Comparison of Mechanical Earnings Forecast Models

Hongxin Lei

Master Thesis at the Department of Corporate Finance

First Supervisor:

Luca Brunke, M.Sc.

Second Supervisor:
Univ.-Prof. Dr. Dieter Hess

Submitted in the Master Program

International Management

At the Faculty of Management, Economics and Social Sciences
of the University of Cologne

Cologne, 31.05.2021

Candidate:
Hongxin Lei
Grügürtelstraße 118
50996 Köln
EMail-Adresse:karolina76@163.com
Student Number: 7348017

Contents

List of Tables

List of Abbreviations

AC	working capital accruals
CPI	consumer price index Square
MIDAS	mixed data sampling regression methods
OLS	ordinary least square
TACC	total accruals

1 Introduction

The earnings forecast matters much to the investment world. Gerakos and Gramacy (2013) mentioned that an earnings forecast is an input for estimating *implied capital cost*, which is in turn an input for corporate valuation. Analysts often implement forecasting using fundamental analysis based on ”forecasting frameworks” which take industry, strategy, financial information, etc., into consideration.[1] Although analysts often produce accurate earnings forecast, [2] analysts are easily exposed to conflict of interests or simple optimism, which therefore raises the concerns about the real accuracy of the analyst forecasts. Moreover, some small companies that are supposed to be evaluated are not usually assessed because analysts only cover specific scopes of large companies[3].

As a result, academic researchers attach more weight to model forecast than analyst forecast. Mechanical models often sum up the relevant financial factors into one regression-based forecasting function to generate the regression results compared to actual earnings and analyst's forecasts. In addition to discussing the accuracy of model forecasts, most previous studies use the forecasted earnings to derive the *implied capital cost*, which is proven to be a proper proxy for stocks' expected return and thus an essential variable for corporate valuation. But this paper limits itself within earnings forecast, discussing exclusively the comparison of mechanical earnings forecast models.

Ball and Watts (1972) find that the accounting income follows a ”sub martingale” process, which means that ” the conditional expectation of the next accounting income in the sequence is equal to the present value.” This implies that accounting income possesses the properties of a time series. The finding serves as a basis for further development of time-series forecast models. For example, Box and Jenkins (1976) discuss using time series regression to model future earning and propose the Box Jenkins method for the model's buildup (henceforth the BJ method/model). Based on the concept of the BJ method, lots of researchers extend and modify the single form of the BJ model to explain the time series properties of quarterly earning per share[4]. The modified time series models can be summarized to mainly include a set of five simple-version firm-specific BJ models, and three parsimonious BJ

[1]See Lambert, Matolcsy and Wyatt (2009)
[2]See O'brien (1988); Hou, Van Dijk and Zhang (2012)
[3]See Evans, Njoroge and Yong (2017)
[4]See Foster (1977), Brown and Rozeff (1979)

models[5]. With respect to forecasts performance, time-series models show good in-sample forecasts. However,these models often fail in forecast ability when applied to out-sample observations (O'Brien 1988). Fama and French (2000) attribute the weak predictability to limited samples in these models, which leads to representativity loss. The samples used in these models are often small with large survivorship bias.

Due to endogeneity problems that could plague time-series model, cross-sectional models have outperformed time-series models. The cross-sectional regression can diminish the effects of survivorship bias and allows for a larger set of empirical samples compared to time-series models. From the preceding, Fama and French (2000) discuss the reasons behind the underperformance of time series models. They carried out regressions on large samples of time-series cross-sectional data to predict profitability/earnings. They conclude that a company's profitability turns out to be diluted under a good financial situation and vice versa, indicating the mean-reverting characteristic of the profitability/earnings. In addition, they find that negative changes and extreme changes in earnings seem to reverse faster than the modeled mean reversion.

Motivated by Fama and French (2000) and Hou, Van Dijk et al. (2010), who have demonstrated that the cross-sectional models perform well capturing the variation in companies' expected profitability, Hou et al. (2012) use the large set of cross-sectional data to model the future earnings (henceforth the HVZ model). They show that their models exhibit higher earnings response coefficient, but nevertheless less accuracy than analysts' forecast. The inaccuracy is further confirmed by Gerakos and Gramacy (2013) and Li and Mohanram (2014) (henceforth as LM (2014)) who find that a simple model of random walk computes even better forecasts results than the HVZ model. While Hou et al. (2012) examine the impact of firm size on model accuracy, LM (2014) separately compare forecast accuracy for firms with and without analyst coverage. In addition, by benchmarking their models against the random walk, the HVZ model and the analyst forecast, LM (2014) develop the models of "earnings persistence" (henceforth the EP model) and "residual income"(henceforth the RI model) which are proved to outperform both the RW and the HVZ model. However, Hess Meuter and Kaul (2019) indicate that both the HVZ and LM model ignore the fact that the earnings in HVZ model are income with special items (henceforth HVZ income), in the LM model scaled income with special

[5]See Lorek (1979)

items (henceforth LM income), in analyst forecast the analyst-adjusted income. As a result, a like-for-like evaluation with in total nine models is done in their work.

More recently, researchers use pooled cross-sectional approach.There is consensus in the literature regarding the choice of balance sheet items that will serve as key variables in the regression model. However, earnings are further divided to recognize the different forecast abilities of various components. For example, Call, Hewitt, Shevlin and Yohn (2016) improve the model in connection with the different persistence of other earnings components. Banker and Chen (2006) construct a regression based on cost variability and cost stickiness. Ball and Ghysels (2017) develop a model based on Mixed Data Sampling Regression Methods (MIDAS), adding various high-frequency time-series data to forecast earnings. Evans, Njoroge and Yong (2017) mentioned that analyst forecast is a better proxy at a shorter horizon as forecasts generated from the mechanical earnings models are all less accurate than consensus analyst forecasts at a one-year forecast horizon. Azevedo, Bielstein and Gerhart (2021), Harris and Wang (2019) add analyst forecast and stock price into the mechanical forecast models. The rationale for doing this is for either the analyst forecast or the stock price to provide additional information compared to only accounting data. Their modelling is found to be more accurate than analysts' forecast.

This paper is closely related to Hess et al. (2019), LM (2014) and Hou et al.(2012). The main objective is to partially replicate the empirical work in these papers and compare the three mechanical earnings forecast models - the HVZ model, the EP model and the RI model. Specifically, following previous studies, I do both coefficient estimation and the calculation of forecast dispersion. In addition, I follow Hess et al. (2019) implementing a like-for-like evaluation using four kinds of earnings definitions instead of three. On the other hand, although Hou et al. (2012) and LM (2014) find that the scaling/unscaling has no impact on forecast accuracy, I have the opposite view: Scaling makes a difference in both coefficient estimation and the calculation of forecast dispersion by examining both the scaled and unscaled variables by defining earnings differently. I find that scaling yields accurate coefficients estimation and improve forecast performance. A detailed explanation is given in the section on data preparation. When using earnings per share excluding special items, this paper converges to LM (2014) who adjust the HVZ's earnings by subtracting special items to make results comparable.

I compare the models in four aspects. Firstly, a theoretical background of the model is presented, and a brief explanation of the selected variables. Secondly, the comparison goes on to the regression models. As done in the literature, I implement a pooled regression using a 10-years rolling window. Then the mean, standard error, and significance of the regression coefficients are examined. Finally, the comparison moves to forecast performance. I use coefficients derived from each rolling window to calculate the forecast for the next year. Unlike Hess et al. (2019), I calculate forecast accuracy and bias as earnings per share. Cross examinations of forecast accuracy and bias between models and earnings definitions are implemented. While forecast bias is the difference between actual and forecasts earnings, forecast accuracy measures the absolute deviation of model forecasts. Following LM (2014) and Hess et al. (2019), I partition samples into analyst-covered and non-covered sub-samples and examine the forecast performance.

The empirical work in this paper shows some apparent limitations due to data manipulation. In addition, I separately examine the modules of both pooled OLS and simple OLS in python to confirm whether the results vary. This is a naive procedure, but it provides evidence to guide future empirical work using the python. I find that coefficients estimation of the large set of time-series cross-sectional data using pooled regression or simple OLS are the same. However, the Newey-West method should be used to correct heteroscedasticity. The Newey-West standard errors of coefficients are five to eight times larger than those derived from the simple OLS. This finding is consistent to Fama and French (2000) who adjust the standard errors using the Fama and MacBeth (1973) method instead.

The estimated coefficients are primarily consistent with the previous studies. Parameters for all selected variable have expected signs and show significance in terms of t-statistics. Specifically, accrual, current earnings, asset, book value of equity and dividend payment are found to be influential factors for earnings forecast. However, different earnings definitions can result in different results. For example, using unscaled earnings leads to coefficient estimation problems and generates forecasts with a large deviation from actual figures. As indicated earlier, the unsatisfactory results can be attributed either to non-comprehensive data preparation or to potential problems on variables. But this limitation again suggests the significant dependence of mechanical models on sample characters.

On the other hand, although the West and Newey (1987) t-statistics of regression coefficients are mostly as significant as the results in the previous studies, both the forecast means, bias and accuracy in my work are larger than in previous results for each earnings definitions. Hess et al. (2019) reports the similar effects due to large extreme results so they show median values. However, the median values in my work are still larger and I can't modify them. Despite limitations due to inadequate forecast results, this paper contributes to the existing literature by concentrating on earnings forecast and providing detailed comparison of different models in terms of theoretical background, empirical method and forecast results. Besides, this paper examines the fourth definitions and incline with the finding of Fairfield, Sweeney and Yohn (1996) who indicates the different persistence of special items and extraordinary items.

The rest of my paper is organized as follows. Section 2 reviews literature of earnings forecast models starting from time series models. Section 3 describes the HVZ, EP and RI models and gives detailed explanation for each. Section 4 provides information on data selection and empirical execution, including discussing the empirical methods used in prior literature. Section 5 compares the model in terms of forecasts bias and accuracy as well as analyst coverage. Section 6 concludes.

2 Literature review

Green (1964) mentions that a good forecast model should capture the behaviours of the data series. Therefore, I divide the current literature into two parts, contributing to the body of knowledge in data series concerning predictor variables and forecast models.

Some studies focus on fundamental and financial statement analysis to explore predictor variables. Lev and Thiagarajan (1993) concentrate on fundamental analysis, offering insight into the earnings persistence and growth. They conclude that while earnings persistence reflects earnings quality earnings growth indicate the firms' potential. They select firstly fundamental signals such as selling general and administrative expenses and gross margin which are usually income statement items. Then a regression of excess stock returns on the chosen fundamental signals and the changes in earnings is set up for empirical analysis. Aside from the great insight offered, the authors also demonstrate how to depict fundamental variables using absolute value, percentage change, or dummy variables. Their models are specified as follows: [6]

$$\Delta E = \alpha_0 \Delta PTE + \alpha_1 \sum_{j=1}^{n} {B_{ji}}^1 \tag{1}$$

Abarbanell and Bushee (1997) examine Lev's fundamental signals and find that the changes in fundamental signals provide indication for changes in earnings. Freeman, Ohlson and Penman (1982) set up a model within LOGIT framework and demonstrate that book rate of return has informative content in terms of change in earnings, especially when there is a deviation from its mean value. Their models are specified as followed: [7]

$$Probability\,(EPS_t > EPS_{t-1}|ROR_{t-1}) = \,(1 + exp\,(-\alpha - \beta ROR_t{-}1)^{-1} \tag{2}$$

Ou and Penman (1989), in comparison, focused on financial statement analysis. They suggest that identifying variables correlated to future payoffs can help to choose predictor variables. The authors select financial descriptors such as percentage changes of current ratio and dividend per share and then run a LOGIT

[6] Delta PTE is the annual change in after teax earnings, Bji is the fundamental signal

[7] ROR is book rate-of-return which is the net earnings scaled by shareholders' equity

regression. Their selection of 16 financial descriptors that show significance provides a good basis for the selection of predictor variables in other models. Moreover, they calculate the probability of an earnings increase. Matching their results in the direction of changes with actual situation leads to over 50% correction rate. The authors demonstrate an innovative earnings forecast model which functions conveniently for the settlement of trading strategies. Their models are specified as follows: [8]

$$Probability\,(E_t > E_{t-1}) = E_{t-1} + \sum_{j=1}^{n} s_{ji} \tag{3}$$

Like Ou and Penman, Fairfield et at. (1996) also analyze the financial statement, but only those items related to earnings. They study the predictive power of different earnings related items, which are like net income, special item, extraordinary items and income from discontinued operation. The authors use a simple linear model, demonstrating empirically that a proper selection of earnings results in better forecast results. They find that special items rather than extraordinary items and discontinued operations have more forecasting power. Their models are specified as followed: [9]

$$E_{t+1} = E_t + \sum_{j=1}^{10} relevant\ item_{ji} \tag{4}$$

The above studies are within the framework of fundamental or financial statement analysis, considering a broad scope of selection. Finger (1994) mentions that current earnings are the most predictive factor for future earnings, as the item already contains the majority of fundamentals information. Therefore, researchers like Sloan (1996), Banker and Chen (2006), Call et al. (2016) and Finger (1994) jump outside the statement analysis and focus on earnings itself by valuing different relevance of earnings components on predictions.[10]

Accounting earnings can be viewed as the cash flow components and accruals components (Sloan 1996). While the cash flow components are indeed backed up with the firms' profitability, accruals are frequently exposed to potential managerial manipulation due to discretion. Penman and Zhang (2002) mention that the discretionary

[8] sij denotes the financial descriptors

[9] Relevant items are different disaggregated earnings-related items

[10] The relevant items in the fourth equation are earnings-related items including those that are not real earnings like gross margin, depreciation, etc. However, the earnings components here denote strictly accounting earnings of different definitions.

accounting policy can work alongside the firms' investment decisions to largely influence the bottom line figures. Therefore, the persistence of earnings from cash flow is greater than that from accruals. However the accruals items according to DeAngelo (1986) can be further separated to discretionary and non-discretionary parts. While non-discretionary accruals result from change of the inner-situation, containing predictive content for future [11], discretionary accruals reduce earnings quality and therefore are less persistent. Graham, Dodd and Cottle (1934) recognize extreme accruals such as arbitrage reserve and abnormal depreciation. Sloan (1996) study these extreme accruals, confirming their significantly lower persistence. The above mentioned relationship between accruals, cash flow and earnings is commonly specified as:

$$Earnings_{t+1} = \chi_0 + \chi_1 Accruals_t + \chi_2 CashFlows_t + \epsilon \tag{5}$$

Banker and Chen (2006) integrate the ideologies of earnings decomposition. They firstly disaggregate earnings into operating plus non-operating income. This process is similar as Fairfield et al. (1996), distinguishing different earnings within financial statement analysis. Then earnings are decomposed to reflect cost stickiness and variability, which are associated with fundamental analysis. Lastly earnings are decomposed into cash flows and accruals components, which reflects the earnings' nature. Their idea of earnings decomposition and disaggregation also serve as a motivation to examine the forecast power of different earnings definitions in mechanical models in the later empirical design. One of their forecast models is specified as followed:

$$E_{t+1} = \chi_0 + \chi_1 Dummy_t + \chi_2 E_t + \chi_3 Sales_{t-1} + \chi_4 Sales_t - \chi_1 Dummy_t + \epsilon \tag{6}$$

Since cash flow is less likely to be manipulated, researchers often deem it unbiased and use it directly to analyze the earnings quality. According to Dechow, Ge and Schrand (2010), persistence, accruals and smoothness are common descriptors of earnings quality. Dichev and Tang (2008) suggest the larger the mismatch of expenses and revenues exists, the lower the level of earnings sustainability. Dichev and Tang (2009) further mention that high earnings volatility can moreover worsen this mismatch. And they find that high earnings volatility has even larger negative effects on predictability than accruals effects studied by Sloan (1996). Freeman et al.

[11] See Jones (1991)

(1982) examine the mean reversion of earnings. They conclude that the smooth and steady earnings contain more predictive content, while volatile earnings offer little predictive power as extreme value often exhibit stronger mean reversion thus being uninformative. The mean reversion model put forward by Freeman et al.(1982) is specified as follow:

$$E_{t+1} = \alpha_0 + \alpha_1 E_t + \epsilon_{t+1}. \tag{7}$$

Where ϵ_{t+1} indicates the existence of heteroscedasticity.

Those intrinsic effects of earnings combined with valuable financial descriptors from financial statement analysis serve as the major basis for the selection of predictors variables in mechanical earnings forecast models. The findings not only lay the foundation but also back up the model improvement. For example, Call et al. (2016) improve the performance of parsimonious cross-sectional forecasting models by adding firm-specific estimates of earnings persistence. Banker and Chen (2016) propose three separate models for different earnings decomposition and find their models outperform models that only uses line items in the balance sheet. Dichev and Tang (2009) introduce accruals, earnings, and earnings persistence into one forecasting model. The earnings persistence is modeled as followed:

$$E_{t+1} = \lambda_0 + \lambda_1 E_t \tag{8}$$

From this paragraph, I review literature with respect to forecast models. Pure discussion of earnings forecasts concentrates on early literature when time series properties of earnings are used to model future earnings. Foster (1977) mentions that time series analysis provides one basis for the selection of forecasting models. Time series models characterize the firms' future earnings with the past earnings series. Compared to the most cross-sectional studies, where annual data are pooled as a large information set, time-series research also uses quarterly data. Dichev and Tang (2009) suggests a classification of earnings forecast as either the short-term or the long-term forecasts. While the short term forecasts need quarterly earnings, long term forecasts use annual data. This suggestion also applies to cross-sectional studies. For example, Lorsbach (2019) examines the HVZ, EP and RI models using quarterly instead of annual time series cross-sectional data to improve the forecast accuracy. Matched to the classification of Dichev and Tang (2009), the data used for examination of the mechanical models in this paper are annual panel data. The earnings forecasts are the long term forecasts for one- through five years ahead.

Ball and Watts (1972) examine the time series behaviours of annual earnings using four models. They conclude that the annual accounting earnings follow a random walk with normally distributed error and a linear trend through time.

$$E_t = E_{t-1} + 4Z_t + 1 \tag{9}$$

Beaver (1970) and Albrecht, Lookabill and McKeown (1977) also examine the properties of annual earnings. They use first and second-order of autocorrelation and find no significant linear relation but only a random walk. They demonstrate that the annual earnings follow random walk and there are no correlations between earnings changes. The random walk model is specified as followed:

$$E_{t+\tau} = E_t + \epsilon \tag{10}$$

Although LM (2014) find random walk model performs better than all the mechanical models at a one-year ahead horizon, Kormendi and Lipe (1987) reject the random walk, arguing that at least random walk does not completely represent the series' properties.

While the annual earnings data are often used for the long term predictions, quarterly data providing more periodic information for the short-term predictions. Lorek, McDonald and Patz (1976) note the importance of modeling either an adjacent component or a seasonal component when predicting the year-end earnings. Their model turn out to be a typical form of an ARIMA. Griffin (1977) models the quarterly earnings using another typical ARIMA form which combines the cross-sectional correlation and the partial autocorrelation. He suggests that the time-series behaviors of earnings follow a first-order autoregressive process with a fourth difference. The ARIMA analysis of the time series datasets also contributes to the adjustment of the serial correlation in forecast error by "simulating" a sub-autoregressive process for serial correlation itself. However, as an ARIMA analysis that incorporates both the feature of AR and MA does not always hold, AR and MA models are also separately examined. For example, Nelson (1974) estimate the MA (1) model, while Gonedes and Roberts (1975) reviewed some issues for the AR (1) model. A simple AR model is shown as follows: [12]

$$\Delta E_q = \beta + \sum_{i=1+\frac{h}{3}}^{I*} \omega_i \Delta E_{q-i} + \epsilon_q \tag{11}$$

Time-series properties can either be studied on an individual firm basis or a pooled cross-sectional basis (Albrecht et al. 1977). While firm specific data are used in individual cases, the pooled time-series data are used for pooled analysis. Ball and Watts (1972), Beaver (1970), Foster (1977) for example, examine the time-series properties of earnings on a pooled cross-sectional basis. By studying the cross-sectional autocorrelation on an individual firm basis, Foster (1977) reports the adjacent and seasonal components case-for-case for his sample firms in terms of earnings, sales, and expense series. However, Griffin (1977) mentions that the firm-specific approach offers only slight improvement for earnings forecast. Watts (1975) favor a parsimonious model with a multiplicative quarterly earnings process. Instead of abandoning the firm-specific ARIMA analysis, Lorek (1979) refers to an alternative to describe the time-series properties of a "representative" firm. Albrecht et al.(1977) prove that a general representative firm model performs equally well as the firm-specific Box-Jenkins models in describing the time-series characteristics of annual earnings for their sample firms. Besides, Brown and Rozeff (1979) combine the firm specific model of Foster (1976) and parsimonious approach of Griffin (1977) and find their version has superior forecast accuracy.

Researchers constantly find new models that could outperform the old ones. However, Lorek (1979) refers to the difficulty of selecting the best model, which holds for cross-sectional parsimonious models as well. Lorek (1979) examines the forecast ability of these time-series models and only finds mixed results. For instance, he disagree with Brown and Rozeff (1979) on the superiority of their model in an earnings forecast, while he demonstrates that the model of Griffin (1977) performs the best with their data. Nevertheless, that the Foster model(1977) performs unevenly is consistent with the Brown-Rozeff's results (1979).

The above paragraphs briefly review the time series literature by introducing different models. Although time series models capture time-series properties of earnings, most of the literature carries out a univariate time series forecast, by excluding many

[12] See Ball and Ghysels(2017)

other variables that may potentially capture helpful information (Foster 1977). Furthermore, the forecast accuracy is inferior to the analyst forecast (Brown and Rozeff 1978). Evidence from Lev (1969) and Freeman et al. (1982) also suggests the statistical invalidity of the time-series model. Lastly, time-series research collects only a few samples; for example, Lorek et al. (1976) use only thirty-two to fifty-two quarterly earnings observations. Although the BJ models already require a time series of 50 minimum observations, it is still few compared to the large set of observations used in the most cross-sectional models.

To check the shortfalls of simple univariate time series models, parsimonious cross-sectional models are often multivariate and are regressed using pooled cross-sectional data that are large and informative. Since variable selection and forecast models have been thoroughly discussed in previous studies, theoretical background is seldom referred to in the recent literature. This part echoes and associates those backgrounds that have probably been neglected. The latest earnings forecast studies usually serve as a sub-topic to reach conclusions concerning the stock price, implied cost of equity, or earnings quality.

The previous evidence already leads to agreement on forecast ability of some fundamental factors and the lagged earnings itself. Sloan (1996) suggests that accruals are non-cash generative earnings that do not promote earnings sustainability. Besides, the book rate return associated with book equity, current ratio related to an asset, earnings persistence reflected by the signs of earnings are incorporated into some forecasting models. Moreover, Miller and Modigliani (1961) mention that dividend amount impacts the future earnings because of its *mark-to-earnings* feature in some companies. On the other hand, Fama and French (1999) demonstrate that a dummy indicator of dividend payment can also be linked to profitability. Evidence for the dummy indicator can be found in the later work of Fama and French (2000), who incorporate both dividend dummy and dividend into their suggestion for expected profitability proxy. And the rest of the variation is captured with a market-to-book ratio. Their models are shown as followed:

$$E_t = \alpha_0 + \alpha_1 M/B_t + \alpha_2 DD_t + \alpha_3 Depr_t/Asset_t + \epsilon_{t+1} \tag{12}$$

In developing another proxy for profitability which is similarly defined by Fama and French (2000), Fama and French (2006) use the original explanatory variables but also add into models accruals, market capitalization, and the measurement of firm strength which Piotroski (2000) and Ohlson (1980) suggest. Apart from a regression summing up the variables, they also fit a cross-section LOGIT regression. Fama and French (2000, 2006) deflate the earnings by the lagged asset to predict profitability, which is different from the deflation in the HVZ, EP, and RI model. The contribution of Fama and French (2000) serves as the basis for the further development of earnings prediction models in the following ways. Firstly, both profitability and earnings are mean reversing. Secondly by calculating the differential in profitability, they find the nonlinearity in the mean reversion of profitability changes. That means a faster reversion of negative than positive changes. They also conclude that the nonlinear mean reversion is the key driver for " the predictable variation in profitability " (Fama and French 2000). Thirdly, although they find mean reversion of profitability is the precondition to predict earnings [13] they cannot attribute the earnings change completely to profitability change. Therefore, as mentioned, Fama and French (2006) examine the prediction by adding new predictors. Although they still find the strongest predictor for future profitability is lagged profitability, asset growth, accruals, firm size are also significant predictors for future earnings. Hou, Van Dijk et al. (2010) conclude that the firm size has robust effect on prediction of profitability as well. Lastly, Fama and French (2000), Fama and MacBeth (1973) enrich the empirical methods by implementing simple OLS regression for analyzing the time series cross-sectional data.

Since previous studies already demonstrate the effectiveness of cross-sectional models to capture expected profitability, Hou et al. (2012) simply uses a cross-sectional model to proxy for the cash flow. They find that the HVZ model shows statistical power and can generate more earnings forecasts than the analysts'. The forecast is lower in accuracy compared to analysts forecast but lower in bias as well. Lastly other variables added into the HVZ model are significant but not contributing to the overall explanatory power. This indicates that the attempt to include all factors into the regression model to depict the best fit isn't a reasonable choice for model improvement.

Hou et al. (2012) develop the HVZ model and offer a solution to generate a wider coverage of earnings forecasts. However, LM (2014) criticize the low performance and high forecast error of the HVZ model by comparing its forecast results with the random walk. They find that the HVZ model performs even worse within

[13]Although both profitability and earnings are earnings related, profitability refers to financial ratios such as ROE, while earnings are simply accounting incomes.

the samples of non-covered firms. Therefore they present two alternatives with interaction items, all capturing the difference persistence of gain and losses. They find their two alternatives generate more accurate forecasted results than the HVZ model and the RI model performs the best among the three.

Based on Hou et al. (2012) and LM (2014), Hess et al. (2019) indicate the inconsistency in earnings used. Therefore they implement regression using different earnings definitions from different samples. They agree on the best performance of the RI model, but state that both the earnings definitions and sample bases affect the forecast performance. Therefore a rough comparison among models without earnings confine is meaningless and misleading.

Swamidass (2000) classifies the forecast error into the standard error and relative error. While standard approach describes error in the same units as data, relative approach reflects a percentage. The typical standard error is the mean error that measures the forecast bias. Swamidass (2000) mentions that bias reflects if the forecast results are always higher or lower than the actual value. However, bias can be misleading as negative and positive forecast errors can net each other. The description of absolute deviation describes the real magnitude of forecast errors. Therefore models that are low in absolute deviation represent for good forecast quality. Instead of using the traditional mean-squared error as evaluation criteria, Hou et al. (2012), LM (2014), Hess et al. (2019) all scaled the individual error before calculating mean. The dollar forecasts are scaled by the market capitalization while per share forecasts are scaled by the price. This measure aims to eliminate the size effect. However,Gerakos and Gramacy (2013) observe that scaling by market equity does not remove all size effect. They propose an adjustment of individual forecast by CPI in the forecasting years (Gerakos and Gramacy 2013, pg.10). [14]

As mentioned earlier, later studies about the cross-sectional models modify the initial mechanical models, for example, alternating different factors in the model. Harris and Wang (2019) add model current and lagged stock price, lagged book value of equity into the model. The model is specified as followed:

$$E_{t+1} = \lambda_0 + \lambda_1 P_t + \lambda_2 E_t + \lambda 3 B_t + \lambda 4 B_{t-1} + \lambda_5 P_{t-1} + \lambda_6 acc_t + \epsilon_{t+1} \tag{13}$$

Combining the analyst forecast and the model analyst is another improvement, exploiting both the advantages of forecast accuracy and forecast bias (Azevedo et al.

[14]consumer price index (CPI)

2017). Their model include gross margins, change between current and ten months before market equity, change between current and last month market capitalization.The model is specified as followed: [15]

$$E_{t+1} = \lambda_0 + \lambda_1 E_t + \lambda_2 GP_t + \lambda_3 r10_t + \lambda_4 r_t + \epsilon_{t+1} \tag{14}$$

Besides, Ball and Ghysels (2017) identify a limitation of forecast models: the low frequency of the data update, therefore, narrow information sets. While pooled cross-sectional regression is an improvement compared to individual time series analysis, MIDAS regression enables using the high-frequency variables along with the lower-frequency variables in one regression model. The author improve the model forecast ability by combining analyst forecast and MIDAS.

[15] E_t is one year ahead forecast on IBES; GP is the gross profit; r10 and r1 are corresponding backward difference in market capitalization

3 The models

3.1 The HVZ model

Hou et al. (2012) forecasts the future earnings using accounting fundamentals, lagged earnings and two dummy variables. They extend the cross-sectional profitability models in Fama and French (2000), Fama and French (2006), Hou and Robinson (2006) and Hou et al. (2010). The model is shown as followed:

$$E_{t+\tau} = \alpha_0 + \alpha_1 A_t + \alpha_2 D_t + \alpha_3 DD_t + \alpha_4 E_t + \alpha_5 NegE_t + \alpha_6 AC_t + \epsilon. \tag{15}$$

$E_{t+\tau}$ is the earnings in t+τ (τ=1 to 5). Originally in the HVZ model, earning refers to income before extraordinary items and special items. This definition of making will be referred to as E1 in my work. However, I also apply another three earnings definitions to the model to examine the difference in the forecast accuracy. A_t denotes the total assets; D_t is the dividend amount ; DD_t is a dummy variable which equals 1 for companies paying dividend and 0 for company that do not pay dividends. E_t is the 1 to 5 years lag of earnings; $NegE_t$ is an indicator variable for loss firms, which equals 1 for the negative earnings and 0 otherwise; AC_t is the working capital accruals. As it is calculated in literature, for the periods before 1988, I use the balance sheet method, where AC is the backward difference of non-cash current assets, current liabilities excluding financial liabilities, and depreciation & amortization. From 1988 onward, the cash flow method is used where accruals are calculated as earnings minus cash flows from operations.

3.2 The EP model

Li and Mohanram (2014) propose the EP model which depicts different persistence in earnings. Gerakos and Gramacy (2013) find the EP model to perform the best among the HVZ, EP and RI model in prediction. Since this model seems to be simple, they conclude that "the simpler, the better." The model is specified as

$$E_{t+\tau} = \beta_0 + \beta_1 NegE_t + \beta_2 E_t + \beta_3 NegE * E_t + \epsilon. \tag{16}$$

Compared to the HVZ model, the EP model shows a simple form of regression by removing all the balance sheet items but keeps only the main items relevant to earnings. The selection of E_t and $NegE_t$ is the same, but NegE*E_t is introduced as an interaction term to model earnings persistence. The interaction term in the EP model reflects different persistence of profit and negative earnings (Li 2011). Fama and French (2000) also find the non-linearity in the mean reversion, implying that negative profitability changes reverse faster than positive ones. Additionally, I expect the coefficients β_1 and β_3 to be negative and β_2 to be positive

3.3 The RI model

$$E_{t+\tau} = \chi_0 + \chi_1 NegE_t + \chi_2 E_t + \chi_3 NegE * E_t + \chi_4 B_t + \chi_5 TACC_t + \epsilon. \tag{17}$$

The last model is the RI model, which exhibits the best forecast accuracy according to LM (2014) and Hess et al. (2019) , however is not quite effective in my work. The model is a modification of the original residual income valuation equation put forward by Feltham and Ohlson (1996). Meanwhile, since stock, prices, returns, lagged/changes in the book value of equity are also proved to be leading indicators of future earnings [16], the RI model is not only mathematically but also intuitively correct. Compared to the HVZ model, the RI model are built with similar variable choices, including indicator variables such as balance sheet variables and lagged earnings. Just as LM (2014) did in the EP model, they introduced the interaction term into the RI model but removed dividend-related variables. On the other hand, instead of using total asset and accruals, the book value of equity and the total accruals from Richardson, Sloan, Soliman and Tuna (2005) are used in the RI model.

Based on the work of LM (2014) and Hess et al.(2019), I expect the coefficients χ_1 χ_3 and χ_5 to be negative, reflecting a lower loss persistence indicated by Li (2011), the quality of earnings and the conservative accounting policy. In contrast, the coefficients χ_2 and χ_4 should be positive because of positive impact of both a firm's book equity (Feltham and Ohlson, 1995) and its current earnings on future earnings.

[16]See Beaver (1970); Weiss, Naik and Tsai (2008), Feltham and Ohlson (1996), Pope and Wang (2005);

4 Data and empirical execution

4.1 Data preparation

Hess et al. (2019) already mention that earnings definitions tend to affect both the coefficients estimation and forecast performance. There are several ways to define earnings and this paper aims to examine the scaling effect, special item effect and extraordinary item effect. Therefore I select four earnings definitions to make a paired comparison. While E1 is unscaled earnings, E3 is scaled counterpart. [17] E2 is the scaled earnings excluding special items, whereas E4 is earnings per share excluding extraordinary items. Comparing E1 and E3, the scaling effect can be examined, while comparing E3 and E2 (or E4) or between E2 and E4, differences in prediction power of special items and extraordinary items can be derived.

Following Hess et al. (2019), LM (2014), Hou et al. (2012), I collect the initial samples from the Compustat North America fundamentals files, IBES and CRSP monthly stock files. The period of 47 years (from 1968 to 2014) is chosen, similar to the time span of significant studies in this field. Like prior studies in which rolling regression method is used, I employ rolling regression as well with a 10 years-window to do both the coefficient estimation and the forecasting. Therefore, the expected forecasting period begins only from 1978[18].

The majority of balance sheet variables are collected from Compustat files including total asset(A_t), dividend(D_t), income(E_t), common shares outstanding, etc. My choice of individual variables is guided by LM (2014), who list the "Xpressfeed variable" in the appendix. However, in the spirit of Hou et al. (2012) and Hess et al. (2019), I make an adjustment to the choice of "Xpressfeed variables". Compared to LM (2014), I calculate the accruals in two ways, as both Hou and Hess do. Therefore, another variable relating to operating cash flow is collected. Another item that is different from the work of LM (2014) but conforms to Hess et al. (2019) is the selection of working capital accruals (AC) rather than total accruals (TACC) in the RI model. In my work both TACC and AC are found to be equally significant so finally TACC is replaced by AC for the sake of comparison. Observations whose values of asset, earnings, book equity are missing are removed, while those with missing accruals and dividend are maintained and filled with "0".

[17] However, due to the sample reasons, E1 and E3 may differ in other respects. E3 is the earnings per share including extraordinary items, whereas E1 is income before extraordinary items. Not all sample firms possess E3 due to change in the GAAP requirement

[18] The actual period begins from 1980

Besides, I scale each numerical variable with the end of year common shares outstanding from CRSP, which is a different choice compared to Gerakos and Gramacy (2013), who choose to scale by market value of equity or the average total asset. However, after robustness test checking for scaling effect, I find the coefficient estimation using either scaled or unscaled variables leads to similar estimates and R squared. To compute price scaled forecast accuracy/bias, the end of June market capitalization and stock price are calculated, with data from the CRSP.

Hess et al. (2019) find that LM (2014) and Hou et al. (2012) differ in the choice of earnings and different earnings choices lead to different forecast results. I follow Hess et al. using four different kinds of earnings definitions as left hand side variables including income before extraordinary items (used by Hou et al. 2012), scaled earnings without special items (used by LM 2014), earnings per share without special items and earnings per share without extraordinary items. Hess et al. (2019) mention that the across-samples comparison is invalid. Therefore I collect the samples only from the Compustat and follow LM (2014) to get the firm level earnings (E1). IBES provides only analyst forecasts of earnings per share.

Concerning handling the outlier effect, most studies winsorize the variables at 1st and 99th percentiles. As discussed by Leone, Minutti-Meza and Wasley (2019), winsorization is a common choice to handle outlier effects in accounting and finance studies. Winsorization is defined as the process of replacing the data that fall outside the interval of chosen percentiles by a specific value on those chosen percentiles. Following the previous literature, I winsorize only the final non-dummy regressors to keep the optimal information in the original datasets at the 1st and 99th percentile. However E1 is insensitive to the 1st and 99th percentile winsorization due to the low skewness.

Considering the impact of winsorzation on forecast performance, Gerakos and Gramacy have found in 2003 that winsorization functions better for scaled variables, while it does not perform well for unscaled variables. I attribute their finding to the issues with data sets. When scaling variables, common shares outstanding are used as the denominator. However, since the initial common shares outstanding contain extremely small values, the newly scaled variables are exceedingly large (or small).[19] And those large variables are further winsorized, which leads to losing a significant amount of information. In addition, there is sometimes huge dispersion between shares figures on the Compustat and the CRSP. In contrast, scaled variables like

[19] Hess et al. (2019) refer to the effect of stock-split or share buy-back

E3 and E4 are accurate records since they are direct items on the income statement. Still, I find that winsorization remains the favorable choice, as observations with extreme values distort the regression coefficients. [20] After processing outliers, datasets are smoother than raw data, leading to proper coefficients estimation. In addition, Gerakos and Gramacy (2013) find that the winsorization has divergent impact on forecast accuracy with regard to estimation horizons and scaling processing. Specifically, when forecasting unscaled earnings with long estimation horizon, unwinsorized predictor variables usually cause accurate forecasts. For forecasting scaled earnings with short estimation horizon, winsorized predictor variables leads to accurate forecasts.

When computing scaled variables, any observations with the positive or negative "infinity" value produced during the data preparation are removed. While E1 corresponds to unscaled variables, E2 to E4 corresponds to scaled ones. Furthermore, because of inconsistent firm identifiers across the different databases, linking different databases leads to a loss of observations. Nevertheless, around 228,563 final observations still provide a strong base for the ordinary least square (OLS) regression. Moreover, I partition data into analyst covered and non-covered subsets for the comparison of forecast accuracy and bias.

Table 1: Summary statistics of all relevant variables. All values except for N in million US-dollar.

Variable	N	Mean	1%	25%	Median	75%	99%	std
E2	194,446	0.89	-4.09	-0.08	0.59	1.65	7.66	1.77
E3	207,636	0.74	-6.77	-0.14	0.55	1.61	7.86	2.03
E4	207,232	0.74	-6.32	-0.13	0.55	1.59	7.47	1.94
NegE_E4	207,232	-0.34	-6.32	-0.13	0	0	0	0.96
NegE_E3	207,636	-0.36	-6.77	-0.14	0	0	0	1.03
NegE_E2	194,446	-0.23	-4.09	-0.08	0	0	0	0.64
ACt_	181,091	-1.09	-12.86	-1.65	-0.49	-0.01	5.97	2.48
Bt_	204,349	10.13	-3.2	2.38	6.99	14.39	55.41	10.75
Dt_	203,329	0.35	0	0	0	0.47	3	0.62
At_	204,497	45.38	0.14	5.35	16.2	43.95	565.25	85.54
prccm	228,562	17.28	0.22	5	12.5	23.75	89.61	17.07

Table 1 reports the summary statistics of the dataset of 228,563 observations containing the complete required information from the CRSP and the Compustat. For each of them, the time series averages of the annual cross-sectional means, medians, standard deviations and selected percentiles of the variables are described. E2, E3 and E4 are earnings definitions mentioned earlier. A_t is the scaled total assets; D_t is the scaled dividends; AC_t is the scaled accruals; $NegE * E_t$s are the products of

[20]For example, a parameter that should have been negative is very positive

the NegE$_t$s, the dummy variables and respective earnings in year t. B_t is the scaled book value of equity. The "prccm" is the value of the end of June stock price.

I summarize the statistics of analyst forecasts in Table 2, where analysts forecasts for only one to three years are reported to save space. Forecast values are earnings per share scaled by the end of June stock price. To prevent the effects of extreme results, the table only shows the median value. N denotes the number of forecasts available at the respective forecasting horizons. The last row represents the statistics with respect to the total year span. The table shows that while the number of analysts forecasts declines at longer-horizon forecasting, the coverage improves over time. In particular, the dataset shows 151,225 forecasts for one-year ahead, 147,596 forecasts for two years ahead, and 89,907 forecasts for three years ahead. Following Hess et al. (2009), Hou et al. (2012), I infer the missing forecasts by using the IBES consensus long-term growth rate and only the latest forecasts. If the forecasts remain unavailable more than a year, they are not inferred by growth rate any more.

Table 2: Summary statistics of earnings forecasts,1982–2009.

fyear_group	N_Et+1	Et+1_median	N_Et+2	Et+2_median	N_Et+3	Et+3_median
1982-1987	14,988	0.80	13,393	1.07	3,639	1.39
1987-1992	19,259	0.68	18,296	0.90	6,564	1.13
1992-1997	24,932	0.66	24,428	0.88	12,223	1.16
1997-2002	29,576	0.64	28,939	0.87	15,441	1.13
2002-2009	43,301	0.77	43,301	0.99	34,551	1.26
1982-2009	151,225	0.73	147,596	0.96	89,907	1.25

4.2 Empirical execution

Before showing empirical results, this section briefly reports the process of empirical execution. The section is structured as follows: the potential issues associated with the empirical methodology employed in the mechanical models are firstly discussed. Then the main empirical procedure used in this paper is described.

Common econometric datasets include cross-section, time series, and panel data[21]. Baltagi (2008) describes that while time-series or cross-section describes data only in one dimension, panel data features both cross-section and time-series characteristics. Pooling panel data provides richer information for parameter estimation. The data

[21]See Baltagi (2008)

type used for the earnings forecast is the time-series cross-sectional data. When analyzing this type of data, potential problems appear. Fama and French (2000), Freeman, Ohlson and Penman (1982) and Fairfield et al. (1996) and West and Newey (1987) have discussed them in their studies.

When building up a model with cross-sectional data, it is necessary to consider whether one model can be applied to the whole sample periods. Freeman et al. (1982) and Fairfield et al. (1996) solve the problem by using pooled time-series cross-sectional regression with inclusion of dummy variables to control power problem. In contrast, Fama and French (2000) mention that both the Fama-Macbeth regression and the year-by-year cross sectional regressions are proved to have equivalent slopes to those of pooled time-series cross-sectional regression.

In addition to power problem, Fama and French (2000) mention that allowing for residual cross-correlation is crucial which is the second problem of such type of data analysis. Cross-correlation refers to the interrelationship between two time-series, possibly measured by the cross correlation of residuals. [22] By pooling large samples, Fama and French (2000) can use average slopes as the coefficients for estimators. And the time series standard errors that reflect the over-year slope variations already incorporate the residual correlation effect, so that the problem can be ignored. Specially, they find that the Fama-Macbeth standard error can also account for residual cross-correlation, contributing to the correct parameter examination as well.

Haugh (1976) mentions that autocorrelation within a time sequence can inflate the cross-correlation between time sequences, suggesting autocorrelation as the third problem. Autocorrelation refers to the correlation between the variables and the " lagged value of itself ". Potentially, the lagged variable itself can also correlate with the residual error, leading to biased and inconsistent estimators due to inaccurate estimation of standard error and therefore the t-statistics. Besides, when there are structural changes during the sample periods, heteroskedasticity may also exist. I implement the traditional BP test and the result of the BP confirms the existence of heteroskedasticity among the samples. Both heteroscedasticity and the autocorrelation violate the OLS assumption. Newey and West (1987) propose the adjustment of standard errors to allow for both autocorrelation and heteroskedasticity. Hou et al. (2012) and Hess et al. (2019) all report the Newey-West time series averages of coefficients and the time-series Newey-West t-statistics.

[22] See "white noise" in Haugh (1976)

Another criterion for pooled cross-section regression is the sample characteristics. According to Wooldridge (2010), fixed effects and random effects are applied to the balanced panel with the same samples of individuals in different periods, while pooled OLS is applied to the unbalanced panel with different individuals for each period. The pooled OLS ignores the entity and time effects, therefore employing constant coefficients referring to both intercepts and slopes[23]. In estimating this type of model, researchers can pool all of the data and run an ordinary least squares regression model (Hiestand 2015).

In the spirit of Fama and French (2000) who discuss in detail the regression method in their work, the later studies omit the methodology part and simply implement pooled cross-sectional regression. Common empirical execution is: regressing the future earnings on chosen regressors through a rolling regression with 10-years' window which aims to get the moving average of coefficients to smooth the structural changes. Then the standard errors are adjusted using the Newey-West method. This paper follows the mainstream method.

Earnings forecasts are derived from three discussed models which are the HVZ model, EP model and RI model. Each model is estimated using four kinds of earnings. The setting helps to examine the finding of Hess et al. (2019) and Fairfield et al. (1996) that different earnings can show forecast differences. The empirical process is divided into three parts: first, I run both pooled cross sectional regression and the simple OLS using earnings on lagged independent variables. The coefficients estimation is the major result of this part. Secondly, considering the regression coefficients of each rolling window and firm information on the benchmark year, I calculate one to five-year ahead model forecasts. The major result of this part is the statistics of model forecasts at each horizon. Thirdly, using the forecasting results, I compute forecast accuracy and forecast bias, which are defined the same as the major studies. I follow LM (2014) and Hess et al. (2019) to calculate the difference between each model under different earnings concepts.

Furthermore, the analysts forecasts are also considered to make comparisons. The major output is the comparison results with denoted statistical significance level. Specifically, forecast bias and accuracy are described below:

$$Forecast\ bias = \frac{(Actual\ Earnings - Earnings\ Forecast)}{June\ market\ capitalization} \tag{18}$$

[23] See Hiestand et al. (2005)

$$Forecast\ bias = \frac{(Actual\ Earnings - Earnings\ Forecast)}{June\ stock\ price} \tag{19}$$

$$Forecast\ accuracy = |Forecast\ bias| \tag{20}$$

Since the number of observations is in decline each time merging one database with the other, I partition data for different purposes. Each dataset comes originally from one database but is handled differently to conserve the optimal information. Besides, the prior literature exhibits an inconsistency (Hess et al. 2019). While Hou et al. (2012) use the income before extraordinary items, LM (2014) excludes special items. They compared the HVZ model with the EP and RI models using earnings per share excluding special items, which is not the original concept of Hou et al. (2012). The problem arises because earnings possess different definitions in accounting and can also be scaled or unscaled. For example, McVay (2006) clearly state that the classification of core earnings and special items is used as a earnings management tool to meet analyst forecast.

On the contrary, in comparison to previous studies, I find that scaling leads to differences in coefficient estimates, causing some variables to be statistically insignificant. However, scaled earnings excluding special items function correctly to estimate coefficients. The main part of the empirical analysis show both scaled and unscaled variables through different earnings definitions. Specifically, I implement a cross-check and a "like-for-like" evaluation when comparing forecast results. The objective of first part of empirical analysis is to run the models and fit the coefficients. Both simple OLS and pooled OLS is implemented in order to compare the difference of residual error as indicated by Fama and French (2000). The second part aims to make earnings forecasts for one through five years ahead using the coefficients from each rolling window. The third part includes the calculation and the comparison of the forecast accuracy and forecast bias. Lastly, I partition samples into covered and non-covered subsets to further analyze the differences in between.

5 Empirical results

5.1 Comparison of regression method

To illustrate the pooled regression method mentioned in the prior studies, I describe briefly in the following paragraphs the estimation of coefficients based on the pooled OLS and simple OLS.[24] The description is only based on the coefficient estimates of the HVZ model for E1 forecasts of one-year ahead without using rolling regression. And the results are not tabulated.[25]

The values of estimated parameters are almost identical in the two regressions with only slight differences. For example, the coefficient of AC_t in the PooledOLS is smaller than that in the simple OLS. And the coefficient of DD_t is larger in the pooled OLS but smaller in the simple OLS. I focus on t-statistics and adjusted/R^2 as the small differences in coefficients are not relevant for statistical interpretation. With respect to R^2, both r-squares are almost identical with only a small difference.

Despite the similarity between the two regression methods, the covariance estimator called Driscoll-Kraay is executed to correct heteroscedasticity problems in the pooled OLS. In the simple regression there is no correction.[26] According to the "linearmodel" guide, Bartlett's kernel is the default kernel, which produces a covariance estimator like the Newey-West covariance estimator (see python-linearmodel guide). With the original dataset around 220,000 observations, the t-statistics in pooled OLS for most coefficients without heteroscedasticity correction can be five to eight times larger than that in regressions with heteroscedasticity correction. Therefore the standard errors without corrections are smaller than those with corrections. This observation reinforces the finding of Fama and French (2000) regarding the Fama-Macbeth standard error (pg.163) and Haugh (1976)'s conclusion regarding the autocorrelation. Another interesting finding is that with a larger dataset (examining from circa. 92,000 to finally circa. 220,000 observations), the heteroscedasticity strengthens as the residual correction leads to larger changes in t-statistics. The finding indicates the necessity to correct for heteroscedasticity in pooled cross-sectional regression.

[24] Python has modules for both the PooledOLS function and the simple OLS function
[25] The results can be examined in "resulttable.xlsx" in the submitted zip-folder
[26] The OLS function in statsmodel can be extended to compute the covariance matrix

This pretest provides empirical evidence that pooled regression analysis can be either done through the simple OLS or pooled OLS on python. The two modules generate unbiased and similar coefficients.

5.2 Coefficients estimation

This subsection includes the results of coefficients estimation, including the coefficients description and the statistics comparison among different models. The table below reports the Newey-West time series averages of regression coefficients for the HVZ, EP and RI model for one-year ahead forecast. By applying four different earnings definitions to each of the three models, I obtain twelve sets of regression coefficients. Results of each set are presented as follows.

Table 3: Coefficient estimates of the three cross-sectional earnings forecasts model using four different earnings definitions, 1968-2014

	ACt	At	DDt	Dt	Et	NegEt	NegE_Et	Bt	R^2
HVZ_E1	-0.1329***	0.0053	-12.4255	0.4988***	0.7125***	-6.7581			68.83%
HVZ_E2	-0.0274***	0.0015***	0.1292***	0.189***	0.7061***	-0.0162***			63.54%
HVZ_E3	-0.05***	0.0019***	0.2339***	0.3235***	0.5305***	0.0268***			44.01%
HVZ_E4	-0.0587***	0.0014***	0.1962***	0.239***	0.608***	-0.0162***			51.47%
EP_E1					0.9059***	-29.417***	-1.2158***		68.50%
EP_E2					0.8511***	-0.192***	-0.4597***		64.51%
EP_E3					0.7927***	-0.2993***	-0.6592***		47.85%
EP_E4					0.8258***	-0.2775***	-0.6064***		54.25%
RI_E1	-0.1823***				0.7627***	0.2927	0.0076	0.0211	68.47%
RI_E2	-0.0195***				0.7615***	-0.1538***	-0.2611***	0.0136***	63.65%
RI_E3	-0.0402***				0.6909***	-0.2772***	-0.4637***	0.0116***	44.46%
RI_E4	-0.0501***				0.7389***	-0.2257***	-0.3766***	0.0084***	51.77%

The estimated coefficients using different earnings definitions exhibit different results. In general while E2, E3, E4 lead to expected and significant coefficients estimation, the unscaled total earnings, E1, in some cases returns estimation that does not conform to a priori expectation. For example the signs of negative indicator NegE_t and the interaction item NegE*E_t in the RI model are unreasonably positive, although they are insignificant. Similarly, A_t ,DD_t and NegE_t in the HVZ model are insignificant. However, the estimations of the EP model with all four earnings definitions are consistent. The coefficients for all its variables exhibit significance and signs that conform with economic intuition, similar to Hess et al(2019).

A plausible reason for the good estimation performance of the EP model using E1, in contrast to the estimation of the HVZ and RI models, is that all of the EP's variables are only related to earnings. Therefore, the independent variables are unified in scale, even though unscaling still shows its impact on the slope coefficients. For example, in the EP model, the slope coefficients for the negative dummy $NegE_t$ reflect exactly the scaling effect. The parameter is -29.417 using E1 compared to -0.2993 using E3. However, the coefficients also vary among other scaled earnings definitions showing no large deviation as between E1 and E3.

For each model and each estimation horizon, the coefficients of E1 are the largest for the same model compared to the other earnings definitions, suggesting that scaling variables reduces information content and earnings become less persistent. Specifically, the coefficients of E1 for the one-year ahead regressions are 0.7125, 0.7627, 0.9059 for the HVZ, the RI and the EP models, respectively. In contrast, the coefficients of E3 are respectively 0.5305, 0.6909, 0.7927. Moreover, E2 has the largest coefficients among E2, E3 and E4, suggesting that earnings excluding special items are more persistent than excluding extraordinary items. That the coefficients of E3 are the lowest also confirms the higher persistence of core earnings rather than earnings without disaggregation.

I also find that R-squared generally declines with longer forecasts horizon. For example using E1 in the HVZ model, the R squared decline from 68.83% for one-year ahead forecast to 48.39% for the five-year ahead forecast, suggesting that the forecast ability of mechanical models decline with a longer forecast horizon. Compared to the previous studies, I did not get as high R-squared as Hess et al (2019), HVZ (2012) and LM (2014), although I use the same scaled variables and same sample periods as Hess et al (2019). Procedure during data preparation can be a possible reason. Also, Hess et al(2019) mention that R^2 using different earnings definitions is almost identical with the same model. In contrast, I find significant differences in R^2 among earnings definitions. Specifically, R^2 is higher using E1 and E2 than E3 and E4. Scaled earnings excluding special items result in higher R^2 than total earnings and those excluding extraordinary items. On the contrary, the variation is not much among different models.

With respect to other variables such as B_t AC_t A_t and D_t, their signs are consistent with previous studies. Although the RI model is derived from one of the valuation model equations (LM 2014), B_t in the RI model can be interpreted as a replacement of A_t from the HVZ model. Therefore, both A_t and B_t are positively correlated with future earnings as expected. In addition it is almost certain that D_t is also positively correlated with the future earnings. AC_t on the contrary has the negative signs. I reach the same conclusion through my empirical studies. However,like Hess et al (2019), I find an one-off unaccountable and significantly positive estimation for $NegE_1$ when using E3 in the HVZ model for regression of one-year ahead. With respect to t-statistics, apart from using E1, all of the variables in my regression show significance at least at 10% level.

5.3 Earnings forecasts

From each rolling window derived coefficients are used to predict the earnings of the next year. Following Hess et al.(2019), I show in the following table the forecast mean and median for the three models using four earnings definitions. Comparing to Hess et al.(2019) who list in the table the number of forecasts generated from mechanical models, I summarize in the table the number of covered firms. The reason is that my initial samples of analysts forecasts are full rather than consensus IBES analysts forecasts, containing multi-party forecasts for one firm. Therefore N in the following table is set to denote the number of covered companies by counting the unique "gvkey"[27] for each forecast periods. Besides I also incorporate the respective statistics of the IBES analyst forecasts from 1982 to 2014. Lastly, forecasts generated using E1 are the dollar forecasts, which are scaled by market capitalization. The rest is earnings per share, which is scaled by the end of June close price.

Table 4: Summary statistics of earnings forecasts, 1980 - 2009.

	Et+1			*Et+2*			*Et+3*		
Model/Analyst	N	Mean	Median	N	Mean	Median	N	Mean	Median
HVZ-E1	9,829	0.0127	0.0493	8,670	0.0398	0.0430	7,600	0.0527	0.0355
RI-E1	9,837	-0.0183	0.0488	8,680	0.0380	0.0519	7,610	0.0461	0.0509
EP-E1	10,983	-0.1228	0.0435	9,707	-0.1266	0.0406	8,499	-0.1385	0.0377
HVZ-E2	9,163	0.0120	0.0438	7,986	0.0218	0.0383	6,959	0.0245	0.0327
RI-E2	9,165	0.0103	0.0449	7,997	0.0206	0.0404	6,974	0.0233	0.0363
EP-E2	9,966	0.0097	0.0441	8,693	0.0154	0.0377	7,572	0.0168	0.0325
HVZ-E3	9,496	0.0033	0.0374	8,353	0.0200	0.0333	7,327	0.0232	0.0274
RI-E3	9,532	-0.0009	0.0384	8,447	0.0147	0.0347	7,463	0.0180	0.0310
EP-E3	10,718	0.0062	0.0396	9,479	0.0193	0.0340	8,306	0.0223	0.0292
HVZ-E4	9,497	-0.0012	0.0386	8,404	0.0146	0.0336	7,421	0.0166	0.0283
RI-E4	9,507	-0.0044	0.0392	8,414	0.0111	0.0355	7,432	0.0146	0.0320
EP-E4	10,693	0.0030	0.0409	9,442	0.0148	0.0352	8,272	0.0182	0.0305
Analyst	13,871	0.0462	0.0389	12,884	0.0744	0.0495	9,011	0.0911	0.0611

To align the forecast periods between model and analyst, I set the forecast period from 1980 to 2009 [28]. I find it challenging to compare forecast results among models based on the statistics of earnings forecasts, because I get some unexpected findings. Firstly, analysts' forecasting covers more companies than mechanical models

[27] One type of company identifier used in Compustat and CRSP
[28] I fail to generate as many forecasts results as the previous studies do, therefore am only able to align results from 1980 to 2009

do. This result differs from other studies, which generally conclude that the mechanical models can generate more forecasts. However, the result is interpretable from two aspects. Firstly, the N in table 4 represents different meanings. Secondly, I assume that there are potential limitations from data transformations. Since I incorporate all earnings definitions in one initial file and observations with any missing explanatory variables are removed. For example, E3 which represents diluted earnings per share including extraordinary items is not available in all samples. Therefore those observations without E3 but with E1 and E2 are eliminated. In addition, I observe that with longer forecast horizon, both mechanical models and analysts decline in the number of forecasts. This finding differs from the Hess et al. (2019); this is probably due to data transformations/manipulations leading to loss of observations. That the EP model generates the most forecasts among all the mechanical models can also infer this explanation, as the EP model requires the least input of fundamental information. [29]

While the median figures generally converge among the three models, the mean figures are quite different, implying the difficulty of comparing the models. Hess et al. (2009) also mention that comparing results among models can be misleading. E1 and E2 generate larger figures than E3 and E4 in term of the median results with respect to earnings definitions. Since analyst often exclude "non-recurring" items from their earnings definitions, I compare the forecast results of E4 with analyst forecasts. Table 4 shows that analysts report higher forecasts than mechanical models, which implies their optimism. While analyst have access to rich information, their forecast behaviours can be influenced by conflict of interests.

5.4 Forecast accuracy

The computation of both forecast accuracy and forecast bias belongs to performance analysis. Comparing the forecast performance among models and earnings definitions provides more purposeful information than earnings forecasts statistics. The accuracy of each model using each type of earnings is examined. For example, the description of the HVZ model includes both the forecast accuracy using different separate earnings definitions and the differences among models and earnings. Furthermore, samples are partitioned into covered and non-covered to be specifically

[29] Analysts often focus on those target companies and can acquire richer information, therefore giving reliable implications on any nonpublic or missing information, while parsimonious models suffer from the lack of public information.

matched with analyst forecast. The section starts from the overall assessment based on the whole sample forecast results. Therefore, the focus is firstly on a "like for like evaluation". Then by matching the IBES "ticker" with the Compustat "gvkey", the forecast accuracy of covered firms is compared. Finally, this paper further compares the statistics of covered and non-covered samples.

Table 5 describes the forecast accuracy of the chosen models for all samples without any partition. The first panel shows results for one-year-ahead forecasts. The comparison of forecast accuracy is implemented among models and earnings definitions. The results conform to the finding of Hess et al (2019) that forecast accuracy varies among earnings definitions. In particular, E2 exhibits the best forecast accuracy while E3 and E4 are almost alike.

Table 5: Comparison of mean forecast accuracy, the whole samples base, 1980 - 2009.

Earnings Definition	HVZ	RI	EP	HVZ-EP	HVZ-RI	RI-EP
Panel A : accuracy of model-based forecasts Et+1 (mean)						
E1	0.2663***	0.2062***	0.2935***	-0.0272*	0.0601*	-0.0873
E2	0.0757***	0.0764***	0.0761***	-0.0004***	-0.0007***	0.0003***
E3	0.1148***	0.1159***	0.1123***	0.0026**	-0.0011***	0.0036***
E4	0.102***	0.1041***	0.1007***	0.0013***	-0.002***	0.0034***
E1-E3	0.1515***	0.0903***	0.1813***			
E2-E4	-0.0263***	-0.0277***	-0.0246***			
E3-E4	0.0128***	0.0118***	0.0116***			
E2-E3	-0.0391***	-0.0395***	-0.0362***			
Panel B : accuracy of model-based forecasts Et+2 (mean)						
E1	0.2803***	0.2692***	0.3999***	-0.1196	0.0112	-0.1307
E2	0.0908***	0.0922***	0.0918***	-0.001***	-0.0014	0.0004**
E3	0.1282***	0.1279***	0.1244***	0.0038***	0.0003***	0.0035***
E4	0.1153***	0.1172***	0.1145***	0.0008***	-0.0019	0.0027***
E1-E3	0.1522***	0.1413***	0.2755***			
E2-E4	-0.0245***	-0.025***	-0.0227***			
E3-E4	0.0129***	0.0107***	0.0099***			
E2-E3	-0.0374***	-0.0357***	-0.0326***			
Panel C : accuracy of model-based forecasts Et+3 (mean)						
E1	0.281***	0.3054***	0.4531***	-0.1721	-0.0244	-0.1476
E2	0.0978***	0.1005***	0.1003***	-0.0026***	-0.0027***	0.0002***
E3	0.1293***	0.1331***	0.1286***	0.0007***	-0.0037***	0.0044***
E4	0.1209***	0.124***	0.1199***	0.001***	-0.0032***	0.0042***
E1-E3	0.1517***	0.1724***	0.3244***			
E2-E4	-0.0231***	-0.0235***	-0.0195***			
E3-E4	0.0084***	0.009***	0.0088***			
E2-E3	-0.0315***	-0.0326***	-0.0283***			

Following Hess et al. (2019) forecast accuracy is the absolute difference between actual earnings and forecasted earnings scaled by end of June price for scaled forecasts and corresponding market capitalization for the unscaled forecasts. Therefore, the forecast accuracy is the absolute deviation from the actual figures. The smaller the figure of forecast accuracy, the better the performance of models is. Unlike Hess et al (2009), the above table shows the Newey-West (1987) time series averages of annual mean rather than median of each earnings definition for each model. Therefore the results shown are bigger than the previous studies due to extreme values.

Panel A in the table reports the adjusted forecast accuracy for one-year-ahead forecasts. The results of cross comparison among models and earnings definitions are shown next to the main comparison. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. I find that the RI model performs the best among the three models when using E1 (0.2061 compared to 0.2663 and 0.2935 for HVZ and EP models, respectively), while the EP model is the best when using E4 (0.1007 compared to 0.1020 and 0.1041 HVZ and RI models respectively). This indicates that earnings definitions affect the comparison of forecast performance among models. Comparing E1 and E3, which are only different in scaling, E3 implies better forecast performance (the results shown in the columns of "E1 - E3" are all positive). Since E1 is the unscaled earning, regression coefficients using observations of large-size firms tend to dominate the results of that rolling window. However, firms of the small size are larger in sample number (Hess et al. 2019). Therefore scaling reduces the impact thus increasing the forecast accuracy. Compared with E4, using E2 leads to the best forecast accuracy in all models. This verifies the hypothesis that special items, which are settled within management control are more discretionary than extraordinary items. Earnings that exclude special items have more persistence and thus more indication for future earnings, since they are within the control of management. By comparing the results in the last four rows within each panel, the sequence of forecast performance of earnings definitions follows E2, E4, E3 and E1, from the best to the worst.

The last three columns compare the forecast accuracy between the three models. Generally, the differences are smaller between models than between earnings definitions. For example, the difference between the HVZ and the EP for one-year ahead regression is only 0.0013 for E4 forecasts. However, for E1 forecasts the difference is -0.0272. On the other hand, the differences between models are also not that significant. In other words, the difference from earnings definitions is more significant than those from various models. Finally, I get some unexpected and contradicting results compared to LM (2014) and Hess et al.(2019). Specifically, I find that the forecast accuracy performance of the RI model is not the best, as the forecasts accuracy of three models approaches.

Table 6 provides the results for comparing model forecasts with analysts' forecasts by showing the forecast accuracy of the partitioned samples of covered and non-covered firms. Since E2 results in the best performance as indicated above, I use E2, which is earnings excluding special items, to compare analyst forecasts. This choice conforms to Hess et al. (2019). Likewise, differences in accuracy between model forecasts and analysts' forecasts are calculated. As shown by my results, analysts' forecasts are more accurate than model forecasts at a shorter horizon in terms of mean values, and dominantly accurate in terms of median values. The analysts' median forecast accuracy is about 0.0042 for one-year-ahead forecasts one-year-ahead while the average median accuracy of models forecasts are 0.0247, far larger than model forecast. However, using mean values I cannot reach the same conclusion. Analysts' forecasts tends to be accurate due to the information advantage that analysts have. Ball and Ghysels (2017) indicate while public press updates are updated on a quarterly base, analysts are privy to high frequency data – daily or weekly.

Table 6: Comparison of mean forecast accuracy, firms not-/covered by analysts, 1980 - 2009.

Forecast	EP(E2)	HVZ(E2)	RI(E2)	Analyst	EP-Analyst	HVZ-Analyst	RI-Analyst
			Panel A: analyst-covered samples accuracy mean				
Et+1	0.0712***	0.0723***	0.0724***	0.0555***	0.0158***	0.0168***	0.0169***
Et+2	0.0865***	0.0879***	0.0882***	0.087***	-0.0005	0.0009***	0.0012***
Et+3	0.0946***	0.0951***	0.0968***	0.1197***	-0.0251	-0.0246	-0.0229
			Panel B: non-covered samples accuracy mean				
Et+1	0.1177***	0.114***	0.1223***				
Et+2	0.1402***	0.1292***	0.136***				
Et+3	0.155***	0.134***	0.1422***				

Lastly, by comparing the forecast accuracy of non-covered and covered firms. I reach similar conclusions as Hess et al. (2019) that models generate accurate forecasts for analyst-covered firms. Table 6 also compares the forecast accuracy between covered and non-covered samples. The forecast accuracy of non-covered samples is, in all instances, bigger than that of covered samples. For example, the HVZ model exhibits 0.114 in forecast accuracy for non-covered samples, while only 0.0723 for covered-samples. As discussed before, the mechanical models require accurate input of fundamental information. Hess et al. (2019) assume that the public information of the analyst-covered firms is more stable and reliable than that of the non-covered firms. Dichev and Tang (2009) mention that earnings volatility leads to less predictable future earnings. Therefore, the volatility in the non-covered firms potentially leads to reduced forecast accuracy.

In conclusion, Table 5 reveals the performance of three mechanical model forecast using different earnings definitions. However, there are many interesting results in my work: For example, I do not find the dominant performance of the RI model. Also, the inferences derived using average mean values of forecast accuracy diverge when using average median values. Though there are dissimilarities compared with previous studies, there are some consistencies with the earlier studies. For example, the performance between models varies slightly, implying that a comparison between earnings definitions is more meaningful than among models. Besides, the earnings excluding special items tend to be more persistent than that excluding extraordinary items. Finally analysts forecasts are generally more accurate than model forecasts.

5.5 Forecast bias

The second part of performance analysis is based on the results of forecast bias. The forecast bias is defined as the difference of the actual earnings minus forecasted earnings scaled by the end-of June price for scaled version and by market capitalization for the unscaled version. Compared to the results of forecast accuracy forecast biases turn to be more accurate, as the values converge to the results in the previous literature. Again, the samples are partitioned into covered and non-covered firms and the analysis begins from a whole sample base.

Table 7: Comparison of mean forecast bias, whole sample base, 1980-2009

Earnings Definition	HVZ	RI	EP	HVZ-EP	HVZ-RI	RI-EP
Panel A : Bias of model-based forecasts Et+1 (mean)						
E1	-0.0435***	-0.0125	0.1062***	-0.1497***	-0.0310***	-0.1187
E2	-0.0037***	-0.0019***	0.0042***	-0.0079***	-0.0019***	-0.0061***
E3	-0.0211***	-0.0169***	-0.0116***	-0.0095***	-0.0042***	-0.0053***
E4	-0.0178***	-0.0145***	-0.0103***	-0.0075***	-0.0033***	-0.0041***
E1-E3	-0.0224***	0.0044***	0.1179***			
E2-E4	0.0141***	0.0126***	0.0145***			
E3-E4	-0.0033***	-0.0025***	-0.0013***			
E2-E3	0.0174***	0.0151***	0.0158***			
Panel B : Bias of model-based forecasts Et+2 (mean)						
E1	-0.0753***	-0.0737***	0.1038***	-0.1791***	-0.0016***	-0.1775***
E2	-0.0136***	-0.0122***	-0.0026***	-0.0110***	-0.0014***	-0.0096***
E3	-0.0374***	-0.0321***	-0.0277***	-0.0097***	-0.0053***	-0.0045***
E4	-0.0342***	-0.0305***	-0.0255***	-0.0087***	-0.0037***	-0.0050***
E1-E3	-0.0379***	-0.0416***	0.1314***			
E2-E4	0.0206***	0.0183***	0.0229***			
E3-E4	-0.0032***	-0.0016***	-0.0021***			
E2-E3	0.0238***	0.0199***	0.0251***			
Panel C : Bias of model-based forecasts Et+3 (mean)						
E1	-0.0901***	-0.0837***	0.1114***	-0.2015***	-0.0064***	-0.1951*
E2	-0.0143***	-0.0129***	-0.0033***	-0.0110***	-0.0015***	-0.0095***
E3	-0.0382***	-0.033***	-0.0311***	-0.0071***	-0.0052***	-0.0019***
E4	-0.0342***	-0.032***	-0.0297***	-0.0045***	-0.0022***	-0.0023***
E1-E3	-0.052***	-0.0507***	0.1425***			
E2-E4	0.0199***	0.0191***	0.0263***			
E3-E4	-0.0040***	-0.0010***	-0.0014***			
E2-E3	0.0238***	0.0201***	0.0278***			

Table 7 reports the models' forecast biases. Moreover, cross comparison for a "like for like evaluation" of earnings definitions among different models is incorporated in the last three columns and last four rows. Except for the two exceptions shown in the EP model, the forecast biases are negative in sign. As a result, I interpret biases in terms of absolute values for most cases. The following are the findings: Firstly, the earnings definitions tend to affect the bias as significantly as the accuracy. Similar to the results in forecast accuracy, E1 results in the largest forecast bias among all the models. For example, the bias of two-year-ahead forecast in the RI model is -0.0737 for E1 forecasts while only -0.0321 for E3 forecasts. The biases for E2 forecasts remain the smallest among all four earnings definitions and exhibit better performance compared to E4 forecasts.

The differences in bias between individual models are pronounced, especially when using E1 and E2 as the earnings. This is in accordance with neither the finding in the section of forecast accuracy nor the finding of Hess et al (2019). For example, the differences in bias between the HVZ and the EP model for the E1 forecast is -0.1497. In contrast, the biases among earnings definitions within the HVZ model have an average of -0.0215. However, I do find that the RI model has the smallest forecast bias. The forecast biases of the RI model are all along smaller than both the HVZ model and the EP model. The finding further supports the conclusion that differences among models, are under most circumstances, highly significant.

By analyzing the samples of covered firms, I show in Table 8 the time series averages of mean bias for three models using E2 and analysts' forecast. Like the model forecasts, the analysts' forecasts are negatively biased at all forecast horizons and are generally much more biased than the model forecasts. For example, the mean of analyst forecasts bias for one-year ahead is -0.0420, while the HVZ counterpart is only -0.0028, far more smaller than analysts' figures. This finding conforms to HVZ (2002) and Hess et al. (2019). As Hess et al. (2019) mention, forecast bias with opposite signs is not comparable, I only show the result of forecast bias of non-covered samples in table below.

Table 8: Comparison of mean forecast bias, firms not-/covered by analysts, 1980-2009.

Forecast	EP(E2)	HVZ(E2)	RI(E2)	Analyst	EP-Analyst	HVZ-Analyst	RI-Analyst
Panel A: analyst-covered samples bias mean							
Et+1	0.003***	-0.0028***	-0.0024***	-0.042***	0.045***	0.0392***	0.0396***
Et+2	-0.0043***	-0.0127***	-0.0123***	-0.0654***	0.0611***	0.0527***	0.0531***
Et+3	-0.0059***	-0.0136***	-0.0134***	-0.089***	0.0831***	0.0754***	0.0756***
Panel B: non-covered samples bias mean							
Et+1	0.0052***	-0.0095***	0.0082***				
Et+2	0.0096***	-0.0234***	-0.0092				
Et+3	0.0189***	-0.0208***	-0.0066				

In summary, in the second part of performance analysis I still get some inconsistent results with the previous studies. The consistent conclusions are firstly the driver roles of different earnings definitions. Secondly, the analysts forecasts are more upward biased.

Finally this paper implements robustness tests as well. Firstly, since E1 and E3 are not strict counterparts (see explanation in footnote 17), E1 is again scaled by common shares outstanding for precise comparison. According to Gerakos and Gramacy (2013), winsorizing unscaled variables causes a reduction in forecast performance. However, this does not hold through my examination. I find quite the opposite results. I examine all three models using E1 for one to three years forecast. For the HVZ model using unwinsorized variables leads to improvement neither on coefficients estimation nor on R^2 [30]. In addition, model forecasts become more biased and less accurate. For the RI model, the coefficients for $NegE_1$ and $NegE^*E_1$ become significant at a 5% level. However, the forecast accuracy is even worse than before, although forecast bias improves slightly for two and three years ahead. For the EP model unwinsorizing variables greatly improve forecast performance in terms of forecast accuracy and forecast bias. The forecasts are more accurate and less biased. One potential reason for this finding is that the EP model needs less information than the other two models and winsorizing reduces observations' information content. Just like Tukey (1960) state, the winsorized observations are not real samples. On the other hand, since all EP inputs are earnings related, the outliers can be ignored.

Table 9: Forecast accuracy and bias after applying scaled E1 to the HVZ, the EP and the RI model

	Mean_forecast_bias	Mean_forecast_accuracy	R-squared
HVZ-E1_1	-0.1352	0.3257	0.6920
HVZ-E1_2	-0.1676	0.3303	0.5728
HVZ-E1_3	-0.2097	0.3566	0.5198
HVZ-E1_4	-0.2691	0.3830	0.4787
HVZ-E1_5	-0.3643	0.4789	0.4294
RI-E1_1	0.0674	0.2796	0.7164
RI-E1_2	0.0405	0.3494	0.5966
RI-E1_3	0.0400	0.4081	0.5393
RI-E1_4	0.0200	0.4297	0.5003
RI-E1_5	-0.0009	0.4134	0.4509
EP-E1_1	0.0254	0.1976	0.7200
EP-E1_2	-0.0019	0.2620	0.5909
EP-E1_3	0.0032	0.3134	0.5089
EP-E1_4	-0.0083	0.3335	0.4877
EP-E1_5	-0.0138	0.3335	0.4525

The other robustness test implemented is related to LM (2014, pg. 1158). The au-

[30]The variables that are insignificant remain insignificant

thors do not strictly follow a fixed rolling window, thus ensuring the same amount of "train samples" for regression at each horizon. However, I find that the adjustment does not improve the model estimation, as there are almost no changes in estimated coefficients nor statistical significance. Besides, no change in forecast bias and forecast accuracy appear. Lastly, employing LM's regression method does not generate more forecast results. The table above shows partially the results of the robustness test.

6 Conclusion

Earnings forecast serves as an essential input for corporate valuation. Thus, studies on this topic are littered in literature. While analysts often generate accurate and reliable earnings forecasts, they are potentially exposed to conflict of interests. For this reason, researchers develop various mechanical models for forecasting earnings. Initial models characterize the future earnings as the past earnings, exploiting the time series properties of the earnings. Examples of these models are AR(1), MA(1) or ARIMA. There are also models based on LOGIT regression to predict the trend of the future earnings. However, since those time-series models suffer from fewer observations, parsimonious models using pooled time series cross-sectional data are developed by Hou et al. (2012), Li and Mohanram (2014).

This paper has the following contributions. Firstly, I summarize the key literature in earnings forecasts, which have been long forgotten. The theoretical background and development path of the mechanical models are thoroughly discussed. Secondly, in addition to following the empirical method in Hess et al. (2019), this paper examines the fourth definition of earnings, "E4" presented in Fairfield et al. (1996). The empirical results suggest that special items and extraordinary items differ in their impact on earnings persistence. Earnings per share excluding special items (E2) is the best predictor for future earnings and should be used in earnings forecast. Its forecast performances are better than earnings per share excluding extraordinary items (E4). Although this finding has been made by Fairfield et al. (1996), I do not find any paper in which the E4 is applied to the HVZ, EP and RI models to examine the difference between two earnings definitions. Thirdly, the scaling effect is also studied. Although Hou et al. (2012) , Li and Mohanram (2014) also study the scaling effect, they do it in the robustness test. By contrast, this paper shows the difference directly by defining E1 (dollar earnings) and E3 (earnings per share) for comparison. The forecast errors are adjusted respectively according to earnings' nature. I find that forecasts using scaled variables lead to better performance, since the scaled variables are steady without too many extreme values. Hence the winsorization does not damage the information content too much.

This paper generally reaches the similar conclusions as to the previous studies. The consistent conclusions are shown as follows. Firstly, the comparison among mechanical models can be misleading, while that among earnings definitions is meaningful. Secondly, the performance of models varies when different earning definitions are used as the model input. However, there are some unexpected results and different findings. The computed forecast accuracy is larger in average than Hou et al. (2012), Li and Mohanram (2014) and Hess et al. (2019). The author did not detect and correct the large values. Thirdly, there are fewer forecasts generated in my work. The models only produce a limited number of forecasts from 1980 to 2009. By contrast, Hess et al. (2019), who employ the same sample periods as in this paper, provide model forecasts from 1982 to 2014. Lastly, I do not find that the RI model is the best performer in forecasting earnings. I incline with Gerakos and Gramacy (2013) that a simple form like the EP model performs the best when using scaled E1, as shown in the robustness test.

The potential reasons for the above mentioned divergence are probably issues with data preparation. This stands as a limitation to this study. There is another limitation. This paper only studies the three typical mechanical models. The other latest models like in Harris and Wang (2019) and Ball and Ghysels (2017) are not examined. The authors argue that their models are outperformers among all the mechanical models and analyst forecasts. Future work can also extend those models which generally add the analysts' forecast into models or use different datasets that feature a high frequency (like the work of Lorsbach 2019).

References

Abarbanell, Jeffery and Brian J. Bushee, 1997, Fundamental Analysis Future Earnings, and Stock Prices, *Journal of Accounting Research* 35, 1–24.

Albrecht, Ws, Ll Lookabill and Jc McKeown, 1977, Time-Series Properties Of Annual Earnings, *Journal of Accounting Research* 15, 226–244.

Azevedo, Vitor, Patrick Bielstein and Manuel Gerhart, 2021, Earnings forecasts: the case for combining analysts' estimates with a cross-sectional model, *Review of Quantitative Finance and Accounting* 56.

Ball, Ryan and Eric Ghysels, 2017, Automated Earnings Forecasts: Beat Analysts or Combine and Conquer?, *Management Science* 64.

Ball, Ray and Ross Watts, 1972, Some Time Series Properties of Accounting Income, *The Journal of Finance* 27, 663–681.

Baltagi, Badi, 2008, *Econometric analysis of panel data.* (John Wiley & Sons).

Banker, Rajiv D and Lei Chen, 2006, Predicting earnings using a model based on cost variability and cost stickiness, *The accounting review* 81, 285–307.

Beaver, William H, 1970, The time series behavior of earnings, *Journal of Accounting Research* pp. 62–99.

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

Brown, Lawrence D and Michael S Rozeff, 1979, Univariate time-series models of quarterly accounting earnings per share: A proposed model, *Journal of Accounting Research* pp. 179–189.

Call, Andrew C, Max Hewitt, Terry Shevlin and Teri Lombardi Yohn, 2016, Firm-specific estimates of differential persistence and their incremental usefulness for forecasting and valuation, *The Accounting Review* 91, 811–833.

DeAngelo, Linda Elizabeth, 1986, Accounting Numbers as Market Valuation Substitutes: A Study of Management Buyouts of Public Stockholders, *The Accounting Review* 61, 400–420.

Dechow, Patricia, Weili Ge and Catherine Schrand, 2010, Understanding earnings quality: A review of the proxies, their determinants and their consequences, *Journal of accounting and economics* 50, 344–401.

Dichev, Ilia D and Vicki Wei Tang, 2008, Matching and the changing properties of accounting earnings over the last 40 years, *The accounting review* 83, 1425–1460.

Dichev, Ilia D and Vicki Wei Tang, 2009, Earnings volatility and earnings predictability, *Journal of accounting and Economics* 47, 160–181.

Evans, Mark E, Kenneth Njoroge and Kevin Ow Yong, 2017, An examination of the statistical significance and economic relevance of profitability and earnings forecasts from models and analysts, *Contemporary Accounting Research* 34, 1453–1488.

Fairfield, Patricia M., Richard J. Sweeney and Teri Lombardi Yohn, 1996, Accounting Classification and the Predictive Content of Earnings, *The Accounting Review* 71, 337–355.

Fama, Eugene F and Kenneth R French, 1999, The corporate cost of capital and the return on corporate investment, *The Journal of Finance* 54, 1939–1967.

Fama, Eugene F and Kenneth R French, 2000, Forecasting profitability and earnings, *The Journal of Business* 73, 161–175.

Fama, Eugene F and Kenneth R French, 2006, Profitability, investment and average returns, *Journal of financial economics* 82, 491–518.

Fama, Eugene F and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Feltham, Gerald A. and James A. Ohlson, 1996, Uncertainty Resolution and the Theory of Depreciation Measurement, *Journal of Accounting Research* 34, 209–234.

Finger, Catherine A, 1994, The ability of earnings to predict future earnings and cash flow, *Journal of accounting research* 32, 210–223.

Foster, George, 1977, Quarterly accounting data: Time-series properties and predictive-ability results, *Accounting Review* pp. 1–21.

Freeman, Robert N, James A Ohlson and Stephen H Penman, 1982, Book rate-of-return and prediction of earnings changes: An empirical investigation, *Journal of accounting research* pp. 639–653.

Gerakos, Joseph and Robert Gramacy, 2013, Regression-based earnings forecasts, *Chicago Booth Research Paper*.

Gonedes, N. J. and Roberts, 1975, Autoregression versus Ordinary Regression: Lessons for Inference, .

Graham, Benjamin, David L Dodd and Sidney Cottle, 1934, Security analysis: Principles and technique, .

Green, David, 1964, Towards a theory of interim reports, *Journal of Accounting Research* pp. 35–49.

Griffin, Paul A, 1977, The time-series behavior of quarterly earnings: preliminary evidence, *Journal of Accounting Research* pp. 71–83.

Harris, Richard DF and Pengguo Wang, 2019, Model-based earnings forecasts vs. financial analysts' earnings forecasts, *The British Accounting Review* 51, 424–437.

Haugh, Larry D, 1976, Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach, *Journal of the American Statistical Association* 71, 378–385.

Hiestand, Thomas et al., 2005, Using pooled model, random model and fixed model multiple regression to measure foreign direct investment in Taiwan, *International Business & Economics Research Journal (IBER)* 4.

Hou, Kewei and David T Robinson, 2006, Industry concentration and average stock returns, *The Journal of Finance* 61, 1927–1956.

Hou, Kewei, Mathijs A Van Dijk et al., 2010, Profitability shocks and the size effect in the cross-section of expected stock returns, *Fisher College of Business Working Paper* pp. 1–3.

Hou, Kewei, Mathijs A Van Dijk and Yinglei Zhang, 2012, The Implied Cost of Capital: A New Approach, *Journal of Accounting and Economics* 53, 504–526.

Jones, Jennifer J, 1991, Earnings management during import relief investigations, *Journal of accounting research* 29, 193–228.

Kormendi, Roger and Robert Lipe, 1987, Earnings Innovations, Earnings Persistence, and Stock Returns, *The Journal of Business* 60, 323–345.

Lambert, D, ZP Matolcsy and A Wyatt, 2009, How do analysts forecast earnings?, in *Accounting and Finance Association of Australia and New Zealand Conference*. AFAANZ.

Leone, Andrew J, Miguel Minutti-Meza and Charles E Wasley, 2019, Influential observations and inference in accounting research, *Accounting Review* 94, 337–364.

Lev, Baruch, 1969, Industry Averages as Targets for Financial Ratios, *Journal of Accounting Research* 7, 290–299.

Lev, Baruch and S Ramu Thiagarajan, 1993, Fundamental information analysis, *Journal of Accounting research* 31, 190–215.

Li, Kevin and Partha Mohanram, 2014, Evaluating Cross-Sectional Forecasting Models for Implied Cost of Capital, *SSRN Electronic Journal*.

Li, Kevin Ke, 2011, How well do investors understand loss persistence?, *Review of Accounting Studies* 16, 630–667.

Lorek, Kenneth S, 1979, Predicting annual net earnings with quarterly earnings time-series models, *Journal of Accounting Research* pp. 190–204.

Lorek, Kenneth S, Charles L McDonald and Dennis H Patz, 1976, A comparative examination of management forecasts and Box-Jenkins forecasts of earnings, *The Accounting Review* 51, 321–330.

Lorsbach, Tobias, 2019, Three essays on the performance of earnings forecasts, bankruptcy predictions and textual analysis, Ph.D. thesis Universität zu Köln.

McVay, Sarah Elizabeth, 2006, Earnings management using classification shifting: An examination of core earnings and special items, *The Accounting Review* 81, 501–531.

Miller, Merton H and Franco Modigliani, 1961, Dividend policy, growth, and the valuation of shares, *the Journal of Business* 34, 411–433.

Nelson, Charles R, 1974, The first-order moving average process: Identification, estimation and prediction, *Journal of Econometrics* 2, 121–141.

O'brien, Patrick K., 1988, The political economy of British taxation, 1660-1815, *The Economic History Review* 41, 1–32.

Ohlson, James A, 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of accounting research* pp. 109–131.

Ou, Jane A and Stephen H Penman, 1989, Financial statement analysis and the prediction of stock returns, *Journal of accounting and economics* 11, 295–329.

Penman, Stephen H and Xiao-Jun Zhang, 2002, Accounting conservatism, the quality of earnings, and stock returns, *The accounting review* 77, 237–264.

Piotroski, Joseph D, 2000, Value investing: The use of historical financial statement information to separate winners from losers, *Journal of Accounting Research* pp. 1–41.

Pope, Peter and Pengguo Wang, 2005, Earnings Components, Accounting Bias and Equity Valuation, *Review of Accounting Studies* 10, 387–407.

Richardson, Scott A, Richard G Sloan, Mark T Soliman and Irem Tuna, 2005, Accrual reliability, earnings persistence and stock prices, *Journal of accounting and economics* 39, 437–485.

Sloan, Richard G, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, *Accounting review* pp. 289–315.

Swamidass, P. M. (ed.), 2000, *Forecast errors*Springer US Boston, MA pp. 224–224.

Tukey, J. W., 1960, A survey of sampling from contaminated distributions, *Contributions to Probability and Statistics* pp. 448–485.

Watts, R.L., 1975, *The Time Series Behavior of Quarterly Earnings.* (Department of Commerce, University of Newcastle).

Weiss, Dan, Prasad Naik and Chih-Ling Tsai, 2008, Extracting Forward-Looking Information from Security Prices: A New Approach, *Accounting Review* 83.

West, Kenneth and Whitney Newey, 1987, A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703–08.

Wooldridge, Jeffrey M, 2010, *Econometric analysis of cross section and panel data.* (MIT press).