The modeling of earnings per share of Polish companies for the post-financial crisis period using random walk and ARIMA models

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

Received: 24 January 2023 / Revised: 6 April 2023 / Accepted: 25 April 2023 / Published online: 11 May 2023

ABSTRACT

The proper forecasting of listed companies' earnings is crucial for their appropriate pricing. This paper compares forecast errors of different univariate time-series models applied for the earnings per share (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 is the seasonal random walk (SRW) model across all quarters, which describes quite well the behavior of the Polish market compared to other analyzed models. Contrary to the findings regarding the US market, this time-series behavior is well described by the naive seasonal random walk model, whereas in the US the most adequate models are of a more sophisticated ARIMA type. Therefore, the paper demonstrates that conclusions drawn for the US might not hold for emerging economies because of the much simpler behavior of these markets that results in the absence of autoregressive and moving average parts.

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

Keywords: earnings per share, time series, random walk, ARIMA, financial forecasting, Warsaw Stock Exchange.

1. INTRODUCTION

It is important to understand which underlying univariate time-series processes may generate earnings per share of listed companies from a practical investment point of view and pure academic perspective. The purpose of the paper is to examine which of the statistical time-series models of random walk and ARIMA types well approximates the behavior of earnings per share for the Polish market. The issue had been investigated in the literature since the late 1960s, mostly for US companies. Various models were examined including the naïve random walk class of models as well as autoregressive integrated moving average type models (Ball & Watts, 1972; Watts, 1975; Griffin, 1977; Foster, 1977; Brown & Rozeff, 1977, 1979b). The results of these studies were mixed and led to ambiguous conclusions. In some works, it was argued that the naive model provided the best results and more advanced mechanical models were not able to

DOI: 10.7172/2353-6845.jbfe.2023.1.2

beat the naive ones, whereas in others different conclusions were drawn. However, for quarterly data, a consensus among researchers arose that ARIMA-type models performed the best (Lorek, 1979; Bathke & Lorek, 1984). Market and academic interest in the subject 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 (Pagach & Warr, 2020).

Because the literature almost exclusively focuses on the US, it is interesting to see how it looks in the other markets. Only a few papers were dedicated to this issue. The author could find only papers focused on companies from Taiwan and the Baltic states (Bao, 1996; Grigaliūnienė, 2013). Unfortunately, the samples of companies used in these papers were of very limited size. From this perspective, it is valuable to examine this problem in the context of the Polish market, which is the deepest among emerging Central European capital markets. This market is not as developed as the US market and has a much shorter history and different institutional framework. Moreover, the good quality of earnings forecasting in these markets is of much higher importance than in the US market because only a small fraction of companies is covered by financial analysts.

It is worth mentioning that all the existing research is limited to the period ending before 2009, i.e. the year prior to the last financial crisis. No paper covers yet the most recent period of stability, i.e. the period between the financial crisis shock and the pandemic shock. From this perspective, this paper is the first to analyze this period. The sample of 267 listed companies and quarterly data for the period from 2010 to 2019 are used for the analysis. The data from Q1 2010 to Q4 2018 are used for the estimation of the time-series model and the period Q1 2019 – Q4 2019 for testing. The robustness check is made to confirm the obtained results using the expanding window approach for the years 2018 and 2017 as hold-out validation samples.

Eight different univariate time-series models are estimated and assessed. The first four models are the naïve time-series models like the random walk model, the random walk with drift model, the seasonal random walk model, and the seasonal random walk with drift model. The next four models are the models of autoregressive integrated moving average type. They are the Griffin-Watts model, the Foster model, the Brown-Rozeff model, and the firm-specific ARIMA model.

Instead of relying on the mean absolute percentage error (MAPE) metric widely used in the previous research, the mean arctangent absolute percentage error (MAAPE) is calculated (Kim & Kim, 2016). The latter overcomes the primary difficulty with the standard MAPE error metric that is related to the explosion of this measure when its denominator is very small, i.e. when actual earnings are close to zero. It is found that the distribution of arctangent absolute percentage forecast errors is similar in all analyzed quarters, which leads to a conclusion that surprisingly the forecast errors do not increase with forecast horizons. The best model, with the lowest rank, is the seasonal random walk (SRW) model across all quarters, which is superior to other models and particularly more advanced ARIMA-type models. This empirical regularity contradicts the results obtained for the US market and studies limited to pre-2008 time periods. The SRW model relatively well captures the behavior of the Polish companies, compared to other models. It turns out that the medians of errors of eight analyzed models differ statistically significantly in almost all quarters.

The superiority of the seasonal random walk model (SRW) implies that the underlying EPS generating process exhibits neither autoregressive nor moving average parts and there is no drift component. The horizontal performance of the stock market index WIG during the analyzed period implies the absence of a trend. In the context of emerging markets, the absence of the moving average part is consistent with the fact that a lower fraction of companies publish the forecasts of their earnings compared to developed markets, and hence not for so many companies past forecast errors result in the correction of the performance of future earnings. The non-existence of the autoregressive part may be related to the dominance of the seasonal component relative to past

DOI: 10.7172/2353-6845.jbfe.2023.1.2

EPS behavior, which might imply that Polish and more generally emerging market companies are more seasonal than those operating in developed markets.

This paper contributes to the literature in four different ways. First, the time-series models are estimated using the most recent data coming from a period of relative earning stability, i.e. ranging from the last financial crisis of 2008–2009 to the pandemic shock of 2020. No academic paper has so far focused on that recent period. Second, the study covers the behavior of emerging market companies, not frequently documented in the literature, like in the papers by Bao (1996) and Grigaliūnienė (2013). Contrary to the US findings, the earnings per share behavior are well described by the seasonal random walk model, whereas in the US the most adequate models are of the more sophisticated ARIMA type. It may result from a different level of advancement as well as an institutional framework of the US market compared to emerging economies. This shows that conclusions drawn for the US might not automatically hold for emerging economies. Also, this result is more important for emerging markets, given that only a small fraction of companies are covered by financial analysts in these markets, contrary to the situation in the US. Third, the research uses a large sample compared to studies mentioned above regarding emerging markets. Consequently, the findings in this paper are supposed to be more statistically meaningful. Fourth, instead of the mean absolute percentage error (MAPE), which is the most popular metric to measure time-series forecast errors, a modification of this measure is used to overcome the standard difficulty of exploding the MAPE metric when earnings are close to or equal to zero. This is done by using the mean arctangent absolute percentage error (MAAPE). Despite the disadvantage described above, the MAPE metric is widely used in literature, e.g. in publications written by Johnson and Schmitt (1974), Lorek (1979), Bathke and Lorek (1984), Collins et al. (1984), Brown et al. (1987), Bao (1996), Lorek and Willinger (2007) and Grigaliūnienė (2013). The paper concludes that for the Polish market the most appropriate model describing earnings per share behavior is the naïve seasonal random walk model. It is consistent with the dominance of a seasonal component over EPS behavior and the fact that only a small fraction of companies publish the forecasts of their earnings.

2. LITERATURE REVIEW

There are four areas of research dedicated to modeling earnings per share using univariate time-series models. The first distinction refers to which earnings are forecasted – annual earnings or quarterly earnings. The research started with annual earnings modeling because this approach required much fewer data points, which was a necessity arising from the computational power at that time. The second cut is based on the statistical techniques used. Three groups of models are applied in the literature starting from naive models relying on random walk processes through the class of autoregressive integrated moving average models, exponential smoothing models, and others.

The literature referring to modeling of companies' earnings using a statistical approach started from the paper by Cragg and Malkiel (1968) in which the authors found that security analysts' predictions performed not much better than those based on past growth rates. Later research was either focused on listed companies' earnings or companies' EPS. Hence, the term earnings is hereafter used interchangeably either for net earnings or for EPS. Beaver (1970) concluded that the underlying process that generates annual earnings was likely to be a mixture of a random walk and a mean-reverting process. Ball and Watts (1972) argued that the measured annual accounting income followed either a submartingale or some very similar process. The famous research by Elton and Gruber (1972) examined the accuracy of forecasts produced by nine mechanical models. It occurred that the additive exponential smoothing with no trend in trend dominated other models. They also found that the differences in forecasts accuracy of mechanical

DOI: 10.7172/2353-6845.jbfe.2023.1.2

models and security analysts' forecasts were not statistically significant. In the work by Johnson and Schmitt (1974), also various mechanical models were tested including naive random walk, moving average model, linear projection model, single double and triple exponential smoothing models, and the accuracy of their forecasts was calculated. The naive model provided the best results and more advanced mechanical models were not able to beat the naive one. Brooks and Buckmaster (1976) applied single, double, and triple exponential smoothing models for different strata of earnings time series, and Albrecht et al. (1977) analyzed various models across three industries (chemical, food, and steel). It is worth emphasizing that only the behavior of annual earnings has been investigated so far.

Watts (1975) focused on modeling quarterly earnings as one of the first. He introduced widely recognized autoregressive integrated moving average (ARIMA) type of models using a method developed some time ago by Box and Jenkins. Later, his findings were extended by Griffin (1977), who accomplished the first parsimonious ARIMA model of quarterly earnings. The model relied on the observation that quarterly earnings could not be adequately described as a random walk or a martingale (sub martingale) and the successive changes in quarterly earnings were not independent. The author concludes that quarterly earnings could be parsimoniously described as a multiplicative combination of two processes: one that reflected adjacent quarters and the other that reflected the seasonal component. The first of the premier models was named the Griffin-Watts (GW) model and was based on the seasonal autoregressive integrated moving average, i.e. SARIMA $(0,1,1)\times(0,1,0)$, framework. The second premier model comes from Foster (1977). He evaluated the predictive ability of six forecasting models: simple and seasonal random walk with or without drift, his model (F) as well as individually estimated Box-Jenkins model (Box and Jenkins, 1976). This firm-specific model is denoted as BJ. After concluding that quarterly earnings did not follow the submartingale process that appeared to adequately describe annual earnings, he found that the model developed by him (F) outperformed other ones. His model referred to the $(1,0,0)\times(0,1,0)$ SARIMA framework. Salamon and Smith (1977) found that there was diversity in time series characteristics of the EPS sequence of individual firms. Finally, the third parsimonious model of the $(1,0,0)\times(0,1,1)$ SARIMA type was published by Brown and Rozeff (1977). Brown and Rozeff (1978) claimed in another paper that their model (BR) and BJ models performed equally well and were superior to other models for short forecast horizons. For longer forecast horizons, accuracy of the BR model deteriorated, but it outperformed BJ models. Brown and Rozeff (1979b) found also in another research that financial analysts' behavior was at a one-quarter-ahead horizon similar to the three primary (BW, F, BR) models. However, the answer to the question of which parsimonious models performed the best was ambiguous. Lorek (1979) indicated that the GW model was the dominant model and three parsimonious models (GW, F, BR) and firm-specific (BJ) models performed better than simplistic random walk models. Hopwood and McKeown (1981) wrote that a transfer function model proposed by them performed the best and BR was the second best. The incorporation of Box-Cox power transformation, which converts non-normal distribution into a normal shape, according to Hopwood et al. (1981), improved on average forecasts made by BW, F, BR, and BJ models. In the work by Bathke and Lorek (1984), it turned out that the BR model dominated the other models across error metrics and quarters. Kao et al. (1996), based on the Dickey-Fuller test, found that net income and EPS series contained a unit root and hence were nonstationary. This was consistent with the hypothesis that income series generally contained both permanent and transitory components. The result showed that the most of quarterly income series contained a substantial moving average part even after seasonality was accounted for.

Some researchers focused on the relation between firm characteristics and forecast accuracy. Bathke et al. (1989) tested the prediction power for large, medium-, and lower-sized firms. Bathke et al. (2004) found also that, while the seasonal random walk (SWR) model did not appear to be descriptively valid for the entire sample of firms, it may, nevertheless, be more appropriate

DOI: 10.7172/2353-6845.jbfe.2023.1.2

for some firms relative to others. Lorek and Willinger (2007) concluded that the choice of an appropriate model was dependent on the business context.

It is noticed that during two decades spanning approximately from 1968 to the late 1980s, many researchers examined whether analysts' forecasts were superior to time-series forecasts. Elton and Gruber found that the differences in forecasts accuracy between forecasts made by mechanical models (especially by those made by the exponential smoothing additive model) and security analysts' forecasts were not statistically significant. Collins and William (1980) wrote that financial analysts provided forecasts more accurately than the statistical models because financial analysts could respond to situations such as strikes or sudden swings in earnings. Conroy and Harris (1987) concluded that, on average, the primary forecasting advantages of analysts over time-series methods appeared to occur over short forecast horizons. The above studies were however based on annual data. Using quarterly time series of earnings, Brown and Rozeff (1979a) argued that at longer horizons the analyst behavior corresponded to autoregressive time-series models rather than moving average models. Hopwood et al. (1981) pointed out that forecasts of quarterly earnings made by time-series models were outperformed by financial analysts. This literature culminated in 1987 with a conclusion in the paper of Brown et al. (1987) that analyst forecasts were superior to time-series forecasts because analysts had an information advantage and a timing advantage. It took two decades to conclude it. Subsequently, there was a sharp decline in research on the properties of times series EPS forecasts. The research by Damodaran (1989) stressed, however, the importance of time-series models to forecast earnings when analyst forecasts were not available. Walther (1997) found that market participants were placing more weight on analyst forecasts relative to time-series models as institutional ownership and analyst coverage increased, which was the proxy for investor sophistication. Bradshaw et al. (2012) point out that the fraction of listed companies in the US uncovered by analysts diminished from 55% in 1980 to around 20% in 2007. It was one of the major reasons for less interest in statically based forecasting of earnings. Recently, Pagach and Warr (2020) re-examined the hypothesis of analysts' forecasts superiority vs. time-series forecasts made by BR and seasonal random walk models by using quarterly data. The general results were consistent with the analysts' dominance; however, more contextual interpretation was suggested. Specifically, they found that for a relatively large number of cases (approximately 40%) ARIMA time-series forecasts of quarterly EPS were equal to or more accurate than consensus analysts' forecasts. Moreover, the percentage of time series superiority increased for longer forecast horizons as the firm size decreased and for high-technology firms. It occurred also that ARIMA models dominated the SRW model.

Mostly, the research was focused on the US-listed companies due to the long earnings history as well as the extensive analyst coverage compared to other markets. Few exceptions to this rule were papers by Bao et al. (1996) and Grigaliūnienė (2013).

In the first of the above papers, the authors referred to the annual earnings of Taiwanese listed companies. The forecast accuracy of a pure mean-reverting process with and without growth component that was a deterministic function of time, random walk, and random walk with drift processes was considered. It appeared that the model that fitted time-series data the most was the random walk model. The second paper suggested that in Baltic countries quarterly earnings followed a simple and seasonal random walk process compared to the three premier models (BW, F, BR). Unfortunately, the samples of companies used in that research are quite small – they consisted only of 48 companies and 8 companies respectively. They were not large enough to draw statistical conclusions.

It is also worth stressing that all the existing research was limited to the period ending before 2009, which is the year of structural change marked by the last financial crisis, and no paper yet covers the most recent period of stability, i.e. the period between the financial crisis shock and the pandemic shock.

DOI: 10.7172/2353-6845.jbfe.2023.1.2

One of the first literature reviews dedicated to modeling earnings using some time-series techniques was provided by Watts and Leftwich (1977). It was followed by Bao et al.'s (1983) work and a very in-depth description of existing research by Bradshaw et al. (2012). One of the lastly published reviews of evolving accomplishments in that field was the paper by Grigaliūnienė (2013).

The literature review and observation indicate that only very few research papers were devoted to the modeling of earnings per share in emerging markets. Moreover, in those papers, the samples of companies used were not of a substantial size. Also, the periods covered in those publications are generally quite old. Hence, my research goal will be to analyze one of the important emerging markets using the most recently available data and having a sizable sample of companies. Apart from that, in almost all existing literature, the mean absolute percentage metric is used to access the accuracy of forecasts. However, this measure has a serious disadvantage to deal with situations when earnings are close to or equal to zero. A modification of this metric will be proposed to address this issue.

3. METHODOLOGY AND DATA

3.1. Methodology

Naïve models

Five naïve time-series models and four seasonal autoregressive integrated moving average (SARIMA) type models are analyzed in the paper. Denoted as Q_t is the realization of *EPS* at the end of quarter *t*. The naïve models include:

1. The random walk model (RW) can be described as:

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

Hence $E_{t-1}(Q_t) = Q_{t-1}$, so the model does not need any estimation of parameters to make the forecasts. To estimate the variance of the disturbance term: $\varepsilon_t = Q_t - Q_{t-1}$, the following calculations have to be made: $\hat{\sigma}^2 = \sum_{t=1}^{T} \frac{(\varepsilon_t - \bar{\varepsilon}_t)^2}{T-1}$, where $\bar{\varepsilon}_t = \sum_{t=1}^{T} \frac{\varepsilon_t}{T}$.

2. The random walk model with drift (RWD) can be described as:

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

Thus $E_{t-1}(Q_t) = \delta + Q_{t-1}$. To make the forecast, we have to estimate the drift parameter as $\hat{\delta} = \bar{\varepsilon}_t$, whereas $\hat{\sigma}^2$ is estimated as above.

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

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

$E_{t-1}(Q_t) = Q_{t-4}$, so the model does not need any estimation of parameters to make the forecasts. To estimate the variance of the disturbance term: $\varepsilon_t = Q_t - Q_{t-4}$, the calculations similar to those described in point 1 have to be made: $\hat{\sigma}^2 = \sum_{t=1}^{T} \frac{(\varepsilon_t - \bar{\varepsilon}_t)^2}{T-1}$, where $\bar{\varepsilon}_t = \sum_{t=1}^{T} \frac{\varepsilon_t}{T}$.

[1] IID – independent, identically distributed.

DOI: 10.7172/2353-6845.jbfe.2023.1.2

4. The seasonal random walk model with drift (SRWD) can be described as:

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

Similarly to the random walk with drift model $E_{t-1}(Q_t) = \delta + Q_{t-4}$. To make the forecast, we have to estimate the drift parameter as $\hat{\delta} = \bar{\varepsilon}_t$, whereas $\hat{\sigma}^2$ is estimated as mentioned in point 4.

SARIMA models

Seasonal autoregressive integrated moving average (SARIMA) models are a class of autoregressive integrated moving average models (ARIMA) with a seasonal component. The next four presented models are of the seasonal autoregressive integrated moving average (SARIMA) type and they generally can be expressed in the following way:

$$\varphi(B)(1-B)^d \Phi(B^S)(1-B)^D Q_t = \theta(B)\Theta(B^S)\varepsilon_t + \theta_0$$

where B and B^S are backshift and seasonal backshift operators, i.e. $BQ_t = Q_{t-1}$ and $B^S Q_t = Q_{t-4}$. The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean, i.e. $N(0, \sigma^2)$. $\varphi(B)$ and $\Phi(B^S)$ are polynomials referring to the autoregressive part, respectively: $\varphi(B) = 1 - \varphi_1 B - \ldots - \varphi_p B^p$ and $\Phi(B^S) = 1 - \varphi_1(B^S) - \ldots - \varphi_P(B^S)^P$. $\theta(B)$ and $\Theta(B^S)$ are polynomials describing moving average parts, so $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$, where θ_0 is a constant term and $\Theta(B^S) = 1 - \Theta_1(B^S) - \ldots - \Theta_Q(B^S)^Q$. D and d parameters are the degrees of freedom of ordinary and seasonal parts required to eliminate the so-called “unit root” problem, i.e. to achieve stationarity of the Q_t series. Thus, any model can be described of order $(p, d, q)\times(P, D, Q)$, where parameters p, P describe the autoregressive, seasonal autoregressive part, parameters q, Q describe the moving average, seasonal moving average part, and parameters d, D describe the order of differencing, seasonal differencing. Parameters of the SARIMA model are estimated using the maximum likelihood estimation (MLE) method. The MLE estimates for SARIMA parameters are consistent, normally distributed, and asymptotically efficient (Asteriou & Hall, 2011).

The considered SARIMA models are described as follows:

5. The Griffin-Watts (GW) model is the SARIMA model of order $(0, 1, 1)\times(0, 1, 1)$ without constant term and can be described as:

$$Q_t = Q_{t-1} + (Q_{t-4} - Q_{t-5}) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \Theta_1\varepsilon_{t-4} - \theta_1\Theta_1\varepsilon_{t-5}$$

so the forecast is:

$$E_{t-1}(Q_t) = Q_{t-1} + (Q_{t-4} - Q_{t-5}) - \theta_1\varepsilon_{t-1} - \Theta_1\varepsilon_{t-4} - \theta_1\Theta_1\varepsilon_{t-5}.$$

6. The Foster (F) model is the SARIMA model of order $(1, 0, 0)\times(0, 1, 0)$ with constant term and can be written in the following way:

$$Q_t = Q_{t-4} + \varphi_1(Q_{t-1} - Q_{t-5}) + \varepsilon_t + \theta_0$$

and the forecast is given as:

$$E_{t-1}(Q_t) = Q_{t-4} + \varphi_1(Q_{t-1} - Q_{t-5}) + \theta_0.$$

DOI: 10.7172/2353-6845.jbfe.2023.1.2

7. The Brown-Rozeff (BR) model is the SARIMA model of order (1, 0, 0)×(0, 1, 1) without constant term and is formulated as follows:

$$Q_t = Q_{t-4} + \varphi_1(Q_{t-1} - Q_{t-5}) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \Theta_1\varepsilon_{t-4}$$

which implies the following forecast:

$$E_{t-1}(Q_t) = Q_{t-4} + \varphi_1(Q_{t-1} - Q_{t-5}) - \Theta_1\varepsilon_{t-4}.$$

8. The firm-specific (BJ) model in which parameters $(p, d, q)\times(P, D, Q)$, as well as the constant term θ_0, are chosen individually for every company. To determine the orders of differencing d and D, the KPSS test described below is performed. The choice of the most appropriate model type is determined by the lowest Akaike's information criterion (AIC)[2] (Asteriou & Hall, 2011). The individual SARIMA models are selected using the stepwise procedure described by Hyndman and Khandakar (2008).

Forecasts accuracy given by the above models is measured in two ways, using mean percentage absolute error and average rank of error.

Stationarity test

To apply the above models, a time series needs to be stationary. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) nonparametric test of stationarity is used (Kwiatkowski et al., 1992) to establish the difference of that time series over 4 quarters concerning the training period.

$$H_0 : \textit{time series is stationary}$$

Mean arctangent absolute percentage error (MAAPE).

Let's call by $A_1^i, \ldots, A_4^i$ the actual EPS realized in the *1-st*,..., *4-th* quarter respectively of 2019 for the *i-th* firm. $F_1^i, \ldots, F_4^i$ are forecasts of this variable in the above periods. An absolute percentage error (APE) of the forecasts for an *i-th* individual company in the *j-th* quarter of 2019 is defined as:

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

Based on this error metric, the so-called mean absolute percentage error (MAPE) is defined, which is widely used in the existing research. This is the most popular measure of a time-series forecast; however, APE has a significant disadvantage: it produces infinite or undefined values when the actual values are zero or close to zero, which is a common occurrence in the forecasting of earnings. If the actual values are very small (usually less than one), APE yields extremely large percentage errors (outliers), while zero actual values result in infinite APEs. To overcome this difficulty, Kim and Kim (2016) introduced a modified APE measure called arctangent absolute percentage error, which is a novel approach in the literature:

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

because *arctan* is a function transforming [–∞, +∞] interval into [–π/2, π/2] interval.

[2] Bayesian information criterion (BIC) has been used alternatively for model selection. This criterion prevents overfitting, which may arise in the case of AIC. The conclusions remain valid regardless of the information criterion being used.

DOI: 10.7172/2353-6845.jbfe.2023.1.2

Hence, the mean arctangent absolute percentage error (MAAPE) for the *i-th* company across all 4 quarters can be written as:

$$MAAPE^i = \frac{1}{4}\sum_{j=1}^{4} AAPE^i_j = \frac{1}{4}\sum_{j=1}^{4} arctan\left(\left|\frac{A^i_j - F^i_j}{A^i_j}\right|\right).$$

And the mean arctangent absolute percentage error (MAAPE) for the *j-th* quarter across all I companies in the sample can be expressed as:

$$MAAPE_j = \frac{1}{I}\sum_{i=1}^{I} AAPE^i_j = \frac{1}{I}\sum_{i=1}^{I} arctan\left(\left|\frac{A^i_j - F^i_j}{A^i_j}\right|\right).$$

Thus, the mean arctangent absolute percentage error (MAAPE) across all 4 quarters and across all I companies in the sample is given by the formula:

$$MAAPE = \frac{1}{I}\sum_{i=1}^{I} MAAPE^i = \frac{1}{4}\sum_{j=1}^{4} MAAPE_j = \frac{1}{4I}\sum_{i=1}^{I} arctan\left(\left|\frac{A^i_j - F^i_j}{A^i_j}\right|\right).$$

For the described models, forecasts are made and for the *m-th* model, $MAAPE(m)_1$,…, $MAAPE(m)_4$ as well as $MAAPE(m)$ are calculated.

The average rank of error

For every firm and quarter combination, absolute percentage errors of the above-mentioned models are ranked. The model with the lowest error is given a rank of 1 and the model with the highest error is given a rank of 8. Then, the average rank of each model across all firms for 1-st quarter-ahead,…, 4-th quarter-ahead forecasts is calculated together with the average rank across 4 quarters and across all companies. Denoted by $AAPE(m)^i_j$, the arctangent absolute percentage error of the forecast for an i-th individual company in the *j-th* quarter of 2019 given by the *m-th* model and $R(m)^i_j$ is the rank of that forecast, where $m = 1,…, 8$. Hence, the average rank of the *m-th* model and the *i-th* company across all 4 quarters can be written as:

$$\bar{R}(m)^i = \frac{1}{4}\sum_{j=1}^{4} R(m)^i_j.$$

The average rank of the *m-th* model and the *j-th* quarter across all I companies in the sample can be expressed as:

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

The average rank of the *m-th* model across all I companies in the sample and all across 4 quarters can be expressed as:

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

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

DOI: 10.7172/2353-6845.jbfe.2023.1.2

The Kruskal-Wallis test

Then, the Kruskal-Wallis one-way H-test (Corder & Foreman, 2009) is made. This is a nonparametric test that avoids difficulties concerning the potential normality of errors. The null hypothesis is that the median AAPEs of all models are equal, i.e the average ranks of 8 models are the same. This is calculated for respective quarters as well as for all forecast quarters and Kruskal-Wallis H statistics with their respective p-values are calculated.

$$H_0 : \textit{medians of AAPEs of all } 8 \textit{ models are the same}$$

The null hypothesis is rejected when the p-value of Kruskal-Wallis H statistics is greater than the assumed significance level[3].

The Wilcoxon test

As the last one, the paired comparison of forecast errors is performed using the nonparametric two-sided Wilcoxon test (Wilcoxon, 1945) to assess the equality of median absolute percentage errors of various models. For each quarter 1,…,4 and all quarters, separate tables are calculated in which above the diagonal p-values of Wilcoxon statistic are presented for all model pairs.

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

The null hypothesis that medians of absolute percentage errors are the same is rejected when the respective p-value is lower than the assumed significance level. It is important that the test is quite 'robust' and does not require any specific assumptions about a probability distribution apart from symmetricity of the difference in scores and independence of observations.

3.2. Data

The Polish stock market is the deepest among counties that joined the European Union after 2004. At the end of 2021, the capitalization of the Warsaw Stock Exchange was USD 197 bn with 774 listed companies and was the largest in the region. Polish stocks are also not so widely covered by financial analysts as the US market or even Western European companies. In Poland, for only a small fraction of companies, EPS forecasts for the next year are released, so time-series models providing a credible forecast would be of paramount importance. At the end of the analyzed period, i.e. 2019, only around 20% out of 711 listed companies were covered by financial analysts. I focus on earnings per share (EPS) data series, since this measure is merger and split resistant. The data source is EquityRT[4], which is a financial analysis platform. The behavior of earnings per share (EPS) of firms listed on the Warsaw Stock Exchange is analyzed spanning from Q1 2010 to Q4 2019, i.e. between two structural shifts of the processes driving earnings. The first event is the financial crisis of 2008 and the subsequent decline in GDP growth of the Polish economy in 2009. The second one is the beginning of the COVID-19 pandemic followed by the lockdown of the economy and a sharp fall in GDP growth in Q1 2020. The data for the period Q1 2010 – Q4 2018 (36 quarters) are used for the estimation of various models, whereas the data from Q1 2019 to Q4 2019 are used as a hold-out validation sample for testing forecast accuracy of 1 quarter-ahead, 2 quarters-ahead, 3 quarters-ahead, and 4 quarters-ahead forecasts. Hence, the companies for which EPS are forecasted require sufficiently long time series (40 observations for training and testing) of earnings and thus are subject to survivorship bias. This bias is however characteristic of the choice of companies to use any more advanced time-series model that requires a series of

[3] It is assumed at 0.05 level.

[4] EquityRT is a product of Turkish company RASYONET.

DOI: 10.7172/2353-6845.jbfe.2023.1.2

data that is long enough. To assert comparability of results, it is required – even for naïve models – to have the same data sample as for more advanced ones. Hence, firms in the sample are likely to be larger and older than the average. Moreover, I decided to eliminate only stocks with splits/reverse splits because such operations influence EPS behavior substantially. There were only cases of splits/reverse splits for 12 companies in the analyzed period.

To validate that results are not specific for 2019 year, other years are chosen as hold-out samples. The models are estimated using the expanding window approach, i.e. the sample Q1 2010 – Q4 2017 is used for their estimation and Q1 2018 – Q4 2018 for their testing. Then, the same procedure is applied taking the year 2017 to validate the results.

After imposing a full time window 2010–2019 coverage and excluding splits, there are 267 such companies on the market. In various studies mentioned in the previous section, firms which were subject to government rate regulation, like utilities and financial sectors, were eliminated from consideration. Because I cannot find a clear reason why the time-series methodology cannot be applied in these cases, I do not exclude them from the sample. Moreover, in many cases, it is hard to determine to what extent the government regulations are shaping the revenue and what portion of revenues can be attributed to market behavior.

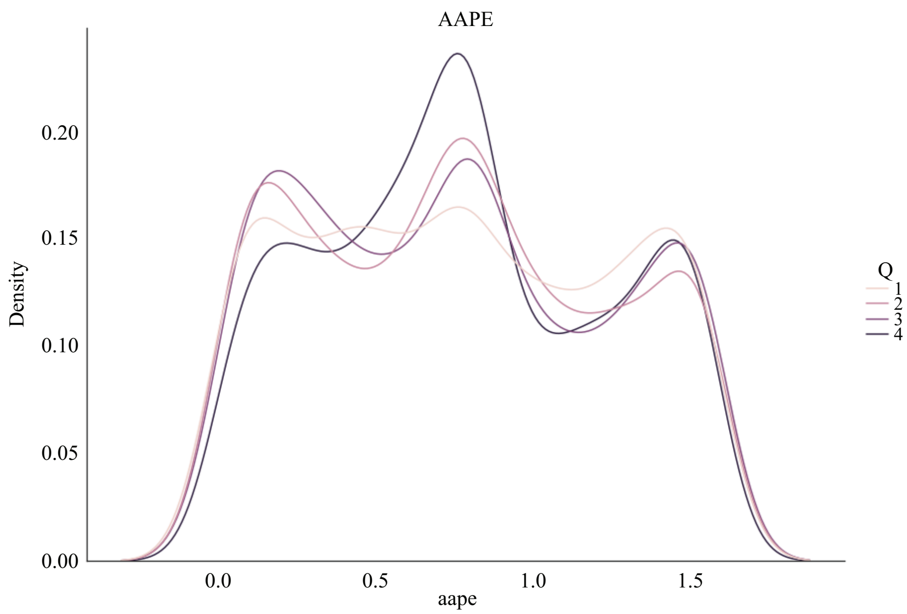
4. RESULTS

4.1. Empirical findings

The time series were analyzed on a level scale. At the beginning of this study, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) stationarity test is applied to the difference over 4 quarters for the analyzed time series to verify if SARIMA type models can be applied. In all cases, the null hypothesis of stationarity cannot be rejected.

Figure 1 suggests that kernel density estimators of arctangent absolute percentage forecast errors are pretty similar in all analyzed quarters. Surprisingly, it implies that forecast errors do not increase with forecast horizons. It is confirmed by the behavior of MAAPE for studied models in Table 1.

Fig. 1.
The kernel density estimators of arctangent absolute percentage errors for forecast quarters



DOI: 10.7172/2353-6845.jbfe.2023.1.2

The model that performs the best and has the lowest rank in respective quarters as well as for all quarters is the seasonal random walk (SRW) model, which is presented in Table 1. The model which in turn performs the worst and has the highest ranks is the seasonal random walk with drift (SRWD) model. Almost the same holds for mean arctangent absolute percentage errors (MAAPEs) except for the 2^{nd} and 4^{th} quarters. In the 2^{nd} and 4^{th} quarters, the models with the lowest mean error are the company specific SARIMA (BJ) models. However, average prediction errors of these models do not differ substantially from the seasonal random walk model (i.e., by only 0.01) in those quarters. The model which in turn performs the worst and has the highest MAAPE is the random walk with drift (RWD) model. It happens in all quarters but not in the 4^{th} quarter as well as jointly for all quarters. In the 4^{th} quarter, the model with the highest rank is the Griffin-Watts (GW) model. The conclusion is that for the Polish market and the forecast period Q1–Q4 2019, the naïve seasonal random walk (SRW) model performs better than more advanced SARIMA type models. This is consistent with the findings of Bao (1996) for Taiwan and Grigaliūnienė (2013) for the Baltic countries that the most appropriate for these markets were naïve models.

Table 1 presents the results of nonparametric Kruskal-Wallis H statistics and its p-value, which does not require a normality assumption. The null hypothesis is that medians of arctangent absolute percentage errors (AAPEs) of all 8 models are the same. The test is made for all quarters respectively and for the entire period. The test shows that the null hypothesis can be rejected in all cases except the 4^{th} quarter. It might derive from the fact that dispersion of ranks between the best and the worst model is much greater for these quarters than for the 4^{th} quarter. In this quarter, the hypothesis that all models generate a similar median of errors statistically cannot be rejected.

In the next step, it is checked if the errors of the best model are statistically significantly different from the results of other models. To do so, the Wilcoxon test nonparametric is calculated for all model pairs. The null hypothesis states that medians of arctangent absolute percentage errors (AAPEs) of a selected pair of models are the same. Tables from 2 to 5 present p-values of the test for all combinations of model pairs in respective quarters. In Table 6, the results for all joint quarters are analogously displayed. The test confirms that the seasonal random walk (SRW) model produces a median of errors statistically significantly lower than other models in the 1^{st} quarter. The only exception to this rule might be the firm-specific SARIMA (BJ) model. It is worth noting that the p-value for the combination of SRW and BJ models is 0.0487, which is only slightly below the assumed significance level of 0.05. Hence, the null hypothesis cannot be rejected at the 0.01 significance level. So, it can be said that BJ forecast errors given by these two models are not so statistically different. In the 2^{nd} quarter, the median of errors of the SRW model are not statistically different from those of the BR (Brown-Rozeff) and BJ models. In the 3^{rd} quarter, only the median of errors of the BR model does not statistically differ from the best SRW model. In the 4^{th} quarter, we can't reject the null hypothesis that the median errors are different for the SRW and BJ models at 0.05 level of significance and for the SRW and BR models at 0.01 significance level. With respect to all quarters, only the SRW and BJ models generate statistically not different medians of errors. It emerges from the above analysis that mainly medians of errors of the firm-specific (BJ) model are statistically the same as the best seasonal random walk (SRW) model in most periods.

It is worth emphasizing that the seasonal random walk model (SRW) is a special case of the Foster (F), Brown-Rozeff (BR), and firm-specific models that assume quarterly seasonality and set all parameters equal to zero. The fact that mean arctangent absolute percentage errors of the SRW model are substantially lower than those of the above-mentioned ARIMA models might emerge from how the models are estimated. Minimization of the maximum likelihood function, which is a standard technique for the model estimation, is not fully consistent with the minimization of any type of absolute percentage error including mean arctangent absolute percentage error.

The superiority of the seasonal random walk model (SRW) implies that the underlying EPS generating process exhibits neither autoregressive nor moving average parts and there is no drift. It means that any older or shorter history than exactly one-year history has no substantial

DOI: 10.7172/2353-6845.jbfe.2023.1.2

influence. The horizontal performance of the stock market index WIG during the analyzed period implies the absence of a trend. The absence of the moving average part means non-existence of an error correction mechanism by which past errors influence the behavior of future earnings. Past errors are deviations of actual EPS numbers from the data forecasted by the model. In the context of emerging markets, it is consistent with the fact that a small fraction of companies publish forecasts of their earnings compared to developed markets. Hence, for not so many companies from these markets, past forecast errors result in the correction of the performance of future earnings. Non-existence of the autoregressive part may in turn be related to the dominance of the seasonal component relative to past EPS behavior, which might imply that Polish and, more generally, emerging market companies are more seasonal than those operating on developed markets. This hypothesis could be examined in further research.

4.2. Robustness check

Table 7 confirms that the seasonal random walk model is characterized by the lowest rank, i.e. gives the best results not only in 2019, but also in 2018 and 2017. Also high values of Kruskal-Wallis H statistics reject the null hypothesis that the model errors are not statistically different in those years. Additionally, the Wilcoxon test is made for all model pairs with the seasonal random walk model and p-values across different years are presented in Table 8. In 2017, they all are lower than the assumed significance level, so the errors of all models are statistically different from the errors of the seasonal random walk model. In 2018, similarly to 2019, only the results of the firm-specific model (BJ) are statistical not different from those of the seasonal random walk model. Hence, it emerges from the above that the superiority of the seasonal random walk model seems to be invariant in time.

5. CONCLUSIONS

The paper describes the forecasting characteristics of eight univariate time-series models applied for quarterly earnings per share of 267 Polish companies in the period 2010–2019. It turns out that counter-intuitively forecast errors do not increase with forecast horizons. The best model, with the lowest rank, is the seasonal random walk (SRW) model across all quarters, which describes quite well the behavior of the Polish market compared to other models. This is consistent with the findings of Bao (1996) for Taiwan and Grigaliūnienė (2013) for the Baltic countries that the most appropriate for these markets were naïve models. The medians of errors of the analyzed models differ statistically significantly in almost all quarters. Medians of errors of the firm-specific (BJ) model are statistically not different from the best seasonal random walk (SRW) model for most analyzed periods. The superiority of the seasonal random walk model seems to be invariant in time.

This finding is consistent with the absence of drift and autoregressive and moving average parts. It might be related to the performance of the stock market index during the analyzed period, the fact that a lower fraction of emerging market companies publish forecasts of their earnings compared to developed markets, and the dominance of the seasonal component compared to past EPS behavior. The above hypothesis can be verified in further research.

Concerning a future research agenda, it would be interesting to verify if another class of time-series models relying on exponential smoothing provides more precise forecasts than naïve random walk models. The relation between forecast efforts and firm size should also be examined. The business context described by the sector in which a company operates may also play an important role in assessing which model most accurately forecasts earnings per share. Additionally, a seasonal pattern described by the SRW model might imply an investment strategy based on this pattern. In further research, such a strategy could be tested in terms of its capability to beat the market. These findings might contradict the weak form of efficient market hypothesis (EMH).

DOI: 10.7172/2353-6845.jbfe.2023.1.2

Table 1
Summary statistics on forecast errors and Kruskal-Wallis test for 2019 quarters

model	Quarters								All Quarters	
	Q1 MAAPE	Q1 Average Rank	Q2 MAAPE	Q2 Average Rank	Q3 MAAPE	Q3 Average Rank	Q4 MAAPE	Q4 Average Rank	MAAPE	Average Rank
RW	0.89	5.21	0.80	4.68	0.83	5.01	0.74	3.97	0.81	4.72
RWD	0.92	5.81	0.84	5.26	0.88	5.59	0.79	4.96	0.85	5.40
SRW	0.66	3.69	0.70	3.98	0.65	3.74	0.74	3.97	0.69	3.85
SRWD	0.70	4.03	0.73	4.35	0.73	4.25	0.80	4.67	0.74	4.33
GW	0.78	4.51	0.80	4.81	0.77	4.52	0.82	4.84	0.79	4.67
F	0.77	4.38	0.75	4.49	0.75	4.35	0.80	4.75	0.77	4.49
BR	0.75	4.16	0.74	4.24	0.71	4.14	0.80	4.62	0.75	4.29
BJ	0.71	4.20	0.69	4.19	0.74	4.40	0.73	4.23	0.72	4.25
H statistics		63.92		19.79		38.18		10.79		36.56
p-value		0.00		0.01		0.00		0.15		0.00

Table 2
P-values of paired Wilcoxon test of forecast errors in Q1 2019

model	RWD	SRW	SRWD	GW	F	BR	BJ
RW	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
RWD		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
SRW			0.0218	0.0001	0.0000	0.0005	0.0487
SRWD				0.0052	0.0129	0.0887	0.5389
GW					0.4606	0.0609	0.0240
F						0.7939	0.1090
BR							0.1573

Table 3
P-values of paired Wilcoxon test of forecast errors in Q2 2019

model	RWD	SRW	SRWD	GW	F	BR	BJ
RW	0.0000	0.0004	0.0210	0.3844	0.0412	0.0066	0.0004
RWD		0.0000	0.0003	0.0541	0.0012	0.0001	0.0000
SRW			0.0036	0.0002	0.0001	0.5705	0.9455
SRWD				0.0215	0.2108	0.9248	0.2197
GW					0.0763	0.0010	0.0007
F						0.4492	0.0856
BR							0.4630

DOI: 10.7172/2353-6845.jbfe.2023.1.2

Table 4
P-values of paired Wilcoxon test of forecast errors in Q3 2019

model	RWD	SRW	SRWD	GW	F	BR	BJ
RW	0.0000	0.0000	0.0001	0.0028	0.0001	0.0000	0.0003
RWD		0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
SRW			0.0001	0.0020	0.0000	0.1113	0.0005
SRWD				0.1770	0.2032	0.2569	0.5654
GW					0.1947	0.0441	0.6852
F						0.1285	0.9419
BR							0.2883

Table 5
P-values of paired Wilcoxon test of forecast errors in Q4 2019

model	RWD	SRW	SRWD	GW	F	BR	BJ
RW	0.0000	0.0000	0.0000	0.0011	0.0000	0.0213	0.7377
RWD		0.0000	0.4202	0.1339	0.8936	0.4939	0.0785
SRW			0.0000	0.0011	0.0000	0.0213	0.7377
SRWD				0.1578	0.6280	0.8037	0.0281
GW					0.2343	0.0502	0.0045
F						0.8973	0.0196
BR							0.0547

Table 6
P-values of paired Wilcoxon test of forecast errors for all quarters 2019

model	RWD	SRW	SRWD	GW	F	BR	BJ
RW	0.0000	0.0000	0.0000	0.0163	0.0003	0.0000	0.0000
RWD		0.0000	0.0000	0.0002	0.0000	0.0000	0.0000
SRW			0.0001	0.0000	0.0000	0.0042	0.0930
SRWD				0.0066	0.0183	0.7726	0.2826
GW					0.0984	0.0008	0.0007
F						0.0282	0.0131
BR							0.3392

DOI: 10.7172/2353-6845.jbfe.2023.1.2

Table 7
Summary statistics on forecast terrors and Kruskal-Wallis for all quarters 2017–2019

model	2017		2018		2019	
	MAAPE	Average Rank	MAAPE	Average Rank	MAAPE	Average Rank
RW	0.83	4.78	0.86	4.97	0.81	4.72
RWD	0.85	5.42	0.88	5.60	0.85	5.40
SRW	0.69	3.86	0.71	3.81	0.69	3.85
SRWD	0.72	4.29	0.76	4.27	0.74	4.33
GW	0.79	4.75	0.80	4.62	0.79	4.67
F	0.75	4.45	0.78	4.41	0.77	4.49
BR	0.74	4.24	0.75	4.19	0.75	4.29
BJ	0.72	4.21	0.73	4.14	0.72	4.25
H statistics		32.07		40.28		36.56
p-value		0.00		0.00		0.00

Table 8
P-values of paired Wilcoxon test of forecast errors for all quarters 2017–2019 and SRW model

year	model	RWD	SRWD	GW	F	BR	BJ
2017	SRW	0.0000	0.0013	0.0000	0.0000	0.0007	0.0233
2018	SRW	0.0000	0.0000	0.0000	0.0000	0.0488	0.1686
2019	SRW	0.0000	0.0001	0.0000	0.0000	0.0042	0.0930

Reference list

Albrecht, W. S., Larry, L. L., & McKeown, J. C., (1977). The time-series properties of annual earnings. *Journal of Accounting Research*, *15*(2), 226–244. https://doi.org/10.2307/2490350
Asteriou, D., & Hall, S. G. (2011). ARIMA models and the Box-Jenkins methodology. In D. Asteriou & S. G Hall (Eds.), *Applied Econometrics* (2nd ed., pp. 265–286). Palgrave MacMillan.
Ball, R., & Watts, R. (1972). Some time series properties of accounting income. *The Journal of Finance*, *27*(3), 663–681. https://doi.org/10.1111/j.1540-6261.1972.tb00991.x
Bao, B.H., Bao, D.H., & Firth, M. (1996). A note on the time series behavior of earnings per share data of Taiwanese firms. *Journal of Business Finance & Accounting*, *23*(4), 641–650. https://doi.org/10.1111/j.1468-5957.1996.tb01030.x
Bao, D. H., Lewis, M.T., Lin, W.T., & Manegold, J.G. (1983). Applications of time-series analysis in accounting: A review. *Journal of Forecasting*, *2*(4), 405–423. https://doi.org/10.1002/for.3980020409
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.
Bathke Jr., A. W., Lorek, K. S., & Willinger, G. L. (1989). Firm-size and the predictive ability of quarterly earnings data. *The Accounting Review*, *64*(1), 49–68.
Bathke, Jr. A. W., Lorek, K. S., & Willinger, G. L. (2004). Differential earnings behavior and the security market assessment of variation in seasonal earnings patterns. *Journal of Accounting, Auditing & Finance*, *19*(4), 463–482. https://doi.org/10.1177/0148558X0401900407

Beaver, W. H. (1970). The time series behavior of earnings. *Journal of Accounting Research*, *8*, 62–99. https://doi.org/10.2307/2674693

Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. Holden-Day.

Bradshaw, M. T., Drake, M. S., Myers, J. N., & Myers, L. A. (2012). A re-examination of analysts' superiority over time-series forecasts of annual earnings. *Review of Accounting Studies*, *17*(4), 944–968. https://doi.org/10.1007/s11142-012-9185-8

Brooks, L. D., & Buckmaster, D. A. (1976). Further evidence of the time series properties of accounting income. *Journal of Finance*, *31*(5), 1359–1373. Wiley-Blackwell. https://doi.org/10.1111/j.1540-6261.1976.tb03218.x

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

Brown, L. D., & Rozeff, M. S. (1977). *Univariate time-series models of quarterly accounting earnings per share: A proposed premier model* (Working Paper No. 77–27). College of Business Administration, University of Iowa.

Brown, L. D., & Rozeff, M. S. (1979a). Adaptive expectations, time-series models, and analyst forecast revision. *Journal of Accounting Research*, *17*(2), 341–351. https://doi.org/10.2307/2490508

Brown, L. D., & Rozeff, M. S. (1979b). Univariate time-series models of quarterly accounting earnings per share: A proposed model. *Journal of Accounting Research*, *17*(1), 179–189. https://doi.org/10.2307/2490312

Collins, W. A., Hopwood, W. S., & McKeown, J. C. (1984). The predictability of interim earnings over alternative quarters. *Journal of Accounting Research*, *22*(2), 467–479. https://doi.org/10.2307/2490659

Collins, W. A., & William, S. H. (1980). A multivariate analysis of annual earnings forecasts generated from quarterly forecasts of financial analysts and univariate time-series models. *Journal of Accounting Research*, *18*(2), 390–406. https://doi.org/10.2307/2490585

Conroy, R., & Harris, R. (1987). Consensus forecasts of corporate earnings: analysts' forecasts and time series methods. *Management Science*, *33*(6), 725–738. https://doi.org/10.1287/mnsc.33.6.725

Corder, G. W., & Foreman, D. I. (2009). *Nonparametric statistics for non-statisticians*. John Wiley & Sons. https://doi.org/10.1002/9781118165881

Cragg, J. G., & Malkiel, B. G. (1968). The consensus and accuracy of some predictions of the growth of corporate earnings. *The Journal of Finance*, *23*(1), 67–84. https://doi.org/10.1111/j.1540-6261.1968.tb02998.x

Damodaran, A., & Landsman, W. R. (1989). Using shrinkage estimators to improve upon time-series model proxies for the security market's expectation of earnings. *Journal of Accounting Research*, *27*(1), 97–115. https://doi.org/10.2307/2491209

Elton, E. J., & Gruber, M. J. (1972). Earnings estimates and the accuracy of expectational data. *Management Science*, *18*(8), B409–B424. https://doi.org/10.1287/mnsc.18.8.B409

Foster, G. (1977). Quarterly accounting data: Time-series properties and predictive-ability results. *The Accounting Review*, *52*(1), 1–21.

Griffin, P. (1977). The time-series behavior of quarterly earnings: Preliminary evidence. *Journal of Accounting Research*, *15*(1), 71–83. https://doi.org/10.2307/2490556

Grigaliūnienė, Ž. (2013). Time-series models forecasting performance in the Baltic stock market. *Organizations and Markets in Emerging Economies*, *4*(1), 104–120. https://doi.org/10.15388/omee.2013.4.1.14261

Hyndman, R. D., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27*(3), 1–22. https://doi.org/10.18637/jss.v027.i03

Hopwood, W. S., & McKeown, J. C. (1981). An evaluation of univariate time-series earnings models and their generalization to a single input transfer function. *Journal of Accounting Research*, *19*(2), 313–322. https://doi.org/10.2307/2490867

Hopwood, W. S., McKeown, J. C., & Newbold, P. (1981). Power transformations in time-series models of quarterly earnings per share. *The Accounting Review*, *56*(4), 927–933.

Johnson, T. E., & Schmitt, T. G. (1974). Effectiveness of earnings per share forecasts. *Financial Management*, *3*(2), 64–72. https://doi.org/10.2307/3665292

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, *32*(3), 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

Kao, Ch., Lee, Ch., & Wu, Ch. (1996). Time-series properties of financial series and implications for modeling. *Journal of Accounting, Auditing & Finance*, *11*(2), 277–303. https://doi.org/10.1177/0148558X9601100207

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, *54* (1–3), 159–178. https://doi.org/10.1016/0304-4076(92)90104-Y

Lorek, K. S. (1979). Predicting annual net earnings with quarterly earnings time-series models. *Journal of Accounting Research*, *17*(1), 190–204. https://doi.org/10.2307/2490313

DOI: 10.7172/2353-6845.jbfe.2023.1.2

Lorek, K. S., & Willinger, G. L. (2007). The contextual nature of the predictive power of statistically-based quarterly earnings models. *Review of Quantitative Finance and Accounting*, *28*(1), 1–22. https://doi.org/10.1007/s11156-006-0001-z

Pagach, D. P., & Warr, R. S. (2020). Analysts versus time-series forecasts of quarterly earnings: A maintained hypothesis revisited. *Advances in Accounting*, *51*. https://doi.org/10.1016/j.adiac.2020.100497

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. https://doi.org/10.1111/j.1540-6261.1977.tb03377.x

Walther, B. R. (1997). Investor sophistication and market earnings expectations. *Journal of Accounting Research*, *35*(2), 157–179. https://doi.org/10.2307/2491358

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

Watts, R. L., & Leftwich, R. W. (1977). The time series of annual accounting earnings. *Journal of Accounting Research*, *15*(2), 253–271. https://doi.org/10.2307/2490352

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 80–83. https://doi.org/10.2307/3001968

DOI: 10.7172/2353-6845.jbfe.2023.1.2