i = \frac{1}{4}\sum_{j=1}^{4} arctan\left(\left|\frac{A_j^i - F_j^i}{A_j^i}\right|\right) \tag{7}$$

In addition, the mean arctangent absolute percentage error (MAAPE) for the *j*-th quarter among all *I* companies in the sample can be represented as:

$$MAAPE_j = \frac{1}{I}\sum_{i=1}^{I} AAPE_j^i = \frac{1}{I}\sum_{i=1}^{I} arctan\left(\left|\frac{A_j^i - F_j^i}{A_j^i}\right|\right) \tag{8}$$

Therefore, the mean arctangent absolute percentage error (MAAPE) for all 4 quarters and among all *I* companies in the sample can be summarized by the following formula:

$$MAAPE = \frac{1}{I}\sum_{i=1}^{I} MAAPE^i = \frac{1}{4}\sum_{j=1}^{4} MAAPE_j = \frac{1}{4I}\sum_{i=1}^{I}\sum_{j=1}^{4} arctan\left(\left|\frac{A_j^i - F_j^i}{A_j^i}\right|\right) \tag{9}$$

Forecasts are generated through the described models and, for each model m, $MAAPE(m)_1, \ldots,$ $MAAPE(m)_4$, along with $MAAPE(m)$ are then calculated.

The average rank of error

For every firm-quarter combination, the absolute percentage errors of the indicated models are sorted. The model with the least amount of error is given a rank of 1, and the one with the highest error gets a score of 4. Then, we calculate the average rank of each model across all firms for forecasts from the first to the fourth quarter of 2019, as well as the average rank across four quarters and companies combined. $AAPE(m)_j^i$ represents the arctangent absolute percentage error of the forecast for an *i*-th company in the *j*-th quarter of 2019 generated by the *m*-th model with $R(m)_j^i$ being its rank (where $m = 1$ to 4). Therefore, we can express the average rank of the *m*-th model and *i*-th company across all four quarters as follows:

$$\bar{R}(m)^i = \frac{1}{4}\sum_{j=1}^{4} R(m)_j^i \tag{10}$$

The average rank of the m-th model and the j-th quarter for all companies represented in our sample can be indicated as follows:

$$\bar{R}(m)_j = \frac{1}{I}\sum_{i=1}^{I} R(m)_j^i \tag{11}$$

The average rank of the m-th model for all I companies in the sample, over all four quarters, can be expressed as such:

$$\bar{R}(m) = \frac{1}{I}\sum_{i=1}^{I} \bar{R}(m)^i = \frac{1}{4}\sum_{j=1}^{4} \bar{R}(m)_j = \frac{1}{4I}\sum_{i=1}^{I} R(m)_j^i \tag{12}$$

For the models described, forecasts are made and the averages $\bar{R}(m)_1, \ldots, \bar{R}(m)_4$ as well as $\bar{R}(m)$ are calculated for the m-th model.

The equality of means tests

To test a statistical significate difference in mean arctangent absolute percentage errors (MAAPEs) across several means, three statistical tests have been used: the one-way ANOVA test, the Alexander--Govern test, and the Kruskal-Wallis test. These tests are described below.

The one-way ANOVA test

To test whether the mean of errors denoted by MAAPEs are statistically different the one-way ANOVA test (Lowry 2014) is used, which is widely applied for testing the equality of means. However, the test has important assumptions that must be satisfied for the associated p-value to be valid, including the independence of variables generating observations in the sample, the homoscedasticity of their variances in different groups, and their normality of distributions.

$$H_0\text{: } \textit{means of AAPEs of all } 4 \textit{ models are the same} \tag{13}$$

The Alexander-Govern test

Unlike the one-way ANOVA test, this test does not assume homoscedasticity, instead relaxing the assumption of equal variances (Alexander, Govern 1994). The rest of the assumptions, including the normality of distribution, hold.

$$H_0\text{: } \textit{means of AAPEs of all } 4 \textit{ models are the same} \tag{14}$$

The Kruskal-Wallis test

After that, a Kruskal-Wallis one-way H-test (Corder, Foreman 2009) is conducted. This is a nonparametric test that bypasses the issues surrounding the potential normality of the errors. The relative closeness of the average ranks of the 4 models implies that the null hypothesis of median

AAPE's equality cannot be rejected. The calculation is done for each respective quarter as well as for all forecast quarters, resulting in Kruskal-Wallis H statistics with the respective p-values.

$$H_0:\ medians\ of\ AAPEs\ of\ all\ 4\ models\ are\ the\ same \tag{15}$$

It is also worth mentioning that in the previous research, the Alexander-Govern test and the Kruskal-Wallis test were not considered.

The Wilcoxon test

Lastly, the paired comparison of forecast errors is completed with the help of a nonparametric two-sided Wilcoxon test (Wilcoxon 1945) to assess the similar median errors of diverse models. For two matched samples, it is a paired difference test. It is essential to note that this test does not demand particular assumptions about a probability distribution, except for the symmetricity of the difference in scores and independence of variables that generate observations. An excellent explanation of the use of the Wilcoxon test in the context of verification if errors produced by various EPS models are statistically different was provided by Ruland (1980). Separate tables for every quarter from one to four, as well as all quarters combined, are produced in which the p-values of the Wilcoxon statistic are located above the diagonal for all model pairs.

$$H_0:\ medians\ of\ AAPEs\ of\ a\ pair\ of\ models\ are\ the\ same \tag{16}$$

The rejection of the null hypothesis of respective tests happened when their p-values were less than the accepted significance level of 0.05, the concept of which is widely applied, among others, by Ruland (1980). The above test statistics and their p-values were calculated using the Scipy library in Python programming language.

If it emerged from the tests that the mean (median) errors of a particular model type were lower and statistically different than other models it would be inferred as a superiority of this model class over the others.

4. Results

4.1. Empirical findings

The seasonal random walk (SRW) model, presented in Table 1, outperforms all other models in every quarter, both in terms of its ranking and its overall performance. The double exponential smoothing (DES) model ranks the lowest among all models and performs the worst, while the triple exponential smoothing (TES) model only slightly outperforms it. The simple exponential smoothing model (SES) performs worse than the best model and better than other exponential smoothing models in all analysed time spans.

Table 1 presents the results of several equality of means tests, including the one-way ANOVA test (F statistic), the Alexander-Govern test (AG statistic), and the Kruskal-Wallis test (H statistics). All those tests confirm in the 1st and 3rd quarters and also for all the joint quarters, that the null hypothesis –

that the means (medians in case of the last test) of arctangent absolute percentage errors (AAPEs) of all 4 models are statistically equal – can be rejected. Interestingly, in the 2nd and 4th quarters, all tests are not able to reject this hypothesis and the respective p-values of these test statistics are far above any accepted significance levels.

To assess whether the errors of the best model differ statistically from the results of the other models, the Wilcoxon nonparametric test was used to compare the medians of the arctangent absolute percentage errors (AAPEs) of all model pairs. As shown in Table 2, the results indicate that the seasonal random walk (SRW) model produces a statistically significant lower median of errors in the 1st and 3rd quarters than other models, which is further confirmed in Table 3 for all joint quarters. This finding is consistent with the above-mentioned outcomes of equality of several means tests. However, in the 2nd and 4th quarters, the hypothesis about a statistically significant difference in medians between the seasonal random walk (SRW) and the simple exponential smoothing model (SES) cannot be rejected. It explains why the equality of several means tests can not reject the null hypothesis in these quarters. This observation paves the way for the conclusion that the seasonal random walk model (SRW) performs the best of all tested models but statistically may not differ in specific quarters from the results of another basic model – the simple exponential smoothing model (SES). It is worth mentioning that both the double exponential smoothing model (DES) and the triple exponential smoothing model (TES) provide statistically the same outcome in all analysed time spans and perform much worse than the best model.

The above findings confirm the superiority of the seasonal random walk model (SRW) for EPS forecasting in Poland, which sometimes can be forecasted using another very simple and old technique – the simple exponential smoothing (SES). The much worse and simultaneously similar performance of more sophisticated exponential smoothing models can be explained by the absence of a trend in the Polish market, which is present in both models. This might coincide with the horizontal performance of the stock market index WIG during the analysed period.

The conclusion made by Brandon, Jarrett and Khumawala (2007) and Jarret (2008) about the superiority of the Holt-Winters model (i.e. the triple exponential smoothing model) in comparison to other methods used for the US market does not hold for Poland, where this model performs as the one of the worst. In Poland, the most efficient is the seasonal random walk model, which sometimes gives an outcome comparable to the simple exponential smoothing model. This might be explained either by the relative simplicity of the Polish stock market compared to the US one or the absence of trends in the Polish EPS data in the examined period. Together with the results obtained by Kuryłek (2023) it proves that no time series models from either ARIMA or exponential smoothing families can do a better job than the simple random walk model. Hence it has a practical consequence, that the application to EPS forecasting for investment purposes of any of the above-mentioned more sophisticated techniques than the ordinary seasonal random walk one makes no sense in Poland. Moreover, assuming that the EPS process is driven by the seasonal random walk, and knowing that the market price of stock comes from a multiplication of the P/E multiple by EPS, it could be inferred that stock prices also behave at least as randomly as EPS. The forecast of the seasonal random walk is simply a value from the respective quarter of the previous year. Thus, it might imply that for predicting future prices the P/E multiple is more important than the next year's earnings of companies. It is consistent with economic theory, saying that this multiple refers more to the expected growth of future earnings, the level of future interest rates, and the market premium which reflects the risk appetite of investors, as the EPS forecast, which refers only to the near future earnings.

4.2. Robustness checks

The robustness checks were calculated with respect to time but also with respect to other popular error metrics.

Table 4 demonstrates that the seasonal random walk model (SRW) obtained the least MAAPE error metric and the lowest rank in 2019, 2018, and 2017. The statistically significant difference in the results of different models is corroborated by the low p-values of all statistical tests carried out: the one-way ANOVA test, the Alexander-Govern test, and the Kruskal-Wallis test. Additionally, the Wilcoxon test was applied to all model pairs with the seasonal random walk model, and the p-values for each year are listed in Table 5. The seasonal random walk model (SRW) achieved statistically better results than all the other surveyed exponential smoothing models in each respective year. Therefore, the seasonal random walk model's superiority appears consistent over time.

In Table 6 the performance of the analysed models was investigated with respect to other error metrics like, the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) for all combined quarters in 2019. The above measures were adjusted to the CPI inflation since a future error must have in nominal term the same present value as the current one. Again, it appeared that the seasonal random walk gave the lowest errors measured either in terms of RMSE or MAE. However, the statistically significant difference between the results of the various techniques was not confirmed by the one-way ANOVA test, the Alexander-Govern test, and the Kruskal-Wallis test. P-values of these test statistics were far above any sensible significance level. This might be explained by the result of Table 7, which indicated that the seasonal random walk (SRW) model made forecasts both in term of RMSE and MAE not statistically different from the simple exponential smoothing (SES) model and simultaneously the outcomes of the double exponential smoothing (DES) and the triple exponential smoothing models were also not significantly different. Nevertheless, the seasonal random walk model gave the lowest forecast error using the other error measures, but that was not statistically different from the prediction of the simplest of explanation smoothing models.

5. Conclusions

The paper examines the forecasting characteristics of four univariate time-series models: the seasonal random walk (SRW), simple, double, and triple exponential smoothing models. When applied to the quarterly Earnings per Share (EPS) of 267 Polish companies from 2010 to 2019, the SRW model obtained the lowest rank and described the behaviour of the Polish market comparatively better than the other models. This is further evidenced by the one-way ANOVA, Alexander-Govern, Kruskal-Wallis, and Wilcoxon tests. However, this contradicts Brandon et al. (2007) and Jarret's (2008) findings that the triple exponential smoothing model is the most appropriate for the US market, as in the case of Poland it was not. This can be attributed to the absence of a trend in the EPS data in Poland, which is consistent with the horizontal nature of the WIG stock market index throughout the analysed period. It could be also due to the relative simplicity of the Polish stock market compared to the US. Additionally, the SRW model appeared to remain superior regardless of time and the other error measures like RMSE or MAE. It has a practical consequence that the application to EPS forecasting for investment purposes of any of the above-mentioned more sophisticated techniques than the ordinary seasonal random walk one makes no sense in Poland. However, relying on seasonal random walk in EPS modelling implies that forecasted stock prices might behave in a very random way. Hence, the prediction of the P/E multiple could be more important than EPS prediction for forecasting future prices.

It would be of interest in a future research agenda to ascertain whether more modern and newer time-series models with a basis in neural networks provide more accurate predictions compared to the naïve seasonal random walk model. Furthermore, the connection between forecasting efforts and firm size could also be considered. The business nature described by the sector in which a company operates can be a major factor when determining which model has the most accurate forecasting of earnings per share. Also, an effect of time series transformation making the EPS distribution more in line with a normal distribution could be investigated. There may be also a seasonal pattern identified by the SRW model, which could hint at possible investment plans. Such a strategy may challenge the "weak form" of the Efficient Market Hypothesis (EMH). Last, it would be interesting to compare the results of the best models with the forecasts made by professional market analysts.

References

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.

Alexander R.A., Govern D.M. (1994), A new and simpler approximation for ANOVA under variance heterogeneity, *Journal of Educational Statistics*, 19(2), 91–101.

Ball R., Watts R. (1972), Some time series properties of accounting income, *The Journal of Finance*, 27(3), 663–681.

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*, 39(1), 1.

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.

Box G.E.P., Jenkins G.M. (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day.

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.

Brandon Ch.H., Jarrett J.E. (1979), Revising earnings per share forecasts: an empirical test, *Journal of Accounting Research*, 17(1), 179–189.

Brandon Ch., Jarrett J.E., Khumawala S.B. (1983), On the predictability of corporate earnings per share, *Journal of Business Finance & Accounting*, 10(3), 373–387.

Brandon Ch., Jarrett J.E., Khumawala S.B. (2007), Comparing forecast accuracy for exponential smoothing models of earnings-per-share data for financial decision making, *Decision Sciences*, 17, 186–194.

Brandon Ch., Jarrett J.E., Khumawala S.B. (2008), 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, 10–15.

Brooks L.D., Buckmaster D.A. (1976), Further evidence of the time series properties of accounting income, *Journal of Finance*, 31(5), 1359–1373.

Brown R.G. (1956), *Exponential Smoothing for Predicting Demand*, National Bureau of Economic Research.

Brown R.G., D'Esopo D.A., Meyer R.F. (1961), The fundamental theorem of exponential smoothing, *Operations Research*, 9(5), 673–687.

Brown L.D., Hagerman R.L., Griffin P.A., 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. (1977), *Univariate time-series models of quarterly accounting earnings per share: a proposed premier model*, Working Paper, 77–27, College of Business Administration, University of Iowa.

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 neural network with genetic algorithm*, 15th Americas Conference on Information Systems 2009, AMCIS 2009.

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.

Chant P.D. (1980), On the predictability of corporate earnings per share behavior, *The Journal of Finance*, 35(1), 13–21.

Conroy R., Harris R. (1987), Consensus forecasts of corporate earnings: analysts' forecasts and time series methods, *Management Science*, 33(6), 725–738.

Corder G.W., Foreman D.I. (2009), *Nonparametric Statistics for Non-Statisticians*, John Wiley & Sons.

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 Junior T. (2021), Re-examining analyst superiority in forecasting results of publicly-traded Brazilian companies, *Revista de Administracao Mackenzie*, 22(1), eRAMF210164.

Granger C.W.J., Newbold P. (1977), *Forecasting Economic Time Series*, Academic Press.

Griffin P. (1977), The time-series behavior of quarterly earnings: preliminary evidence, *Journal of Accounting Research*, 15(1), 71–83.

Groff G.K. (1973), Empirical comparison of models for short range forecasting, *Management Science*, 20(1), 22–31.

Holt Ch.C. (1957), *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*, Carnegie Institute of Technology.

Holt Ch.C. (2004), Forecasting seasonals and trends by exponentially weighted moving averages, *Journal of Economic & Social Measurement*, 29(1–3), 123–125.

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. (2023), The modelling 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.

Lorek K.S. (1979), Predicting annual net earnings with quarterly earnings time-series models, *Journal of Accounting Research*, 17(1), 190–204.

Lowry R. (2014), *Concepts and Applications of Inferential Statistics*, https://onlinebooks.library.upenn.edu/webbin/book/lookupid?key=olbp66608.

Makridakis S., Hibon M., Moser C. (1979), Accuracy of forecasting: an empirical investigation, *Journal of the Royal Statistical Society*, Series A (General), 142(2), 97–145.

Pagach D.P., Warr R.S. (2020), Analysts versus time-series forecasts of quarterly earnings: a maintained hypothesis revisited, *Advances in Accounting*, 51(C), 51.

Ruland W. (1980), On the choice of simple extrapolative model forecasts of annual earnings, *Financial Management*, 9(2), 30–37.

Salamon G.L., Smith E.D. (1977), Additional evidence on the time series properties of reported earnings per share: comment, *The Journal of Finance*, 32(5), 1795–1801.

Watts R.L. (1975), *The time series behavior of quarterly earnings*, Working Paper, April, University of New Castle, Department of Commerce.

Winters P.R. (1960), Forecasting sales by exponentially weighted moving averages, *Management Science*, 6(3), 324–342.

Wilcoxon F. (1945), Individual comparisons by ranking methods, *Biometrics*, 1, 80–83.

Appendix

Table 1

Summary statistics on forecast errors and mean equality tests for 2019 quarters

Model	Quarters								All quarters	
	Q1 MAAPE	Q1 Rank	Q2 MAAPE	Q2 Rank	Q3 MAAPE	Q3 Rank	Q4 MAAPE	Q4 Rank	MAAPE	Rank
SRW	0.66	2.08	0.70	2.32	0.65	2.19	0.74	2.42	0.69	2.25
SES	0.77	2.42	0.72	2.36	0.76	2.39	0.77	2.37	0.76	2.38
DES	0.83	2.84	0.77	2.64	0.83	2.76	0.80	2.61	0.81	2.71
TES	0.82	2.66	0.79	2.67	0.82	2.66	0.80	2.60	0.81	2.65
F statistics	7.33		1.80		7.61		1.34		6.96	
F p-value	0.00		0.15		0.00		0.26		0.00	
AG statistics	23.03		5.37		23.72		3.90		23.07	
AG p-value	0.00		0.15		0.00		0.27		0.00	
H statistics	21.52		5.11		21.26		4.09		16.12	
H p-value	0.00		0.16		0.00		0.25		0.00	

Table 2

P-values of paired Wilcoxon test of forecast errors in respective quarters of 2019

	Q1				Q2		
Model	**SES**	**DES**	**TES**	**Model**	**SES**	**DES**	**TES**
SRW	0.00	0.00	0.00	SRW	0.41	0.02	0.00
SES		0.00	0.05	SES		0.01	0.00
DES			0.37	DES			0.70
	Q1				**Q4**		
Model	**SES**	**DES**	**TES**	**Model**	**SES**	**DES**	**TES**
SRW	0.00	0.00	0.00	SRW	0.45	0.05	0.01
SES		0.00	0.02	SES		0.03	0.11
DES			0.48	DES			0.91

Table 3
P-values of paired Wilcoxon test of forecast errors for all quarters of 2019

Model	SES	DES	TES
SRW	0.00	0.00	0.00
SES		0.00	0.00
DES			0.42

Table 4
Summary statistics on forecast errors and mean equality tests for all quarters 2017–2019

Model	2017		2018		2019	
	MAAPE	Rank	MAAPE	Rank	MAAPE	Rank
SRW	0.69	2.24	0.71	2.24	0.69	2.25
SES	0.75	2.37	0.78	2.46	0.76	2.38
DES	0.80	2.74	0.82	2.72	0.81	2.71
TES	0.80	2.65	0.80	2.58	0.81	2.65
F statistics	6.00		4.82		6.96	
F p-value	0.00		0.00		0.00	
AG statistics	19.70		15.80		23.07	
AG p-value	0.00		0.00		0.00	
H statistics	13.94		12.59		16.12	
H p-value	0.00		0.01		0.00	

Table 5
P-values of paired Wilcoxon test of forecast errors for all quarters 2017–2019 and SRW model

Year	Model	F	BR	BJ
2017	SRW	0.00	0.00	0.00
2018	SRW	0.00	0.00	0.00
2019	SRW	0.00	0.00	0.00

Table 6
Summary statistics on forecast errors for RMSE and MAPE in all quarters 2019

Measure	SRW	SES	DES	TES	F stat	F p-val	AG stat	AG p-val	H stat	H p-val
RMSE	0.94	1.14	1.22	1.17	0.21	0.89	1.21	0.75	4.29	0.23
MAE	0.70	0.94	1.02	0.95	0.34	0.80	2.25	0.52	5.91	0.12

Table 7
P-values of paired Wilcoxon test of forecast errors for RMSE and MAE in 2019

Model	SES	DES	TES	Model	SES	DES	TES
	RMSE				MAE		
SRW	0.22	0.00	0.04	SRW	0.10	0.00	0.00
SES		0.00	0.00	SES		0.00	0.00
DES			0.71	DES			0.59

Czy wygładzanie wykładnicze może dawać lepsze wyniki niż błądzenie losowe w prognozowaniu zysków na jedną akcję w Polsce?

Streszczenie

Dokładna prognoza zysków spółek notowanych na giełdzie odgrywa kluczową rolę w pomyślnym inwestowaniu. Motywacją do tego badania był artykuł Brandona, Jarretta i Khumawala (2007), w którym autorzy wykazali przydatność dość staromodnego modelu wygładzania wykładniczego stworzonego przez Holta i Wintera do przewidywania zysków na jedną akcję (EPS) dla próby firm amerykańskich. W przypadku prognoz krótkoterminowych model ten zapewniał stosunkowo dokładne prognozy w porównaniu z innymi metodami. Według autorów model miał stanowić opłacalną alternatywę dla bardziej czasochłonnych i kosztownych technik. Wyniki te zostały potwierdzone w pracy Brandona i in. (2008) oraz przez Jarreta (2008).

W niniejszym badaniu przeanalizowano jakość prognostyczną modelu sezonowego błądzenia losowego i różnych modeli wygładzania wykładniczego zastosowanych do danych zysków na jedną akcję (EPS) dla polskich spółek giełdowych ze stabilnego okresu między kryzysem finansowym 2008–2009 a szokiem wywołanym pandemią (2020 r.). Zamiast korzystać ze standardowo wykorzystywanego średniego względnego błędu procentowego (MAPE), który wykazuje wartości niewspółmiernie duże, gdy mianownik jest mały (tj. gdy rzeczywiste zyski są bliskie zera), wykorzystano średni arcus-tangens względnego błędu procentowego – MAAPE (Kim, Kim 2016).

Niniejszy artykuł ma trzy cele. Pierwszym jest sprawdzenie, czy ostatnie wnioski o wyższości wygładzania wykładniczego są aktualne również dla rynku polskiego. Drugim jest zbadanie powyższej istotności przy użyciu danych kwartalnych, ponieważ wszystkie istniejące badania opierały się na danych rocznych. Trzeci polega na zmodyfikowaniu powszechnie stosowanej w analizie miary błędu prognozy (MAPE), aby poradzić sobie z kłopotliwymi sytuacjami. Modelem dającym we wszystkich kwartałach i latach najmniejsze błędy okazało się sezonowe błądzenie losowe (SRW), które na polskim rynku sprawdza się dość dobrze w porównaniu z bardziej skomplikowanymi modelami wygładzania wykładniczego.

Statystyczną różnicę między prognozami tego modelu a prognozami innych modeli (jeśli przyjmiemy MAAPE jako miarę błędu prognozy) potwierdzają dodatkowo jednoczynnikowe testy ANOVA, testy Alexandra-Governa, Kruskala-Wallisa i Wilcoxona. Ponadto model SRW okazał się lepszy niezależnie od testowanego okresu prognostycznego czy pozostałych miar błędu prognozy, aczkolwiek nie był statystycznie istotnie różny od błędu najprostszego z modeli wygładzania wykładniczego. Jest to sprzeczne z wcześniejszymi ustaleniami dla rynku amerykańskiego. Można to wiązać z brakiem wyraźnego trendu w danych o EPS spółek giełdowych w Polsce, co jest zgodne z horyzontalnym zachowaniem indeksu giełdowego WIG w analizowanym okresie. Może to również wynikać ze względnej prostoty polskiego rynku akcji w porównaniu z rynkiem amerykańskim. Praktyczną konsekwencją tego jest to, że stosowanie do prognozowania EPS w celach inwestycyjnych którejkolwiek z wyżej wymienionych technik, bardziej wyrafinowanych niż zwykłe sezonowe błądzenie losowe, nie ma w Polsce sensu. Poleganie na sezonowym błądzeniu losowym w modelowaniu EPS oznacza jednak, że ceny akcji mogą się zachowywać w bardzo losowy sposób. Przewidywanie mnożnika P/E może być zatem ważniejsze niż prognoza EPS dla prognozowania przyszłych cen, co jest ugruntowanym poglądem w teorii ekonomii.

Słowa kluczowe: zysk na jedną akcję, błądzenie losowe, wygładzanie wykładnicze, prognozowanie finansowe, Giełda Papierów Wartościowych