Earnings Prediction Using Recurrent Neural Networks

MORITZ SCHERRMANN,* RALF ELSAS†

Abstract

Firm disclosures about future prospects are crucial for corporate valuation and compliance with global regulations, such as the EU's MAR and the US's SEC Rule 10b-5 and RegFD. To comply with disclosure obligations, issuers must identify nonpublic information with potential material impact on security prices as only new, relevant and unexpected information materially affects prices in efficient markets. Financial analysts, assumed to represent public knowledge on firms' earnings prospects, face limitations in offering comprehensive coverage and unbiased estimates. This study develops a neural network to forecast future firm earnings, using four decades of financial data, addressing analysts' coverage gaps and potentially revealing hidden insights. The model avoids selectivity and survivorship biases as it allows for missing data. Furthermore, the model is able to produce both fiscal-year-end and quarterly earnings predictions. Its performance surpasses benchmark models from the academic literature by a wide margin and outperforms analysts' forecasts for fiscal-year-end earnings predictions.

---

*   Institute for Finance & Banking, Ludwig-Maximilians-Universität München, Ludwigstr. 28 RB, 80539 Munich, Germany. E-mail: scherrmann@lmu.de

†   Institute for Finance & Banking, Ludwig-Maximilians-Universität München, Ludwigstr. 28 RB, 80539 Munich, Germany. E-mail: elsas@lmu.de

1. Introduction

Firm disclosures regarding future prospects are critical for determining corporate valuation, necessitating global regulations such as the European Union's Market Abuse Regulation (MAR) and the United States' Securities and Exchange Commission's Rule 10b-5 and Regulation Fair Disclosure (RegFD). These regulations mandate the accurate and timely disclosure of material nonpublic information. However, issuers encounter difficulties in identifying nonpublic information that could impact security prices, as only new, relevant and unexpected information causes price fluctuations in efficient markets. Consequently, a benchmark for "new" news is essential for issuers to comply with disclosure obligations.

In this context, financial analysts are often assumed to represent public knowledge concerning firms' earnings prospects. The German security markets supervisory authority (BaFin) recently even designated analysts' predictions as their regulatory measure for capital market expectations on firms' earnings. Given financial analysts' important role in advising investors and providing insights into earnings prospects through the publication of estimates on future earnings and related key ratios, such a "benchmark" initially appears reasonable.

However, reliance on analysts has certain limitations. Firstly, their estimates may be subject to biases arising from conflicts of interest and existing research shows limited predictive power in explaining stock price reactions to earnings disclosures (see Abarbanell, 1991; O'brien, 1988; Bradshaw, Drake, Myers, and Myers, 2012). Secondly, analyst coverage is limited, particularly outside the US.[1] Moreover, recent regulatory compliance burdens have prompted analyst firms and financial intermediaries to further reduce coverage, especially for smaller, less prominent companies.

In this study, we develop a neural network to forecast future firm earnings, utilizing approximately four decades of financial data. Machine learning models can effectively process and analyze large amounts of data to predict corporate earnings. By employing advanced algorithms, these models can quickly sift through historical data, financial statements and market trends to generate accurate forecasts. This automation potentially addresses the gaps in analysts' coverage and the model may even reveal hidden insights by decoding complex patterns in the input data. Consequently, such an earnings prediction model provides a more comprehensive benchmark for the capital market's expectation concerning firms' earnings

---

[1]   For example, in Germany in the period from 1998 to 2022, less than 50% of CDAX companies received coverage for fiscal-year-end earnings and less than 20% for quarterly earnings.

prospects, thereby assisting issuers in better adhering to disclosure regulations and improving guidance for investors' investment decisions.

Key features that differentiate our model from previous statistical approaches in the literature are that we avoid selectivity biases through high input data demand, as we strive for a parsimonious specification and explicitly allow for missing data in the time-series of accounting data used as predictors. This approach ensures that our model is not biased towards firms with more complete or longer available data, thus avoiding survivorship biases, which is a common issue in previous studies. Also different from most of the academic literature, we provide forecast results not just for fiscal-year-end predictions but also for quarterly earnings. This allows us to capture and study the dynamics of earnings over shorter time periods and is particularly useful for issuers and investors, as quarterly earnings announcements are the most important source of information that determines stock prices.

The structure of our study is organized as follows. In Section 2, we provide a brief overview of the existing literature on earnings prediction and discuss how our study contributes to this field. Section 3 describes the design of the neural network used for earnings prediction, the training procedure and the benchmarks used to validate the model's predictive power. We emphasize that the model's performance is always measured using out-of-sample data not used in the model's training. In Section 4, we describe the data used in our study, including its selection and provide descriptive statistics.

Section 5 contains the empirical results. We report the performance of our model as well as benchmark models, including a random walk as a time-series model, a cross-sectional model by Hess and Lorsbach (2019) and analysts' forecasts. We report performance results for both end-of-fiscal-year and quarterly earnings predictions, for the overall period of analysis as well as subsamples. These subsamples differentiate between predictive performance for different firm sizes, industry roots of firms and the time-period before, during and after the COVID-19 pandemic. This last conditioned subsample allows us to test model performance during a period characterized by a severe and unexpected (i.e., exogenous) shock to firm operations, as the COVID-19 pandemic led to shutdowns with regard to operations, the rise of home-office-work, associated changes in consumer demand and temporary financial squeezes. Lastly, we report results for a matched-firm approach that allows us to test model performance while controlling for the selectivity introduced by running all previous tests on the subsample of firms with analyst coverage. Section 6 concludes our study.

3

In terms of results, our predictive model is at least at par with analysts' forecasts, but performance quite generally is higher for fiscal-year-end forecasts, also when conditioning on industries, firm size and the time-period of the COVID-19 pandemic. Used benchmark models are consistently outperformed by both our model as well as analysts' forecasts by a wide margin.

2. Literature Review

We broadly divide the stream of literature that compares the accuracy[2] of model-based forecasts and analysts' forecasts by the used methodology, i.e. cross-sectional and time-series models and discuss cross-sectional predictive models first.

The literature stream of cross-sectional models starts with Fama and French (2000), where the authors predict future annual profitability and earnings using year-by-year cross-sectional regressions between 1964 and 1996. They use the last annual earnings result, a firm's market value, total dividends as well as a dividend dummy to predict the next years earnings. Hou, Van Dijk, and Zhang (2012) extend the idea of Fama and French (2000) by regressing future annual earnings on total assets, dividends, earnings and accruals. Li and Mohanram (2014) use in addition past earnings and book values instead of dividends as predictors. Despite the fact that all these modelling steps lead to improvements in terms of accuracy and earnings response coefficients[3], none of these models is able to outperform the predictive performance of analysts.

Hess and Lorsbach (2019) contend that the enhanced performance of analysts relative to cross-sectional models can be attributed to their informational advantage, particularly as these studies often overlook quarterly data that analysts routinely utilize. Hess and Lorsbach (2019) use the information of already reported quarterly results to predict the sum of all quarters that are not reported yet. Hence, the authors annual earnings prediction is the sum of reported and predicted quarters. By incorporating quarterly earnings results, Hess and Lorsbach (2019) strongly improve

---

[2]   Accuracy in this context measures the difference between the predicted and actual earnings. We will use the Median Percentage Difference (MPD) and Median Absolute Percentage Difference (MAPD) in our analyses. See Section 3.4.

[3]   The Earnings Response Coefficient (ERC) is one measure of predictive model performance which we will also rely upon in our analyses. The ERC is calculated by regressing the stock price's abnormal return (the return that exceeds the expected return) on the unexpected earnings (the difference between actual earnings and expected earnings). The predictive model's earnings prediction is then used as the expected earning. A model performs better, all else equal, when it explains more of the abnormal returns on average.

forecast accuracy over short- and long-term horizons. However, analyst forecasts are still superior in terms of accuracy.[4]

There is rich literature on time-series models for earnings prediction between 1978 and 1995.[5] The main conclusion of that literature is that analysts are better at predicting future earnings than time-series models (Bradshaw et al., 2012). In addition, the studies were criticized due to relying on only small sample sizes, comprising only large, established companies and due to their data intensity, see e.g. Kothari (2001).

To identify a time-series model for benchmarking purposes, we discuss three recent studies. Bradshaw et al. (2012) use a random walk model to predict annual earnings per share (EPS) of US companies between 1983 and 2008. They find that the EPS forecasts of the simple random walk model are more accurate than analysts' forecasts over longer horizons, for smaller or younger firms, when analysts forecast are negative or if there were large changes in EPS.

Ball and Ghysels (2018) augment classical quarterly accounting variables with high-frequent market information as stock returns, stock volatility or oil prices. The authors find that their time-series forecasts are more accurate than analysts' and that combining time-series forecasts with analysts' forecasts systematically outperforms analysts alone.

Similar to our analysis, Elend, Tideman, Lopatta, and Kramer (2020) use a machine learning model to predict future, quarterly earnings. More precisely, they use a long short-term memory neural network (Hochreiter and Schmidhuber, 1997) processing a time-series of 20 periods with 19 accounting predictors from Compustat and 11 stock market predictors from CRSP.

Ball and Ghysels (2018) and Elend et al. (2020) are two of the few studies which predict quarterly instead of annual earnings results. However, due to the large number of predictors required in both approaches, the samples contain mostly large and established firms, thereby likely inducing survivorship bias.

In summary, the previous literature on predicting firm earnings statistically is characterized by using a large number of predictor variables (both cross-sectionally and

---

[4] Azevedo, Bielstein, and Gerhart (2021) use an annual cross-sectional model with analyst forecasts as predictors. Their model outperforms the above discussed benchmarks from the literature in terms of forecast accuracy, but relying on analyst forecast prohibits the model to serve as an alternative to coMPDnsate for low and further reducing coverage by analysts.

[5] See Bradshaw et al. (2012) for a detailed summary of time-series literature comparing time-series and analyst forecasts.

in the time-series), which is likely to introduce selection biases in the results. Also, most studies focus on predicting fiscal-year-end results, ignoring quarterly earnings (both as an input as well as the predicted variable). Our study addresses both issues by relying on cross-sectional and time-series data but allowing for missing data and heterogeneous time-series' length. The model thus mitigates selectivity and survivorship bias. In addition, our neural network will be trained to predict both quarterly and fiscal-year-end earnings.

3. Methodology

3.1. Model

We think that various types of variables contribute to the prediction of a company's future earnings. These range from accounting data available at a quarterly frequency to daily market data, as well as time-invariant factors such as the company's industry or the quarter in which an earnings announcement is MPDe. Classical models like cross-sectional regressions or time series models are not able to handle different input types or they suffer from survivorship or success biases. We aim to create a model which is able to handle time-series inputs as well as time invariant variables at the same time.

Our model is a combination of a recurrent neural network (RNN) and three feed-forward layers. The recurrent neural network consists of two layers of gated recurrent units (GRUs), introduced by Cho, Van Merriënboer, Gulcehre, Bahdanau, Bougares, Schwenk, and Bengio (2014). The layers have 76 and 38 hidden neurons, respectively. We design the RNN to process five years of quarterly accounting data, resulting in a time series of 20 quarters. The RNN outputs a time series of 38 dimensional arrays of which we use only the last one, since this array contains all the information of its predecessors. For market data input, we avoid using its time series structure by considering only the changes in the corresponding variable within the most recent quarter. We also test a design which preserves the time series structure of the market data using a second RNN. However, this model is more complex and data intensive without outperforming the approach using the reduced market data.

We combine the RNN output with the market data and pass it through the feed-forward layers. The model output consists of two values, a quarterly and a yearly earnings prediction. We add dropout and batch normalization between each layer to reduce the likelihood of overfitting. We zero pad each time dependent accounting
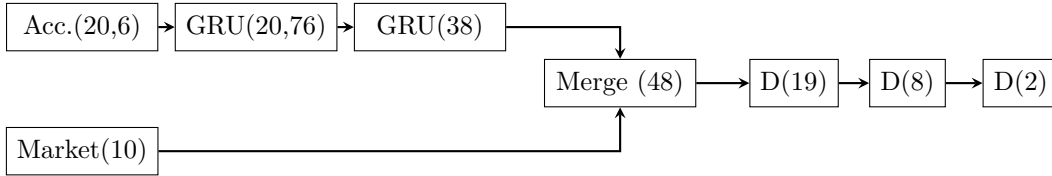
Figure 1: Visualization Model Architecture

input to avoid survivorship biases. To allow for non linear relationships, we add hyperbolic tangent activation functions between the layers. We select 5 years of accounting data to ensure that the model is able to recognize time dependent patterns like seasonal businesses, market cycles or company crises; without harming performance with data which is too old or irrelevant for the actual business. However, we also test other choices for the time series length without changing the results. The same holds for other design choices, like using long short-term memory cells (LSTM, Hochreiter and Schmidhuber (1997)) or different number of hidden neurons. Figure 1 illustrates the model architecture. In the following, we will denote our model with *RNN*.

3.2. Training

We use the adaptive moment estimation method of Kingma and Ba (2014), called Adam, to train the base model with a batch size of 512, a learning rate of 0.0075, exponential decay rates for moment estimation of 0.9 and 0.999 and a dropout rate of 0.001. We evaluate and track the model performance after every epoch on the validation set. An exponential moving average with smoothing factor 0.2 measures the general trend of the validation score. As soon as the moving average increases from one epoch to another, the training stops. In that way, we prevent the model from overfitting on the training set. We train five models with the same setup, but with different random initial model weights. Our final model prediction is the mean of all trained models, as this design reduces the likelihood of a model that performs above or below average just by chance. The loss function is the mean absolute error. We also test a mean squared error loss, but that choice leads the model to fit unusual, extreme earnings better during training, with the drawback of fitting the common and usual earnings less precisely. As our aim is to train a model which is able to predict market expectations of future earnings, we are interested in estimating ordinary earnings results more precisely at the expense of decreasing the forecasting performance of rather unusual results.

3.3. Benchmarks

We test our model against three benchmarks. First, we compare our model to quarterly and annual analyst forecasts. For quarterly forecasts, we use the most recent mean analysts forecast prior to the reporting date of the actual quarterly earnings results of a specific company. This approach yields at most four quarterly analyst forecasts per company and year, one for every quarter. Regarding the annual analyst forecasts, we operate the same way. Note that this approach leads to at most four different annual analyst forecasts per year and company, too. All of these annual forecasts refer to the same annual earnings report, but with different levels of information. In that way, we ensure that neither our model nor the analysts have some informational advantage.

The second benchmark is the augmented residual income model of Hess and Lorsbach (2019). We only choose the residual income model, since the results are very similar across all models that were considered in the study. As we have seen in section 2, the study of Hess and Lorsbach (2019) is the most recent and best performing model within the whole literature stream using cross-section panel regression models, at least if we require no dependency on analyst data. Hess and Lorsbach (2019) use all available information up to the most recent, quarterly or annual, earnings result to predict the cumulative quarterly earnings of the coming quarters until the next annual earnings result. Their annual earnings forecast is the sum of the already known quarterly results and the following, predicted quarterly results. We divide the predicted cumulative earnings for future quarters by the number of the future quarters to get a prediction for the next quarterly earnings result. In that way, we also get a benchmark for the quarterly predictions. We know the model of Hess and Lorsbach (2019) is not designed for quarterly predictions, but as there is no other literature using cross-sectional models that covers quarterly earnings predictions, we assume this approach to be the best proxy for quarterly earnings predictions of cross-sectional models. We will denote this benchmark with *Regression.*

Lastly, we choose the naive random walk model of Bradshaw et al. (2012) as time-series model benchmark, since the authors find strong results despite its simplicity. The other two mentioned models of Ball and Ghysels (2018) and Elend et al. (2020) would have been possible benchmarks too, but their high data requirements complicate an objective comparison in this study. We denote this benchmark with *Random Walk.*

3.4. Performance Measures

We compute all of our performance measures using relative or percentage errors, which means that we divide the difference between true earnings result and model forecast by the true earnings result. The same logic is applied to all benchmarks. The reason for this step is that the difference between the actual and the predicted EPS number alone is not an indication of how much the predicted value has been exceeded. For one company, a difference of one dollar could be very high, while the same difference for another company could be negligible. We assume that the percentage estimation error is more meaningful and enables the comparison of companies with different absolute earnings numbers and different numbers of shares outstanding.

3.4.1. Median Percentage Difference & Median Absolute Percentage Difference

Most of the studies mentioned in section 2 use the forecast bias and accuracy as performance measures for their models. We will use these measures too, but with different names, since bias and accuracy have other meanings in the machine learning literature (Sokolova and Lapalme, 2009). We apply accuracy and bias with their machine learning interpretation in one of our empirical analyses, which would result in a naming conflict. Therefore, we define the forecast bias as the *Median Percentage Difference (MPD)*. It indicates whether a prediction systematically over- or underestimates a ground truth. Accordingly, we define forecast accuracy as *Median Absolute Percentage Difference (MAPD)*, which measures how precise a model is able to predict future earnings.

3.4.2. Sign Prediction

The sign prediction tests the models performance in a scenario where we are only interested in the direction of the change of the earnings results within one year. For quarterly results, this approach implies that we compare changes between the same quarters of subsequent years, so that we account for seasonal businesses. The target is a categorical variable representing either a negative change, no change or a positive change in the true earnings results. We define that no change occurs if the absolute value of the relative earnings change is smaller than 5%, e.g. $|\frac{EPS_t - EPS_{t-1}}{EPS_{t-1}}| < 0.05$. In the same way, we define the categorical earnings change predictions for our model and the benchmarks.

We compute average accuracy, macro precision, macro recall and macro F1 as defined in Sokolova and Lapalme (2009) for our model and all benchmarks. We think that the ability of a model to predict the direction of an earnings change is a key feature since the direction of an earnings change is an important proxy of whether a fiscal year was successful or not.

3.4.3. Earnings Response Coefficient

According to the efficient market hypothesis of Fama (1970), stock prices should react quickly to surprising earnings announcements. We define an earnings surprise as a situation where the market expectation of an earnings results is different to the realized earnings. Therefore, if the market reacts accordingly to high deviations between model prediction and realized earnings, we conclude that the model prediction is a suitable proxy for the market expectation. To do so, we first estimate the abnormal returns at the earnings announcement dates with a three day event window using the market model:

$$r_t^i - r_t^f = \alpha_0 + \alpha_1(r_t^m - r_t^f) + \sum_j \alpha_2^{i,j} D_t^{i,j} + \epsilon_t^i, \tag{1}$$

where r^i is the return of company i, r^f is the risk-free rate (3 months LIBOR), r^m is the return of the S&P 500 index and $D^{i,j}$ is the event dummy of company i at earnings announcement j. Accordingly, $\alpha_2^{i,j}$ is the cumulative abnormal return of company i and earnings announcement j.

Finally, we estimate a year- and firm-fixed effects model with clustered standard errors, specified as follows:

$$
\begin{aligned}
\alpha_2^{i,j} =& \beta_0 surprise_{i,j} \\
&+ \beta_1 ln(TotalAssets_{i,j}) + \beta_2 ln(TotalAssets_{i,j}) \times surprise_{i,j} \\
&+ \beta_3 TobinsQ_{i,j} + \beta_4 TobinsQ_{i,j} \times surprise_{i,j} \\
&+ \sum_y D_{i,j}^y + \epsilon_{i,j},
\end{aligned}
\tag{2}
$$

where the earnings surprise is defined as $\frac{EPS_{i,j} - \widehat{EPS}_{i,j}}{\widehat{EPS}_{i,j}}$ with $EPS_{i,j}$ being the realized earnings per share of company i and earnings announcement j, whereas $\widehat{EPS}_{i,j}$ is its prediction. Coefficient β_0 is the so called *Earnings Response Coefficient (ERC)*. The higher the ERC, the stronger is the relationship between the estimated earnings sur-

prise and the market reaction. $D_{i,j}^{y}$ is the year dummy of year y and announcement j of company i. For quarterly results, we also add quarterly dummies.

4. Data

4.1. Data Selection

Our sample includes all NYSE, Amex and Nasdaq listed securities with stock code 10 or 11 that are in the intersection of the CRSP annual returns file and the Compustat fundamentals annual file between 1990 and 2021. The analyst data originates from the I/B/E/S summary file.

In order to be comparable with other studies, we adapt the accounting data selection from the literature stream using cross-sectional models as Fama and French (2000), Hou et al. (2012), Li and Mohanram (2014), Hess and Lorsbach (2019) or Azevedo et al. (2021). Accordingly, our earnings number is the earnings before extraordinary items, but after dividends paid on preferred stock. As explanatory variable we use lagged earnings as former studies find evidence that this variable has a high explanatory power for future earnings (Bradshaw et al. (2012)). We measure a companies financial stability using the equity ratio, defined as stockholders equity divided by Compustat's total assets. Since Lewellen and Resutek (2019) find a negative relationship between accruals and subsequent earnings, we also add accruals, defined as by Hou et al. (2012). We add total cash dividends as input to our model. Finally, we use total assets as a proxy for size. We scale earnings, dividends and accruals by the weighted average common shares outstanding. The accounting input is a time series of the last 20 quarters for which all of the mentioned variables are available. We use zero padding in cases without sufficient data. The minimum requirement is one observation, to ensure the model to be minimal data intensive. Since occasional quarters can be missing, we add a time variable that indicates how many quarters have passed, expressed as a fraction of a year. In that way we ensure that the model is still able to learn seasonal and economic cycles.

Our company specific market variables are the individual stock returns, volume per share and Tobins's q. The former two variables are proxies for the companies general development and the markets interest in the company. Tobin's q is defined as:

$$Tobin's\ q = \frac{Market\ Value\ of\ Equity\ +\ Book\ Value\ of\ Liabilities}{Book\ Value\ of\ Equity\ +\ Book\ Value\ of\ Liabilities} \quad (3)$$

It is a proxy for the change in market expectations on future earnings of the company. In addition, we incorporate market wide variables as the S&P 500 Index to cover market wide developments and the CBOE Volatility Index to cover market stress. We observe these variables over the last quarter (63 days) prior to the earnings announcement we like to predict.

We compress all stock and market variables as follows: We compute quarterly log returns of the companies' stock price and the market index. We cumulate the companies' volume per share. Tobin's q and the Volatility index are captured as the absolute difference between the start and end of the quarter. If there are less than 63 days of information available, we work with the maximum period length for which data is available. As the period length might have an influence on all the mentioned market variables, we express these market variables on a per-day basis. We add a time variable which denotes the length of the time period.

Finally, we add quarter dummies to make sure that the model is able to learn seasonal effects. We log transform total assets, total cash dividends and the absolute changes in the Volatility Index and Tobins'q to reduce the skewness in these variables. We winsorize all variables at the $1st$ and $99th$ percentile to minimize the effect of outliers. Tables 11 through 14 in the Appendix offer definitions and the winsorization thresholds[6] for the accounting and market variables utilized, in addition to the corresponding Compustat and CRSP variables.

Additionally, we studentize all non-binary inputs as:

$$\tilde{x}_i = \frac{x_i - \bar{x}}{\sqrt{\frac{1}{n} \sum_i (x_i - \bar{x})^2}}, \tag{4}$$

where $\bar{x}$ is the mean of x_i.

4.2. Data Preparation

Table 1 lists all data preparation steps. The upper half of the table describes general steps that we apply to the raw Compustat data. We start with 1,826,340 earnings observations, which contain both annual and quarterly earnings results. After conducting essential cleaning steps, as coverage in CRSP, filtering missing information about the year or quarter of the observation or filtering observations with missing total assets, shares outstanding, earnings or book value of equity, we end up with 669,637 observations.

---

[6]   We use rounded valued close the the respective $1st$ and $99th$ percentiles to simplify replications.

Table 1: Data Preparation

This table presents the data preparation steps for the RNN model and the regression model as defined in section 3. The upper half of the table consists of cleaning steps that belong to both models. The lower half of the table represents steps that are necessary to the specific models.

Preparation Step	Num. Obs.	Num. Obs. Removed
Raw Compustat Data	1,826,340	
No CRSP Match	899,662	-926,678
Missing Year Or Quarter	899,379	-283
Missing Total Assets Or Weighted Average Common Shares	701,786	-197,593
Missing EPS Or BVE	669,637	-32,149

RNN			Regression		
Step	Num. Obs.	Num. Obs. Removed	Step	Num. Obs.	Num. Obs. Removed
Remove Yearly Data	533,430	-136,207	Remove Incomplete Years	651,595	-18,042
Remove Duplicate Quarters	533,219	-211	Remove Q4	521,276	-130,319
Require >1 Obs. Per Comp.	533,065	-154	Remove Year Gaps	459,656	-61,620
Remove Comp. First Obs.	518,713	-14,352			
Remove NaN Or Inf. Obs.	395,681	-123,032			

As firm coverage is a key point in our analysis, we emphasize the data intensity of our model and compare it with the benchmark regression model. Therefore, the lower half of Table 1 compares all cleaning steps for both models separately. Regarding the RNN, we first remove annual observations as we only consider quarterly results for our earnings forecasts.[7] We minimize the data requirements by allowing inputs with only one historical observation. This implies that we remove companies that only have one observation in the sample. We also remove the first observation of each company since these observations have no predecessors which can serve as predictors. Finally, we remove duplicates and missing observations.

Regarding the cross-sectional approach of Hess and Lorsbach (2019), we have to remove all Q4 observations by design of their model. This step is comparable to our removal of annual observations. As the authors use all the earnings data of one year to predict the next year, they have to remove incomplete or missing years out of their sample.

We end up with a sample size of 395,681 for our model compared to a sample size of 459,656 for the benchmark model, which indicates that the data intensity of our model is comparable to the one of the benchmark. This is in contrast to prior time-series models, which have in general higher data requirements than cross-sectional models. Table 2 presents the summary statistics of the accounting and market data.

We split the data set in a training set, a test set and a validation set. The training set contains the oldest 70% of the earnings results, based on the report date, which results in 277,035 observations between 1990 and 2010. This is the data on which we train our model to adjust the models' weights. The test set contains the latest 20% of all earnings results, which implies 79,049 observations between 2014 and 2021. We use the test set to evaluate the performance of the final model and all benchmarks. Accordingly, the validation set contains the remaining 10% of the data, so that this set contains 39,597 observations between 2010 and 2014. The purpose of the validation set is to find the best hyperparameter setup for our model. The reason why we do not use a random split of the data is that through the use of market wide predictors like the S&P 500 or the volatility index, the model might be able to learn patterns in market wide events like the Covid pandemic and may

---

[7]   We employ historical quarterly earnings data to forecast both future quarterly and annual earnings outcomes. To ensure accurate annual predictions based on quarterly results, we make the fundamental assumption that the summation of quarterly diluted earnings per share is equivalent to the annual diluted earnings per share. Notably, our findings reveal that the absolute limits of the 90% confidence interval for the disparity between annual EPS and the sum of quarterly EPS are below 0.2.

Table 2: Summary Statistics

This table presents the summary statistics of the accounting data (upper half) und market data (lower half) as defined in section 4.1. We compute selected percentiles as well as mean, standard deviation and number of observations. The accounting data is given to our model as time-series of the length of 20, which is why we get more than 395,681, as stated in Table 1. The market variables contain compressed information about the last quarter before earnings announcement, defined as follows: We cumulate daily stock returns, market index returns and daily volume. We calculate the absolute change of volatility index and tobins q. The period length, expressed as a fraction of a year, is the minimum of 63 days and the number of days for which information is available. Since some periods are shorter than 63 days, we normalize all market variables (except the period length) on a daily basis. We report the market variables as percentage numbers, except for the period length.

	Count	Mean	Std	1%	25%	50%	75%	99%
Accounting Data								
Earnings Per Share	416,585	0.05	1.59	-3.88	-0.07	0.07	0.28	2.59
Total Assets (Billions)	416,585	2.86	14.13	0.00	0.05	0.21	1.05	49.18
Equity Ratio	416,585	0.54	0.23	0.06	0.36	0.53	0.72	0.96
Dividend Dummy	416,585	0.34	0.47	0.00	0.00	0.00	1.00	1.00
Accruals Per Share	416,585	-0.54	2.11	-9.61	-0.65	-0.12	0.05	3.68
Time To Announcement	6,325,744	2.44	1.52	0.21	1.17	2.27	3.64	5.19
Market Data								
Average Daily Stock Return (%)	395,681	0.09	0.53	-1.17	-0.14	0.07	0.30	1.65
Average Daily Volume Per Share (%)	395,681	0.72	0.83	0.02	0.21	0.47	0.92	4.04
Average Daily Market Return (%)	395,681	0.04	0.13	-0.33	-0.00	0.06	0.12	0.26
Average Daily Change Volatility Index (%)	395,681	0.15	12.00	-35.00	-4.38	-0.59	3.38	36.95
Average Daily Change Tobins q (%)	395,681	-0.08	2.38	-7.83	-0.32	0.00	0.30	5.62
Period Length	395,681	0.25	0.02	0.24	0.25	0.25	0.26	0.27

apply this knowledge to out-of-sample earnings results within the time frame of the market wide event.

5. Empirical Results

This section summarizes our study's empirical results. Sections 5.1 and 5.2 compare the model with all benchmarks with respect to the median absolute percentage difference and the median percentage difference. The former section yields the results for the whole test set, the latter splits them on a quarterly basis. Section 5.3 lists the results per firm sizes, measures by total assets, whereas Section 5.4 compares the performance between different industries. Section 5.5 covers the sign prediction performance. Sections 5.6 and 5.7 each compare the models and benchmarks performance in two contrastive scenarios: Section 5.6 compares the performance results for firms with analyst coverage with firms without analyst coverage. Section 5.7 contrasts the performance measures before the Covid crisis with the measures during the crisis. Finally, section 5.8 contains the results for the earnings response coefficient of all models.

To make the models as comparable as possible, we filter all observations where some data is missing for one or more models. Furthermore, we remove penny stocks, which are observations for which the respective company has a stock price smaller than 5 \$. Finally, as we use percentage errors, we reduce the test set by the 5% smallest and largest prediction errors, since in cases where the predicted earnings result is very small or even zero, the percentage error gets extraordinary high.

5.1. Overall Performance

Table 3 contains the overall performance with respect to MAPD and MPD (see section 3.4.1) of our model and the benchmarks on the test set. Regarding the median absolute percentage error, we see that our model outperforms all benchmarks by at least 4 percentage points for the annual predictions. Additionally, our model is with a median percentage difference of 0.82% less biased than the analyst forecasts with -12.24% and the other benchmarks with 9.76% and 7.55%, respectively. Note that the negative bias in analyst forecasts implies an overly optimistic expectation of future earnings, which is in line with the prior literature (see Abarbanell, 1991; O'brien, 1988; Bradshaw et al., 2012).

For quarterly predictions, we see that the RNN model is less precise than the analysts. However, both the random walk model and the cross-sectional regression are outperformed by more than 8 and 22 percentage points, respectively. Analyst forecasts are, with a median percentage difference of -6.82%, more biased than our

Table 3: Overall performance

This table presents the median absolute percentage difference (MAPD) and the median percentage difference (MPD) between realised and predicted earnings per share, as defined in section 3.4.1, for the RNN model (section 3.1) and all benchmarks. The benchmarks are analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3. We report the results for annual and quarterly predictions. We test for every benchmark whether its median value is significantly different to the median value of the RNN model using the Wilcoxon rank test (Wilcoxon, 1992).

	Annual		Quarterly	
	Median Absolute Percentage Difference	Median Percentage Difference	Median Absolute Percentage Difference	Median Percentage Difference
RNN	21.34 %	0.82 %	30.04 %	0.01 %
Analyst	25.21 %	-12.24 %	26.27 %	-6.82 %
Regression	29.11 %	9.76 %	52.72 %	14.67 %
Random Walk	35.48 %	7.55 %	38.24 %	2.83 %
Wilcoxon Rank P-Val. RNN-Analyst	0.00 %	0.00 %	0.00 %	0.00 %
Wilcoxon Rank P-Val. RNN-Regression	0.00 %	0.00 %	0.00 %	0.00 %
Wilcoxon Rank P-Val. RNN-RW	0.00 %	0.00 %	0.00 %	0.00 %
Num. Obs.	48,192		47,863	

model with 0.01%. The other models are also more biased. To test whether the mentioned differences are statistically significant, we perform Wilcoxon signed-rank tests (Wilcoxon, 1992). Wee see that all differences are statistically significant at the 1% level.

5.2. Performance Per Quarter

Table 4 presents the overall performance results, separated per quarter. For annual predictions, we recognize that the median absolute percentage errors of the RNN, the regression model and the analysts forecasts decrease with increasing quarter. This seems intuitive since the amount of information increases during a fiscal year, leading to more precise predictions. In the same vein, it is plausible that the random walk models is not improving its performance with increasing quarters, as this model makes no use of additional information. However, we see that the analysts performance improvement is not as large as our models improvement with increasing quarters. Between quarter one and four, our model improves its performance about 19 percentage points, whereas the analysts only improve about 11 percentage points. This indicates that our model is more efficient in processing new information during

Table 4: Performance per Quarter

This table presents the median absolute percentage difference (MAPD) and the median percentage difference (MPD) between realised and predicted earnings per share, as defined in section 3.4.1, partitioned by fiscal quarter. We compare the RNN model (section 3.1) with analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3. We report the results for annual and quarterly predictions. For annual predictions, we test for all models whether their median value in quarter 4 is significantly different to the median value of quarter 1 using the Mann-Whitney U test (Mann and Whitney, 1947).

	Quarter 1		Quarter 2		Quarter 3		Quarter 4		P Val. 1-4	
	MAPD	MPD	MAPD	MPD	MAPD	MPD	MAPD	MPD	MAPD	MPD
Panel A: Annual										
RNN (%)	33.0	1.4	24.2	1.0	18.2	0.6	14.4	0.7	0.0	42.1
Analyst (%)	30.4	-15.2	27.2	-13.4	23.4	-11.7	19.6	-9.6	0.0	6.6
Regression (%)	40.6	6.6	34.4	10.6	28.1	11.9	18.6	9.1	0.0	0.0
Random Walk (%)	34.9	7.1	35.1	7.3	35.8	7.9	36.4	8.0	0.2	20.2
Num. Obs.	11,574		11,940		12,151		12,527			
Panel B: Quarterly										
RNN (%)	32.2	-3.1	27.5	-0.2	25.5	1.4	35.5	1.3		
Analyst (%)	25.5	-6.2	25.0	-6.4	23.5	-5.4	32.5	-10.2		
Regression (%)	42.4	-6.1	44.7	11.0	52.5	25.8	65.6	39.0		
Random Walk (%)	46.6	-5.4	34.8	10.6	31.4	4.5	42.9	-2.5		
Num. Obs.	11,382		12,003		12,165		12,313			

a fiscal year than the analysts. Looking at the bias of the annual forecasts, we see an improvement with increasing quarters too, from 1.4% in the first quarter to 0.7% in the fourth quarter. The analyst forecasts also show less bias in the later quarters, although the absolute level of bias remains strictly higher in all quarters compared to the RNN forecasts. The other benchmark models do not decrease their forecast bias at all. To test if there is a statistically significant improvement between quarter one and four, we conduct a Mann-Whitney U test (Mann and Whitney, 1947). For the accuracy of the predictions, we see that the differences between the first and the last quarter of all models are statistically significant at the 1% level. Regarding the bias of the predictions, there is no statistically significant difference for the RNN model, which is likely due to the already low bias in the first quarter.

Looking at the quarterly predictions, we observe for all but the regression model the general trend that quarters one and four seem to be harder to predict than quarters two and three, since the median absolute percentage errors are higher for quarter one and four. We also observe that the weaker performance on quarterly results of our model compared to analysts stems mostly from the first quarter. Although the precision of RNN predictions is slightly lower than that of the analyst forecasts in

Table 5: Annual Performance On Different Firm Sizes (Median Absolute Percentage Difference)

This table presents the median absolute percentage difference (MAPD) between realised and predicted annual earnings per share, as defined in section 3.4.1, partitioned by firm size. We define size deciles using total assets. We compare the RNN model (section 3.1) with analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3.

Size Decile	RNN	Analyst	Regression	Random Walk	Num. Obs.
1	20.34 %	20.70 %	29.00 %	36.67 %	4,820
2	22.13 %	25.36 %	30.66 %	39.66 %	4,819
3	23.16 %	28.63 %	30.12 %	39.53 %	4,819
4	23.43 %	32.21 %	32.12 %	37.84 %	4,819
5	21.33 %	28.03 %	28.45 %	36.52 %	4,819
6	22.18 %	29.06 %	30.33 %	36.47 %	4,819
7	21.22 %	25.18 %	29.03 %	33.24 %	4,819
8	21.04 %	23.72 %	28.07 %	32.53 %	4,819
9	20.70 %	24.07 %	27.43 %	31.19 %	4,819
10	19.12 %	18.77 %	26.64 %	29.99 %	4,820

quarters two, three and four, the results demonstrate a substantial gap of 6.7 percentage points during the first quarter. Looking at the median percentage difference, we see that our model is in all quarters less biased than all other benchmarks.

For the following analyses, we move the quarterly results to the appendix, since similar patterns emerge as on an annual basis.

5.3. Performance Per Firm Size

Next, we investigate the performance of all models with respect to different firm sizes, measured by total assets. To do so, we split the test set in size deciles and we compute the median absolute percentage error per decile and model. Table 5 shows the results. As a general pattern, the models tend to exhibit a decline in performance in the first three to four deciles, depending on the specific model. However, performance subsequently improves in the following deciles. Our model outperforms all benchmarks for all deciles, with the only exception being the analyst forecasts in the last decile (19.12% vs. 18.77%). It should be noted that the performance difference between the RNN model and the analyst forecasts ranges between 3 and 9 percentage points for deciles 2-6, indicating that our model surpasses the analyst forecasts particularly for small to medium-sized companies.

Table 6: Annual Performance On Different Industries (Median Absolute Percentage Difference)

This table presents the median absolute percentage difference (MAPD) between realised and predicted annual earnings per share, as defined in section 3.4.1, partitioned by industry. We use the economic sector class of The Refinitiv Business Classifications (TRBC) for industry classification. We compare the RNN model (section 3.1) with analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3.

Industry	RNN	Analyst	Regression	Random Walk	Num. Obs.
Academic & Educational Services	25.35 %	28.05 %	27.52 %	29.61 %	246
Basic Materials	26.86 %	28.00 %	34.18 %	47.44 %	2,950
Consumer Cyclicals	22.42 %	23.77 %	28.91 %	36.35 %	8,556
Consumer Non-Cyclicals	17.18 %	20.00 %	23.36 %	26.59 %	2,822
Energy	50.32 %	49.17 %	50.46 %	89.29 %	2,399
Financials	18.34 %	20.89 %	26.02 %	26.12 %	745
Healthcare	18.44 %	17.41 %	27.99 %	32.37 %	9,468
Industrials	19.24 %	20.60 %	27.25 %	30.56 %	8,832
Other	29.09 %	22.70 %	35.46 %	22.89 %	30
Real Estate	32.02 %	22.87 %	32.64 %	44.37 %	314
Technology	24.15 %	51.96 %	31.80 %	42.50 %	9,850
Utilities	11.99 %	7.30 %	20.75 %	14.80 %	1,980

5.4. Performance Per Industry

Table 6 contains the MAPD results for different industries. For industry classification, we use the economic sector class of The Refinitiv Business Classifications (TRBC). If there is no class label for a given observation, we assign it to the "Other" class. The results of the previous analyses confirm in a way that the RNN model outperforms all benchmarks for seven out of twelve industries. The five exceptions are the energy, healthcare, real estate and the utilities sector as well as the others category. For energy, healthcare and utilities, the differences are all smaller than 5 percentage points. For the real estate and the others class, the difference is higher, but these classes are rather small. It is noticeable that the median absolute percentage errors in the energy sector are clearly higher compared to all other industries. This effect holds for all models.

5.5. Sign Prediction

Table 7 shows the results of the sign prediction analysis, as explained in section 3.4.2. We see that our model outperforms all other benchmarks with respect to accuracy, precision, recall and F1. Our model outperforms the analysts by 1-2 percentage

Table 7: Sign prediction (Annual)

This table presents the results for the sign prediction analysis for annual data, as defined in section 3.4.2, for the RNN model (section 3.1) and all benchmarks. This analysis examines the models ability to predict the direction of an earnings change from the previous to the actual period. We define three classes: nositive (percentage change > 5%), negative (percentage change < -5%) and neutral (abs(percentage change) < 5%). The benchmarks are analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3. We calculate macro accuracy, precision, recall and F1 for all models, as defined in Sokolova and Lapalme (2009).

	RNN	Analyst	Random Walk	Regression
Accuracy	78.90 %	77.88 %	38.97 %	73.37 %
Precision	56.78 %	55.81 %	8.45 %	49.06 %
Recall	57.80 %	55.84 %	33.33 %	48.03 %
F1	56.79 %	55.60 %	15.58 %	46.98 %
Num. Obs.		40,739		

points for all measures. The difference becomes even clearer compared to the other two benchmarks.

5.6. Performance Covered vs. Uncovered Firms

We have already seen that our model performs better than analyst forecasts if we consider annual earnings forecasts. However, all of the previous results are evaluated on a sample for which observations for every model are available. A natural question is whether our model performs equally well for observations without analyst coverage. As a first approach, we investigate the difference in performance of our model and the remaining benchmarks in both cases, analyst coverage and no analyst coverage. Table 18 in the appendix shows the results. For annual predictions, we see that the MAPD of our model increases about 9 percentage points if we consider data that is not covered by analysts. A similar decrease in performance can be observed for the benchmarks, too. Regarding the MPD, we observe the opposite pattern. For all models, the predictions get less biased if the data is not covered by analysts. All of these effects are statistically significant on the 5% level.

A drawback of the approach of Table 18 is that both samples, i.e. the covered and uncovered firms, are structurally different. As Li and Mohanram (2014) argue in their study and which is in line with our findings, analyst forecasts are available only for a subset of firms, with almost half of all firms not having analyst coverage in most years. Most of the firms without analyst coverage are typically small and young firms. Therefore, in a second step we try to reduce the structural difference between covered and uncovered firms, so that we end up with a clean comparison

Table 8: Performance Analyst Coverage vs. Similar No Analyst Coverage (Annual)

This table presents the median absolute percentage difference (MAPD), as defined in section 3.4.1, of the annual RNN model EPS predictions (see section 3.1) and realised EPS values for not analyst-covered firms and similar analyst-covered firms. Two firms are similar if they are in the same industry and if the eucledian distance between their arrays consisting of standarized values of total assets and tobins q is smaller than 0.01. The sample consists of all non-covered firm observations for which there is at least one similar covered firm observation. If there are more than one similar firm observations, we take the most similar one. We test whether the median value for non-covered firms is significantly different to the median value for similar covered firms using the Wilcoxon rank test (Wilcoxon, 1992).

	Count	MAPD Covered	MAPD Uncovered	P Value
Academic & Educational Services	68	27.5 %	21.7 %	40.3 %
Basic Materials	704	29.9 %	26.6 %	6.2 %
Consumer Cyclicals	3,846	28.0 %	37.1 %	0.0 %
Consumer Non-Cyclicals	471	30.2 %	19.6 %	0.2 %
Energy	743	55.4 %	63.1 %	0.8 %
Financials	114	41.6 %	39.7 %	54.0 %
Healthcare	6,369	17.6 %	20.8 %	0.0 %
Industrials	3,939	23.0 %	27.0 %	35.4 %
Other	4	89.1 %	158.1 %	37.5 %
Real Estate	37	50.1 %	24.7 %	0.1 %
Technology	4,401	29.5 %	33.5 %	0.0 %
Utilities	125	11.6 %	9.8 %	54.7 %
Total	20,821	23.8 %	28.0 %	0.0 %

of the performance of our model in both situations. The idea is to only compare covered with uncovered firms that are similar with respect to industry, total assets and Tobin's q. To achieve this, both companies must belong to the same industry. In addition, the Euclidean distance between the vectors consisting of the standardized values of total assets and Tobin's q must be smaller than a certain limit, which we set to 0.01. Within all the uncovered companies that fulfill that property, we take the one with the minimal distance. If there is no uncovered company that fulfills this property, the covered observation will be removed from the analysis.

Table 8 shows the results of that analysis. To keep things well arranged, we only report the MAPD results for our RNN model. We see that for annual predictions the increase in the median absolute percentage difference is now only 4.2 percentage points instead of 10 percentage points as in Table 18. Even if the differences between covered and uncovered firms are still statistically significant at the 1% level, we see that considering similar firms increases the models performance on uncovered data strongly. Looking at the specific industries, we see that only for four out of twelve industries, the RNN performs statistically significantly worse (at the 10% level) if a firm is not covered by analysts. For all other industries, the difference is either insignificant, or, in case of the industries *Basic Materials*, *Consumer Non-Cyclicals* and *Real Estate* the predictive performance of the RNN is even statistically significantly better for uncovered firms. This indicates that our model yields robust results even for firms not covered by analysts.

5.7. Performance Pre vs. During Covid

Another question we would like to answer is how robust our model predictions are during a crisis. The only crisis that occurred during our test data period is the Covid pandemic. Even though other events as the dot-com bubble in 2001, the subprime crisis in 2007/2008 or the European debt crisis in 2009/2010 are contained in our data, the fact that these events are already seen by our model during training would distort the comparison to the benchmark models. We divide the test set in a time prior the Covid pandemic and in a time during the Covid pandemic. We do not include a period that covers the time after the Covid pandemic, since it is not clear whether the pandemic was already over until the end of 2021. We set the 18th of February 2020 as the starting point for the pandemic.

Looking at Table 9, we observe a performance drop with respect to MAPD of 3 percentage points for the annual predictions of our model when we focus on ob-

Table 9: Performance Pre VS. During COVID 19 (Annual)

This table compares the median absolute percentage difference (MAPD) and the median percentage difference (MPD), as defined in section 3.4.1, for annual earnings reports before and during the Covid pandemic. We compute these measures for the RNN model (section 3.1), analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3. We set February 18, 2020 as the starting point for the pandemic. For all models, we test whether their median value for observations prior to the pandemic is significantly different to their median value for observations during the pandemic using the Mann-Whitney U test (Mann and Whitney, 1947).

	Pre-Covid		During-Covid		P Value	
	MAPD	MPD	MAPD	MPD	MAPD	MPD
RNN	20.44 %	0.26 %	23.88 %	2.67 %	0.00 %	0.00 %
Analyst	24.59 %	-13.21 %	26.82 %	-9.54 %	0.06 %	0.00 %
Regression	29.38 %	10.53 %	28.35 %	7.40 %	1.08 %	0.06 %
Random Walk	32.48 %	6.89 %	42.94 %	10.07 %	0.00 %	0.00 %
Num. Obs.	35,918		12,274			

servations during the pandemic, indicating that the RNN predictions are robust in times of a crises. The drop in performance is comparable to the analysts (more than 2 percentage points) and better than the random walk model (10 percentage points). However, the regression model even performs 1 percentage point better during the crisis. Furthermore, our model is slightly more biased during the Covid crisis (2.67%) than before (0.26%), but the absolute level is still small. However, the MPD of the analysts and the regression model are even decreasing about 4 and 3 percentage points, respectively. We test the performance differences of the models before and during Covid for statistical significance. All differences are statistically significant at the 5% level.

5.8. Earnings Response Coefficient

Our last analysis covers the earnings response coefficient, following the approach from section 3.4.3. In contrast to the previous analyses, we only consider the most recent prediction prior to the report date of an earnings report. Otherwise, it would not be clear why to expect an announcement effect at the end of a fiscal year if the proxy of market expectation was measured after, for example, quarter one. This step roughly quarters the sample for annual predictions. Another necessary filtering step is the requirement of the existence of the exact report date of an earnings result, as we try to link earnings surprises to stock market effects. Therefore, we remove all observations where the respective Compustat variable (rdq) is not available.

Table 10: Earnings Response Coefficient (Annual)

This table presents the annual results for the earnings response regression of section 5.8. We fit a year- and firm-fixed effects model with clustered standard errors. Earnings surprise is defined as $\frac{EPS - \widehat{EPS}}{\widehat{EPS}}$ with EPS being the realized earnings per share and $\widehat{EPS}$ is its prediction.

	RNN	Analyst	Regression	Random Walk
surprise (%)	4.89***	1.83**	4.71***	0.65*
lnTA (%)	-1.93***	-1.93***	-1.91***	-1.87***
lnAssets x surprise (%)	-0.42***	-0.15*	-0.40***	-0.05
tobinsQ (%)	-0.53***	-0.55***	-0.53***	-0.50***
tobinsQ x surprise (%)	-0.05	-0.10**	-0.09	-0.05
2015 (%)	0.78**	0.79**	0.81**	0.75**
2016 (%)	1.32***	1.32***	1.36***	1.25***
2017 (%)	0.63*	0.67**	0.69**	0.59*
2018 (%)	1.17***	1.22***	1.25***	1.14***
2019 (%)	0.83*	0.92*	1.01**	0.83*
2020 (%)	-0.52	-0.50	-0.40	-0.60
2021 (%)	1.04	0.99	0.94	0.89
2022 (%)	2.47***	2.55***	2.70***	2.44***
R^2 (%)	2.30	1.65	2.34	1.54
Num. Obs.	12,446	12,446	12,446	12,446

Table 10 covers the results of the earnings response analysis. For annual predictions, we see that the earnings response coefficient (ERC), i.e. the coefficient of the earnings surprise variable, is with 4.89% more than 3 percentage points higher than the ERC of the analysts, almost 0.2 percentage points higher than the ERC of the regression model and more than 4 percentage points higher than the random walk model. All coefficients are statistically significant. The R^2 of our model amounts to 2.3%, compared to 1.65% for the analysts, 2.34% for the regression model and 1.54% for the random walk. Total assets and the interaction effect of total assets with the surprise variable have both a significant negative effect on the abnormal return for all models (except random walk), which implies that increasing firm size decreases the impact of surprising earnings announcements on the abnormal return. Tobin's q has also a significant negative effect on the abnormal return for all models. However, the interaction effect is not significant (except for the analysts). The results for the year dummies are mixed: We get significant positive effects for the years 2015, 2016, 2017, 2018, 2019 and 2022 for all models.

6. Conclusion

This study addresses the limitations associated with financial analysts' representation of public knowledge regarding firms' earnings prospects. Previous research has revealed the existence of biases in analysts' estimates due to conflicts of interest, resulting in limited predictive power in explaining stock price reactions to earnings disclosures. Furthermore, analyst coverage remains constrained, both inside and particularly outside the United States. Several subsequent research studies have endeavored to tackle the aforementioned issues, but they themselves exhibit certain weaknesses. These studies employ models that either generate less accurate earnings forecasts compared to financial analysts (Hou et al., 2012; Li and Mohanram, 2014; Hess and Lorsbach, 2019), incorporate analyst forecasts in their estimations without resolving the problem of uncovered firms (Azevedo et al., 2021; Ball and Ghysels, 2018), or rely heavily on data-intensive approaches, thereby introducing selection biases (Bradshaw et al., 2012; Elend et al., 2020).

To overcome these challenges, this chapter introduces a novel approach by developing a recurrent neural network model for forecasting future firm earnings. Using four decades of quarterly financial information allows the model to learn time dependent patterns in earnings results, which reduces the informational advantage of analysts. Additionally, we avoid the data intensity of previous time-series models where survivorship biases are commonly observed. Our study goes beyond the traditional focus on fiscal-year-end predictions and extends its forecast capabilities to include quarterly earnings, as these are recognized as a crucial source of information that significantly influences stock prices.

The results of this study for annual earnings results indicate that our model surpasses the considered benchmarks, including analyst forecasts, the best cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model, in terms of both accuracy and bias. Furthermore, our results demonstrate that the strong predictive power of our model extends to uncovered firms, effectively addressing the issue of low analyst coverage for small or young firms and those operating outside the United States. This outcome signifies the broader applicability and reliability of our model in capturing and forecasting earnings for a diverse range of firms. In addition, our model exhibits robustness in the face of crises, as indicated by our analysis of the change in predictive power during the Covid-19 crisis. This resilience further underscores the effectiveness and stability of our model in providing accurate earnings predictions even during periods of economic uncertainty. Notably,

when compared to the benchmarks, our model stands out by producing the highest earnings response coefficient. This result signifies that our model's earnings predictions align more closely with market expectations, highlighting its ability to provide precise and reliable insights into firms' future earnings prospects.

Regarding quarterly earnings predictions, our RNN model exhibits lower bias compared to all the benchmarks and more accurate predictions than the regression and the random walk model, although its forecasts are slightly less precise than analyst predictions. Nevertheless, the overall findings remain consistent with the outcomes observed in annual predictions. Given the limited analyst coverage, particularly for quarterly earnings data, the RNN model retains its value as a valuable tool for forecasting uncovered firms. This is especially important considering that the earnings response coefficient of the RNN model is comparable to that of the analysts. Thus, our RNN model serves as a reliable and effective means of generating forecasts for firms that lack analyst coverage, ensuring comprehensive insights for investors and decision-makers.

References

Abarbanell, Jeffery S, 1991, Do analysts' earnings forecasts incorporate information in prior stock price changes?, *Journal of Accounting and Economics* 14, 147–165.

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, 545–579.

Ball, Ryan T, and Eric Ghysels, 2018, Automated earnings forecasts: Beat analysts or combine and conquer?, *Management Science* 64, 4936–4952.

Bradshaw, Mark T, Michael S Drake, James N Myers, and Linda A Myers, 2012, A re-examination of analysts' superiority over time-series forecasts of annual earnings, *Review of Accounting Studies* 17, 944–968.

Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, 2014, Learning phrase representations using rnn encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078.

Elend, Lars, Sebastian A Tideman, Kerstin Lopatta, and Oliver Kramer, 2020, Earnings prediction with deep leaning, in *German Conference on Artificial Intelligence (Künstliche Intelligenz)*, 267–274, Springer.

Fama, Eugene F, 1970, Efficient capital markets: A review of theory and empirical work, *The journal of Finance* 25, 383–417.

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

Hess, Dieter, and Tobias Lorsbach, 2019, Incorporating quarterly earnings information into cross-sectional earnings forecast models, SSRN Electronic Journal.

Hochreiter, Sepp, and Jürgen Schmidhuber, 1997, Long short-term memory, *Neural computation* 9, 1735–1780.

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.

Kingma, Diederik P, and Jimmy Ba, 2014, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

Kothari, SP, 2001, Capital markets research in accounting, *Journal of accounting and economics* 31, 105–231.

Lewellen, Jonathan, and Robert J Resutek, 2019, Why do accruals predict earnings?, *Journal of Accounting and Economics* 67, 336–356.

Li, Kevin K, and Partha Mohanram, 2014, Evaluating cross-sectional forecasting models for implied cost of capital, *Review of Accounting Studies* 19, 1152–1185.

Mann, Henry B, and Donald R Whitney, 1947, On a test of whether one of two random variables is stochastically larger than the other, *The annals of mathematical statistics* 50–60.

O'brien, Patricia C, 1988, Analysts' forecasts as earnings expectations, *Journal of accounting and Economics* 10, 53–83.

Sokolova, Marina, and Guy Lapalme, 2009, A systematic analysis of performance measures for classification tasks, *Information processing & management* 45, 427–437.

Wilcoxon, Frank, 1992, Individual comparisons by ranking methods, in *Breakthroughs in statistics*, 196–202 (Springer).

Table 11: Definition Compustat Variables

This table presents the definitions of the used Compustat variables.

Abbreviation	Name	Compustat Item Nr.	Unit of Measurement
che	Cash & Short-Term Investments	1	Millions
act	Current Assets	4	Millions
lct	Current Liabilities	5	Millions
ta	Total Assets	6	Millions
dp	Depreciation & Amortization Expenses	14	Millions
spi	Special Items	17	Millions
ib	Income Before Extraordinary Items	18	Millions
csho	Common Shares Outstanding	25	Millions
dlc	Debt in Current Liabilities	34	Millions
mib	Minority Interest	38	Millions
txp	Income Taxes Payable	71	Millions
ibc	Income Before Extraordinary Items	123	Millions
dv	Total Cash Dividends	127	Millions
lt	Total Liabilities	181	Millions
seq	Stockholder's Equity	216	Millions
oancf	Cash Flow From Operations	308	Millions

Appendix

Variable definitions

This section defines the used accounting and market variables. In a first step, we give the definitions for the variables from the used databases. Table 11 yields definitions for Compustat variables and Table 12 provides the definitions for CRSP variables.

Tables 13 and 14 outline the accounting and market variables employed in this study. Furthermore, we delineate the lower and upper winsorization thresholds for these variables to facilitate their straightforward application in future research endeavors. Additionally, as we transition the aggregation of market variables from daily to quarterly frequency, we specify the aggregation method for each market variable.

---

[8]   No winsorization, but data removal

Table 12: Definition CRSP Variables

This table presents the definitions of the used CRSP variables.

Abbreviation	Name	Unit of Measurement
RET	Holding Period Total Return	-
SHROUT	Shares Outstanding	Thousands
CFACSHR	Split-adjustment factor (shares)	-
CFACPR	Split-adjustment factor (prices)	-
VOL	Volume Traded	-
PRC	Price or Bid/Ask Average	-

Table 13: Definition Accounting Variables

This table presents the definitions of the used accounting variables in this study.

Name	Definition	Winsorization Thresholds
Earnings per share	$\frac{\text{ib}-\text{dvp}}{\text{csho}}$	$[-30, 30]$
Equity ratio	$\frac{\text{bv}}{\text{ta}}$	$[0, 1]^8$
Accruals per share (*Until 1988*)	$\frac{\triangle\text{act}-\triangle\text{che}-\triangle\text{lct}-\triangle\text{dlc}-\triangle\text{txp}-\text{dp}}{\text{csho}}$	$[-20, 6]$
Accruals per share (*Since 1988*)	$\frac{\text{ib}-\text{oancf}}{\text{csho}}$	$[-20, 6]$
Dividend dummy	$\text{dv} > 0$	-
Total assets per share	$\ln\left(\frac{\text{at}}{\text{csho}}\right)$	$[0.28, 530]$
Time between predicted and actual observation	$\frac{t_{\text{prediction}}-t_{\text{observation}}}{365.25}$	-

Table 14: Definition Market Variables

This table presents the definitions of the used market variables in this study. A period is 63 trading days (approximately one quarter). If less than 63 trading days available, we take the maximum of available days. Winsorization is applied before period aggregation.

Name	Definition	Period Aggregation	Winsorization Thresholds
Market value of equity (MV)	$\frac{\text{PRC}}{\text{CFACPR}} \times \text{SHROUT}$	-	-
Stock return (log)	RET	Average	$[-1, 1]$
Volume per Share	$\frac{\text{VOL}}{\text{SHROUT}}$	Average	> 0.1
Market Index Return (log) (S&P 500)	marketIndex	Average	-
Volatility Index (VIX)	volatilityIndex	$\frac{\text{Absolute Change}}{\text{Period Length}}$	-
Tobin's q	$\frac{MV + lct}{seq + lct}$	$\frac{\text{Absolute Change}}{\text{Period Length}}$	< 15
Length of period (after compression)	$\frac{t_{\text{period end}} - t_{\text{period start}}}{365.25}$	-	-

Performance Per Firm Size On Quarterly Earnings

The analysis of the model performance with respect to quarterly earnings results reveals similar patterns as the analysis of annual data from section 5.3, as we see in Table 15. The general pattern of a declining performance in the first three to four deciles and improving performance in the following deciles is still present for all models. For all but the third decile, analyst forecasts outperform the RNN. However, similar to the annual case, the RNN model performs best compared to the analyst forecasts for the small to medium sized deciles two, three, four and five, where the gap in performance is between 0 and 3 percentage points. For the remaining size deciles, the disparity becomes increasingly pronounced, for decile one as well as the higher deciles. Notably, the largest difference is observed in the 10th decile, nearing a margin of approximately 6 percentage points. The RNN outperforms the other benchmarks for every decile.

Performance Per Industry On Quarterly Earnings

Table 16 presents the performance evaluation for quarterly earnings predictions of the RNN model and all benchmarks segmented by industry. We see that our model outperforms the random walk and the regression model for almost every industry. However, comparing to quarterly analyst forecasts, the results indicate that the

Table 15: Quarterly Performance On Different Firm Sizes (Median Absolute Percentage Difference)

This table presents the median absolute percentage difference (MAPD) between realised and predicted quarterly earnings per share, as defined in section 3.4.1, partitioned by firm size. We define size deciles using total assets. We compare the RNN model (section 3.1) with analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3.

Size Decile	RNN	Analyst	Regression	Random Walk	Num. Obs.
1	29.48 %	25.69 %	47.11 %	31.62 %	4,787
2	30.81 %	28.29 %	51.51 %	34.89 %	4,786
3	31.95 %	32.01 %	53.75 %	40.50 %	4,786
4	33.68 %	32.34 %	56.68 %	42.68 %	4,786
5	31.86 %	28.58 %	54.35 %	40.44 %	4,786
6	31.59 %	27.94 %	55.18 %	43.44 %	4,787
7	29.47 %	24.72 %	53.59 %	39.66 %	4,786
8	28.98 %	23.85 %	52.92 %	40.41 %	4,786
9	27.89 %	23.31 %	52.39 %	38.62 %	4,786
10	24.68 %	18.75 %	50.36 %	33.97 %	4,787

Table 16: Quarterly Performance On Different Industries (Median Absolute Percentage Difference)

This table presents the median absolute percentage difference (MAPD) between realised and predicted quarterly earnings per share, as defined in section 3.4.1, partitioned by industry. We use the economic sector class of The Refinitiv Business Classifications (TRBC) for industry classification. We compare the RNN model (section 3.1) with analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3.

Industry	RNN	Analyst	Regression	Random Walk	Num. Obs.
Academic & Educational Services	30.05 %	22.04 %	49.55 %	49.75 %	228
Basic Materials	38.39 %	25.54 %	57.59 %	46.94 %	2,914
Consumer Cyclicals	32.10 %	24.80 %	54.87 %	50.01 %	8,467
Consumer Non-Cyclicals	24.86 %	20.73 %	46.82 %	35.49 %	2,796
Energy	68.44 %	49.67 %	83.11 %	68.64 %	2,394
Financials	25.41 %	18.37 %	54.06 %	34.33 %	756
Healthcare	24.91 %	21.15 %	48.30 %	25.31 %	9,524
Industrials	26.99 %	21.13 %	50.96 %	34.82 %	8,784
Other	21.03 %	24.06 %	44.54 %	33.16 %	29
Real Estate	39.79 %	24.50 %	57.42 %	47.64 %	288
Technology	32.52 %	47.30 %	54.48 %	36.78 %	9,742
Utilities	21.94 %	14.21 %	52.00 %	53.36 %	1,941

Table 17: Sign prediction (Quarterly)

This table presents the results for the sign prediction analysis for quarterly data, as defined in section 3.4.2, for the RNN model (section 3.1) and all benchmarks. This analysis examines the models ability to predict the direction of an earnings change from the previous to the actual period. We define three classes: nositive (percentage change > 5%), negative (percentage change < -5%) and neutral (abs(percentage change) < 5%). The benchmarks are analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3. We calculate macro accuracy, precision, recall and F1 for all models, as defined in Sokolova and Lapalme (2009).

	RNN	Analyst	Random Walk	Regression
Accuracy	76.99 %	78.99 %	76.79 %	68.48 %
Precision	52.99 %	56.46 %	51.83 %	42.28 %
Recall	53.50 %	57.38 %	51.42 %	40.50 %
F1	52.94 %	56.79 %	51.35 %	38.75 %
Num. Obs.		40,478		

analyst forecasts outperform the RNN model for all industries, with differences varying between 19 percentage points for the energy sector and 3.8 percentage points in the healthcare sector. The only exception are the technology and the others sectors, where our model performs more than 14 and 3 percentage points better than the analysts, respectively. The technology sector is the largest industry in the sample with 9,742 observations.

Sign Prediction On Quarterly Earnings

Table 17 shows the results for the sign prediction for quarterly earning results of the RNN model and all benchmarks. We see that the RNN model is outperformed by the analyst forecasts with respect to accuracy, precision recall and F1. The RNN model performs comparable at slightly better as the random walk model. Only the regression model performs worse than the other models.

Performance Covered vs. Uncovered Firms

Table 18 investigates the difference in performance of our model and the remaining benchmarks in both cases, analyst coverage and no analyst coverage. For annual predictions, we see that the MAPD of our model increases about 9 percentage points if we consider data that is not covered by analysts. A similar decrease in performance can be observed for the benchmarks, too. Regarding the MPD, we observe the opposite pattern. For all but our RNN model, the predictions get less biased if the

Table 18: Performance Analyst Coverage vs. No Analyst Coverage

This table compares the median absolute percentage difference (MAPD) and the median percentage difference (MPD), as defined in section 3.4.1, for analyst-covered and not analyst-covered firms. We compute these measures for the RNN model (section 3.1), analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3. We report the results for annual (Panel A) and quarterly (Panel B) predictions. For all models, we test whether their median value for covered firms is significantly different to their median value for uncovered firms using the Mann-Whitney U test (Mann and Whitney, 1947).

| | Covered | | Uncovered | | P Value | |
	MAPD	MPD	MAPD	MPD	MAPD	MPD
Panel A: Annual						
RNN	21.88 %	0.84 %	31.25 %	-2.08 %	0.00 %	0.00 %
Regression	29.11 %	9.76 %	37.79 %	5.56 %	0.00 %	0.00 %
Random Walk	35.73 %	7.52 %	48.36 %	2.93 %	0.00 %	0.00 %
Num. Obs.	51,274		8,455			
Panel B: Quarterly						
RNN	30.79 %	-0.33 %	45.16 %	-2.59 %	0.00 %	1.86 %
Regression	52.72 %	14.67 %	61.68 %	10.72 %	0.00 %	8.59 %
Random Walk	38.54 %	2.42 %	53.88 %	-0.00 %	0.00 %	0.39 %
Num. Obs.	50,949		8,416			

data is not covered by analysts. All of these effects are statistically significant on the15% level.

Looking at the quarterly predictions, we see the same pattern. The MAPD of all models increases between 8 and 15 percentage points if the data is not covered. For all but the RNN model, the bias decreases for uncovered data. All these effects are statistically significant at the 10% level.

Table 19 presents the results of the comparison between the RNN's predictive performance for covered and similar uncovered firms, based on quarterly earnings results. The exact procedure is described in section 5.6. We see that the RNN's MAPD of 38% for uncovered is comparable to the 34% for covered firms. Especially if we compare this to the naive comparison of covered and uncovered firms, without considering firm similarities. In this case, the difference in the models predictive performance on covered and uncovered firms is more than 14 percentage points. Regarding the results on industry level, we see that the model achieves compatible results for similar uncovered firms from most of the considered industries.

Table 19: Performance Analyst Coverage vs. Similar No Analyst Coverage (Quarterly)

This table presents the median absolute percentage difference (MAPD), as defined in section 3.4.1, of the quarterly RNN model EPS predictions (see section 3.1) and realised EPS values for not analyst-covered firms and similar analyst-covered firms. Two firms are similar if they are in the same industry and if the euclidian distance between their arrays consisting of standarized values of total assets and tobins q is smaller than 0.01. The sample consists of all non-covered firm observations for which there is at least one similar covered firm observation. If there are more than one similar firm observations, we take the most similar one. We test whether the median value for non-covered firms is significantly different to the median value for similar covered firms using the Wilcoxon rank test (Wilcoxon, 1992).

	Count	MAPD Covered	MAPD Uncovered	P Value
Academic & Educational Services	56	35.3 %	63.2 %	1.4 %
Basic Materials	704	45.9 %	39.5 %	62.6 %
Consumer Cyclicals	3,833	39.7 %	51.3 %	0.0 %
Consumer Non-Cyclicals	470	52.6 %	35.0 %	0.0 %
Energy	746	73.5 %	84.0 %	1.1 %
Financials	111	48.1 %	54.9 %	37.6 %
Healthcare	6,425	24.3 %	31.5 %	0.0 %
Industrials	3,884	34.4 %	29.5 %	0.6 %
Other	3	71.5 %	57.1 %	75.0 %
Real Estate	31	52.4 %	53.7 %	88.5 %
Technology	4,321	41.0 %	46.1 %	0.0 %
Utilities	125	23.3 %	21.3 %	59.0 %
Total	20,709	34.0 %	38.0 %	0.0 %

Table 20: Performance Pre VS. During COVID 19 (Quarterly)

This table compares the median absolute percentage difference (MAPD) and the median percentage difference (MPD), as defined in section 3.4.1, for quarterly earnings reports before and during the Covid pandemic. We compute these measures for the RNN model (section 3.1), analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3. We set February 18, 2020 as the starting point for the pandemic. For all models, we test whether their median value for observations prior to the pandemic is significantly different to their median value for observations during the pandemic using the Mann-Whitney U test (Mann and Whitney, 1947).

	Pre-Covid		During-Covid		P Value	
	MAPD	MPD	MAPD	MPD	MAPD	MPD
RNN	27.98 %	-0.35 %	36.12 %	1.14 %	0.00 %	0.00 %
Analyst	24.79 %	-7.44 %	30.51 %	-5.21 %	0.00 %	0.00 %
Regression	52.02 %	16.39 %	54.64 %	8.54 %	0.00 %	0.00 %
Random Walk	37.85 %	3.04 %	39.39 %	2.14 %	1.27 %	45.52 %
Num. Obs.	35,494		12,369			

Table 21: Earnings Response Coefficient (Quarterly)

This table presents the quarterly results for the earnings response regression of section 5.8. We fit a year- and firm-fixed effects model with clustered standard errors. Earnings surprise is defined as $\frac{EPS - \widehat{EPS}}{\widehat{EPS}}$ with EPS being the realized earnings per share and $\widehat{EPS}$ is its prediction.

	RNN	Analyst	Regression	Random Walk
surprise (%)	5.81***	5.52***	2.06***	2.51***
lnTA (%)	-2.46***	-2.46***	-2.65***	-2.56***
lnAssets x surprise (%)	-0.48***	-0.46***	-0.12***	-0.21***
tobinsQ (%)	-0.68***	-0.68***	-0.69***	-0.65***
tobinsQ x surprise (%)	-0.11***	-0.11***	-0.09***	-0.01
Q1 (%)	-0.11	-0.41	0.12	-0.19
Q2 (%)	0.09	-0.07	0.29	-0.17
Q3 (%)	0.12	0.00	0.26	0.01
2015 (%)	0.33***	0.29***	0.32***	0.27***
2016 (%)	0.66***	0.66***	0.65***	0.55***
2017 (%)	0.17	0.15	0.24*	0.12
2018 (%)	1.30***	1.32***	1.46***	1.34***
2019 (%)	1.10***	1.11***	1.25***	1.04***
2020 (%)	1.49***	1.30***	1.61***	1.27***
2021 (%)	1.45***	1.40***	1.69***	1.58***
2022 (%)	2.55***	2.43***	3.00***	2.64***
R^2 (%)	4.19	3.64	2.63	2.70
Num. Obs.	47,776	47,776	47,776	47,776

Performance Pre vs. During Covid On Quarterly Earnings

Table 20 shows the results for the comparison of the predictive performance of RNN and benchmarks before and during the Covid crisis for quarterly earnings results. In contrast to results for annual earnings in section 5.7, the RNN model reacts more sensitive to the Covid crisis, as the MAPD increases from 27.98% before Covid to 36.12% during Covid. The MAPD of the analyst forecasts only increases about 6 percentage points. The other benchmarks are even more robust to the Covid crisis. The same holds for the bias of the RNN model. The MPD is increasing during Covid for the RNN model, while it is decreasing for all benchmarks. These differences are all statistically significant at the 1% level, except for the random walk model.

Earnings Response Coefficient On Quarterly Earnings

Table 21 contains the results of the earnings response analysis, as described in section 5.8, for quarterly earnings. We see that the results of the RNN model and the analyst forecasts are quite similar. The analysts ERC is slightly dmaller compared

Table 22: Performance Per Year

This table presents the median absolute percentage difference (MAPD) and the median percentage difference (MPD) between realised and predicted earnings per share, as defined in section 3.4.1, partitioned by fiscal year. We compare the RNN model (section 3.1) with analysts forecasts, the cross-sectional regression model of Hess and Lorsbach (2019) and the random walk model (Bradshaw et al., 2012), see section 3.3. We report the results for annual (Panel A) and quarterly (Panel B) predictions.

	RNN		Analyst		Regression		Random Walk		Num.
Year	MAPD	MPD	MAPD	MPD	MAPD	MPD	MAPD	MPD	Obs.
				Panel A: Annual					
2014	17.2 %	0.9 %	21.8 %	-12.2 %	25.6 %	13.8 %	27.6 %	7.9 %	5,775
2015	18.2 %	-2.3 %	24.0 %	-16.1 %	27.3 %	8.9 %	29.2 %	3.6 %	6,069
2016	17.6 %	-0.7 %	21.3 %	-14.2 %	28.4 %	11.5 %	29.0 %	6.6 %	5,838
2017	28.0 %	3.6 %	31.4 %	-7.3 %	35.6 %	13.1 %	39.3 %	12.7 %	5,743
2018	22.8 %	3.4 %	23.9 %	-11.0 %	32.1 %	9.8 %	39.4 %	12.6 %	5,984
2019	19.9 %	-1.3 %	25.1 %	-17.1 %	26.8 %	3.8 %	30.7 %	0.5 %	6,006
2020	26.0 %	-0.7 %	31.6 %	-13.7 %	29.3 %	6.8 %	41.2 %	-4.6 %	5,933
2021	23.7 %	5.9 %	24.6 %	-6.6 %	29.3 %	9.0 %	45.5 %	25.2 %	6,503
				Panel B: Quarterly					
2014	25.7 %	0.3 %	23.3 %	-7.2 %	49.7 %	22.6 %	35.0 %	4.2 %	5,713
2015	25.9 %	-3.5 %	23.9 %	-9.1 %	51.0 %	14.5 %	35.7 %	0.2 %	5,955
2016	25.7 %	-1.6 %	23.2 %	-7.9 %	51.5 %	18.1 %	35.4 %	3.1 %	5,794
2017	32.0 %	1.0 %	28.4 %	-5.3 %	56.0 %	17.4 %	42.5 %	5.8 %	5,765
2018	31.0 %	5.3 %	25.7 %	-5.5 %	55.4 %	17.7 %	43.6 %	3.3 %	5,820
2019	28.8 %	-3.1 %	25.9 %	-10.6 %	50.6 %	5.1 %	37.5 %	0.4 %	5,935
2020	40.0 %	-2.1 %	33.1 %	-6.2 %	57.9 %	0.5 %	45.4 %	3.0 %	5,900
2021	33.0 %	5.2 %	27.4 %	-3.2 %	50.5 %	15.5 %	33.7 %	3.0 %	6,639

to our model (5.81% and 5.52%, respectively). Also the R^2 of the RNN is with 4.19% higher than the analysts' R^2 of 3.64%. The other benchmarks perform worse, around 3 percentage points worse with respect to ERC and more than 1 percentage point worse in terms of R^2. The effects for total assets, Tobin's q and the respective interaction terms are similar to the annual case, with the only exception that the interaction of Tobin's q is statistically significant for all but the random walk model. The quarter dummies are insignificant for all models. The dummies of every year except 2017 are significant and positive for all models.

Performance Per Year

In Table 22 we investigate the predictive performance of the RNN model and the benchmarks for each year in the test set separately. For annual predictions, we see

that our model outperforms the analyst forecasts in every year with margins between 1 and 7 percentage points. The same holds, with even higher margins, for the other benchmarks. There is a general trend of a slightly decreasing performance with increasing years for the estimated models, i.e. the RNN and the regression model. This observation signifies that the model acquires temporal patterns during the training phase, which progressively diminish or undergo alterations in the test set as the temporal distance between the observation and the training dataset increases. Additionally, we see that for all years, the RNN model is less biased than all benchmarks.

Repeating the same analysis for quarterly earnings results, we deduce that the RNN performs worse than the analyst forecasts in all years with respect to MAPD. However, it outperforms all other benchmarks. The trend of a slightly decreasing performance with respect to MAPD exists for quarterly earnings predictions too.