Rotman School of Management
UNIVERSITY OF TORONTO

Rotman

Group Number: 10

Assignment Title: Earnings Prediction

Course Code: RSM8244

Course Title: Analytical Insights Using Accounting and Financial Data

Instructor Name: Scott Liao

In submitting this **group** work for grading, we confirm:

- That the work is original, and due credit is given to others where appropriate.
- That all members have contributed substantially and proportionally to each group assignment.
- That all members have sufficient familiarity with the entire contents of the group assignment so as to be able to sign off on them as original work.
- Acceptance and acknowledgement that assignments found to be plagiarized in any way will be subject to sanctions under the University's Code of Behaviour on Academic Matters.

Please **check the box and record your student number** below to indicate that you have read and abide by the statements above:

☐	1002140541	☐	1005944677
☐	1006118372	☐	1004654527
☐	1000675780	☐	

If submitting this assignment online please use the following "Standard File Naming Convention":

Full Course Code *(including Section)*–Group Name or Number–Assignment Title

Example: RSM1234HS.2016-0101-Group1-Homework1

Introduction

Earnings prediction is simply the act to determine future value of a stock, or any financial instrument, of a company. Accurate and successful prediction of a stock's future prices can yield significant profits. According to the efficient-market hypothesis, stock prices are a function of information and rational expectations, where any newly revealed information about a company is almost immediately reflected on current stock price. Burton Malkiel (1973), claims that therefore stock prices cannot be inherently predicted by looking at price history, and that stock prices follow a "random walk", where deviations are random and unpredictable. However, there are pitfalls to this hypothesis. Some include differences in valuing stocks amongst investors, the fact that a wide range of returns are attained by all investors, and most notable, that some investors are able to attain higher than market or average annual returns. Thus, many believe that earnings can indeed be predicted.

There are generally two approaches to predicting earnings. The first one being a DuPont type of financial statement analysis, the second one to be time-series and cross-sectional approaches which take advantage of earnings properties. In the past, most research on earnings prediction used a time-series model to predict future earnings. However, new researches have used cross-sectional models, which incorporates auto-correlation properties of earnings. Some of the key predictors that have been used in past research include ratio of dividends to book equity and market-value-to-assets ratio. Our report will focus on the use of a cross-sectional model to predict future earnings.

Recreation and Evaluation of Earnings Prediction Models

In this report, we begin by looking at a total of five models to predict future EPS (earnings per share) or earnings, using the data from 2010 to 2015. We then choose the model which generates the most precise forecast and add additional predictors onto it to improve the

prediction accuracy. Before we build each of the five models, we begin by preparing all the variables we need.

Data Selection and Preparation

We start by selecting variables from compa.funda and compa.company to form two datasets: funda and company. The funda dataset mainly contains the company's annual financial information, for example, Net Income. The company data set contains the company's general information, for example, S&P Quality Ranking.

For the funda dataset, the variables we selected from compa.funda are as below:

Variable	**Definition**
exchg	Stock Exchange Code
gvkey	Global Company Key
fyear	Data Year - Fiscal
oiadp	Income from Continuing Operations
ni	Net Income
csho	Common Shares Outstanding
at	Total Asset
act	Total Current Asset
che	Cash and Short-Term Investments
lct	Total Current Liability
dlc	Total Debt in Current Liabilities
txp	Income Tax Payables
dp	Depreciation and Amortization
prcc_f	Prie Close - Annual - Fiscal
sich	Standard Industrial Classification - Historical
dvt	Total Dividends

lt	Total Liabilities
ceq	Book Value

For the dataset company, the variables we selected from compa.company are as below:

Variable	**Definition**
gvkey	Global Company Key
sic	Standard Industry Classification Code
conml	Company Legal Name
spcsrc	S&P Quality Ranking - Current

To create a complete dataset, we merged the two above by matching the variable gykey. We then create lag variables (variables from the previous year) for variables at, act, che, lct, dlc and txp. Then, we calculated the average total assets, change in current asset, change in cash, change in current liability, change in debt included in current liabilities, change in income taxes payable and depreciation using the below formulas:

$$Average\ Total\ Assets = \frac{AT_t + AT_{t-1}}{2}$$

$$\Delta CA\ (Change\ in\ Current\ Assets) = ACT_t - ACT_{t-1}$$

$$\Delta Cash\ (Change\ in\ Cash) = CHE_t - CHE_{t-1}$$

$$\Delta CL\ (Change\ in\ Current\ Liabilities) = LCT_t - LCT_{t-1}$$

$$\Delta STD\ (Change\ in\ Debt\ Included\ in\ Current\ Liabilities) = DLC_t - DLC_{t-1}$$

$$\Delta TP\ (Change\ in\ Income\ Taxes\ Payable) = TXP_t - TXP_{t-1}$$

$$Dep\ (Depreciation\ and\ Amortization\ Expense) = DP_t$$

We also calculate the Accruals, Accrual Component, Cash Flow Component, Earnings, and EPS (earnings per share) using the following formulas:

$$Accruals = (\Delta CA - \Delta Cash) - (\Delta CL - \Delta STD - \Delta TP) - Dep$$

$$Accrual\ Component = \frac{Accruals}{Average\ Total\ Asset}$$

$$Cash\ Flow\ Component = Earnings - Accrual\ Component$$

$$Earnings = \frac{OIADP}{Average\ Total\ Asset}$$

$$EPS = \frac{Net\ Income}{Common\ Shares\ Outstanding}$$

We also had to create other variables:

1. Dummy variable for firms that pay dividends by using variable dividends payment:
    - 1 for firms pay dividends, and 0 for firms didn't pay dividends
2. Dummy variable for firms with negative earnings by using variable Earnings:
    - 1 for firms with negative earnings, and 0 for firms with non-negative earnings
3. Dummy variables for firms with negative EPS (Earnings per Share) by using variable epsfx:
    - 1 for firms with negative EPS, and 0 for firms with non-negative EPS
4. Target variables: Eps_lead and Earnings_lead (next year's actual EPS and Earnings) by using variables EPS and Earnings.

Model Selection and Evaluation

We conduct a total of 5 models: Random Walk (RW), HVZ, Earnings Persistence (EP), Residual Income (RI) and Earnings = Cash Flow Components + Accrual Components. The detailed equation for each model is as below:

1. $Random\ Walk\ Model: EPS_{lead} = EPS + \in$
2. $HVZ: \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 DD_{i,t} + \alpha_4 E_{i,t} + \alpha_5 NegE_{i,t} + \alpha_6 AC_{i,t} + \in_{i,t+\tau}$
3. $EP: E_{t+\tau} = \beta_0 + \beta_1 * NegE_t + \beta_2 * E_t + \beta_3 * NegE * E_t + \in$
4. $RI: E_{t+\tau} = X_0 + X_1 * NegE_t + X_2 * E_t + X_3 * NegE * E_t + X_4 * B_t + X_5 + TACC_t + \in$
5. $Earnings = Cash\ Flow\ Components + Accrual\ Components$

For model RW, EP and RI, we want to predict future EPS, therefore the target variable is Eps_lead. For the other two models, we want to predict future earnings so the target variable will be Earnings_lead. The formulas and regression summary tables are shown as below:

Model	1. RW	2. HVZ	3. EP	4. RI	5. Earnings
Intercept	1.6751 (0.514)	-0.0011 (0.033)	2.4870 (0.656)	2.3741 (0.681)	-0.1041 (0.019)
EPS	0.2473*** (0.007)	-	0.3025*** (0.008)	0.3025*** (0.008)	-
Earnings	-	0.5695*** (0.005)	-	-	-
A	-	5.4e-07 (2.39e-06)	-	-	-
D	-	-1.57e-05 (7.25e-05)	-	-	-
DD	-	0.0715 (0.043)	-	-	-
Neg_E1	-	-0.3919*** (0.044)	-	-	-
AC	-	0.1403*** (0.005)	-	-0.0035 (0.12)	0.7146*** (0.005)
Neg_E	-	-	-3.0227 (1.038)	-2.9272 (1.05)	-
Neg_E:EPS	-	-	-0.5920*** (0.025)	-0.592*** (0.025)	-
B	-	-	-	4.728e-05 (7.75e-05)	-

CashFlow	-	-	-	-	0.5768*** (0.005)
R-Squared	0.046	0.468	0.069	0.069	0.466

For HVZ, A is Total Asset, D is Dividend Payments, DD is the dummy for firms that pay dividends, NegE is the dummy for firms with negative earnings, AC is Total Accruals and E is Earnings. For EP, B represents Book Value.

Although the HVZ model has the highest R-Squared (0.468), it is used to predict earnings instead of EPS. In order to control the effect of firm sizes, we want to predict EPS instead of the actual earnings and thus we decided to focus on the RW, EP and RI models.

Out of the three models, EP and RI have the higher R-Squared (0.069), which means 6.9% of the data fit the regressions. Out of the EP and RI model, we decided to choose the EP model as the best model mainly due to three reasons:

1. The p-value for two predictors (NegE and NegE*E) are close to 0, which means they are highly significant. RI model added two predictors (B and TACC) on the EP model, but the p-value for B and AC are 0.542 and 0.976 respectively. The two p-values are both much greater than the threshold 0.05, which means B and TACC are not significant and do not help to predict EPS in our case.
2. EP model has less number of variables but can achieve the same R-squared as RI model. Including fewer variables can reduce the variance and the dimensionality of the model.
3. Using a simpler model like EP will give us more choice to decide on adding which factors to improve the model in the next step.

Model Improvement

Although the EP model results in a relatively high accuracy rate among the other two models that predict EPS, this report also investigates a range of variables that are not included in the

original EP model that have the potential to contribute to a higher prediction accuracy. For instance, we believe market value could be helpful in predicting earnings because companies that have higher market value potentially have larger market share, which could translate to a higher profitability level. After an initial screening based on accounting knowledge, we have decided to experiment model accuracy contribution for the following five variables: Year End Stock Price (Prcc_f), Dividend(D), Dividend Dummies (DD), Negative Accrual dummy variable(Neg_AC), Stock Rating (spcsrc).

Firstly, we investigated the model accuracy contribution of Dividend by including it in the EP model and found that it failed to improve model performance. Since the associated P-value of the variable Dividend is at 0.809, which exceeds a 5% threshold, we conclude that it is not statistically significant. Therefore, we decided to not include Dividend as an explanatory factor to our new model.

Although addition of the magnitude of dividends issued proved to be statistically insignificant, whether companies issue dividends or not seems to provide valuable information in predicting future annual EPS. Therefore, we included the dividend dummy into the model. Although including dividend dummy results in similar model accuracy in terms of R-squared, dividend dummy is still a fairly valuable predictor in this model considering the coefficient is large enough (0.3) that would provide a great insight about the directional impact of the EPS (positively or negatively). Such finding indicates that the action of issuing dividend contributes to higher annual EPS. It is likely that the ability to issue dividends reflects solid financial health and management confidence in future growth, which indicates higher EPS potential in the future.

We then tried another variable. We added a grouped stock rate to our new model. We grouped stock rates into 3 dummy variables. All the ‘A+’, ‘A’, ‘A-’, is categorized to “stock_rate_A”; All the “B+”, “B”, “B-” is categorized into “stock_rate_B”; while all others is grouped into

"stock_rate_others". The result is very negative with a decrease in R-squared to 0.06 and no change in training accuracy (mean-square-error). Therefore, none of these dummy variables is a good explaining factor to predict earnings per share, so they are not to be included in the final model.

Then, we tried another variable which is Negative Accrual Indicator. It is derived from accruals that if the accruals is negative it will be valued as 1, vise versa. After removing missing values and infinite values, all the mean-square-error remains the same, while R-squared decreased to 0.06. Moreover, the coefficient is -1.1, which can be interpreted to be a very negative relationship with the earnings of next year. If the accrual of a company is negative, then the earnings per share of the following year is likely to be decreasing. Moreover, it may even be decreasing more than 10%.

Lastly, we included Year End Stock Price to our prediction model. The p-value is very low (smaller than 0.001), which indicates that it is statistically significant. In addition, we have also decreased the mean-square-error to 5983. Therefore, it is statistically important to add this variable to the improved new model.

After experimenting with adding different variables to the EP model, we decided to add the following three variables in order to achieve the highest possible model accuracy: Year End Stock Price (Prcc_f), Dividend dummy variable (DD) and Negative Accrual dummy variable(Neg_AC). The improved model achieved a higher R-squared from 0.069 to 0.071. The summary of the additional factors is listed in the table before:

	Factors	coefficient	P Values	R-Squared	Adj. R-Square	MSE
0	D	0.000244	8.088931e-01	0.069048	0.068888	5994.287521
1	prcc_f	0.043114	1.321352e-10	0.070698	0.070538	5983.661660
2	DD	0.301968	7.851872e-01	0.069049	0.068888	5994.283444
3	Neg_AC	-1.151739	3.400057e-01	0.069082	0.068922	5994.067710
4	stock_rate_A	-0.228476	9.051109e-01	0.069046	0.068886	5994.298947
5	stock_rate_B	0.605246	5.889862e-01	0.069057	0.068897	5994.227289
6	stock_rate_others	-0.640648	6.038962e-01	0.069057	0.068896	5994.233164

The result of the final improved model is shown in the graph below:

OLS Regression Results

Dep. Variable:	Eps_lead	R-squared:	0.071
Model:	OLS	Adj. R-squared:	0.071
Method:	Least Squares	F-statistic:	294.8
Date:	Fri, 20 Mar 2020	Prob (F-statistic):	0.00
Time:	21:26:10	Log-Likelihood:	-1.3401e+05
No. Observations:	23237	AIC:	2.680e+05
Df Residuals:	23230	BIC:	2.681e+05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.0757	1.255	1.654	0.098	-0.384	4.536
Neg_E	-1.7705	1.121	-1.579	0.114	-3.968	0.427
EPS	0.2974	0.008	38.684	0.000	0.282	0.312
Neg_E:EPS	-0.5736	0.025	-22.865	0.000	-0.623	-0.524
prcc_f	0.0435	0.007	6.466	0.000	0.030	0.057
DD	0.0817	1.115	0.073	0.942	-2.103	2.266
Neg_AC	-1.5112	1.214	-1.245	0.213	-3.891	0.868

Omnibus:	87246.562	Durbin-Watson:	1.615
Prob(Omnibus):	0.000	Jarque-Bera (JB):	79576313786.032
Skew:	83.807	Prob(JB):	0.00
Kurtosis:	9067.279	Cond. No.	273.

In all, the final model included a few new factors: Year End Stock Price (Prcc_f), and Dividend Dummy Variable (DD); while the other factors we tried such as grouped stock rate dummy variables and Amount of Dividend were not included in the final model. We have successfully

increased R-squared by 2.9% (0.069 to 0.071) and successfully decreased mean-square-error from 5994.3 EPS to 5984.1 EPS.

2017 Annual EPS Forecast

Since the modified model improves the R-squared by 2.9%, it is important for us to analyze the predicted results a bit more deeply. In order to analyze the results more efficiently, we choose a sample of the population in our dataset to adequately represent the entire financial market. In terms of evaluating the performance of the model, “random.choice” function was performed to pick the 10 firms for further analysis about the forecast accuracy. The information about randomly selected companies are listed below:

gvkey	Company	Industry	Stock ratings	Closing price	Negative Accrual	Market Value	EPS	Eps_2017	Predicted_Eps_2017	Error%
9667	Sherwin-Williams Co (The)	General building materials	A+	268.74	0	38496	12.18	18.88	17.79	6%
13323	Sanderson Farms Inc	Poultry farming	B	89.98	1	3411	8.33	12.27	5.65	117%
23812	Regeneron Pharmaceuticals Inc	Pharmaceuticals	B	367.09	0	40462	8.45	11.14	23.15	52%
6335	Kansas City Southern	transportation	B+	84.85	1	10842	4.48	9.34	4.73	97%
121077	F5 Networks Inc	Technology	B+	124.64	1	7546	5.60	6.72	7.31	8%
165675	Viacom Inc	media	B+	38.1	1	11641	3.62	4.66	1.76	165%
7316	Michaels Cos Inc (The)	Retail	B+	19.67	1	4888	1.96	2.15	0.37	486%
6379	Kelly Services Inc.	Professional services	B	22.92	1	1051	3.16	1.86	0.76	144%
184256	Ironwood Pharmaceuticals Inc	Pharmaceuticals	C	15.29	1	2255	-0.56	-0.78	-0.26	202%
14163	McClatchy Co (The)	Publishing	C	13.18	1	69	-4.51	-43.19	-1.32	3181%

As we used a random approach in picking the 10 firms, they differ from each other in terms of the business, industry and information environment. For example, Sherwin-Williams Co is a general building material company while Ironwood Pharmaceuticals Inc is a pharmaceutical company. Company sizes also vary across the 10 businesses. Firms such as Sherwin-Willians Co and Regeneron Pharmaceuticals Inc have a relatively high market value while McClatchy Co’s market value is a fraction of theirs. The collection of 10 firms also include stocks from a variety of ratings from A+ to C and the majority of the companies have negative accrual.

In terms of the accuracy on the predicted 2017 EPS, the overall error rate is quite high, which is expected. However, there are a couple of the companies from different industries that the modified model performed quite well, specifically Sherwin-Williams Co and F5 Networks Inc. Sherwin-Williams Co is meant to be one of the companies that is relatively easier to be

forecasted with the modified model. The first reason one is that Sherwin-Williams Co is an American Fortune 500 company with businesses in many countries around the world. Second, the homebuilding industry has been developing rapidly in the last decades meaning there has been enormous demands from building material companies like Sherwin-Williams Co. Therefore, Sherwin-Williams Co has had steadily positive returns over the few years since the 2008 financial crisis. F5 Network Inc is a technology company where it focuses on delivery, security, performance, and availability of web applications. As the technology evolves, there has been again a tremendous demand from companies like F5 Network that can provide data security and web applications. Thus, the company has had a positive return for the last few years and therefore the modified model performed well. On the other hand, the modified model performed poorly on companies like Michaels Cos Inc and McClatchy Co. Michaels Cos Inc is North America's largest provider of arts, crafts, framing, floral & wall décor. The company heavily relies on smooth inflow and outflow inventories and since the marginal profit is fairly low, indicating that they consistently require a large number of consumers in order to run a profitable business. McClatchy Co, a publishing company, is relatively a small-sized company compared to others that were previously discussed. One of the biggest downsides for this company is that as the internet users grows, more and more people start to adapt digital reports. This likely has led to a lower demand in the publishing industry and the company’s earnings have been moving downward in recent years. Ultimately, the modified model performs poorly due to the information that fluctuates the earnings not being captured by the model.

Company	Eps_2017	Predicted_Eps_2017	Error%	Analyst's forecast	Analyst's error%
Sherwin-Williams Co (The)	18.88	17.79	6%	9.43	100%
Sanderson Farms Inc	12.27	5.65	117%	4.54	170%
Regeneron Pharmaceuticals Inc	11.14	23.15	52%	8.70	28%
Kansas City Southern	9.34	4.73	97%	7.34	27%
F5 Networks Inc	6.72	7.31	8%	4.65	45%
Viacom Inc	4.66	1.76	165%	4.40	6%
Michaels Cos Inc (The)	2.15	0.37	486%	1.06	103%
Kelly Services Inc.	1.86	0.76	144%	1.02	83%
Ironwood Pharmaceuticals Inc	-0.78	-0.26	202%	-0.19	309%
McClatchy Co (The)	-43.19	-1.32	3181%	-2.49	1634%

According to the comparison above, it appears that analyst forecasts (meanest from IBES.statsum_epsus) only outperform our model around 50% of the time based on the 10 firms we have selected. Due to the complex nature of earning forecasts, even analysts find it difficult to consistently produce accurate predictions despite the access to more information, domain knowledge and extensive modeling capabilities. Similar to our model, analyst's forecast can fall far away from the actual annual EPS, such is the case of McClatchy Co, in which analyst forecast error reaches 1634%. However, it is also possible that the portfolio of the ten companies is not representative of all publicly traded companies, and therefore we cannot reach the conclusion that analyst's predictions are systematically on par with our model predictions.

Ten Highest EPS Firms Prediction

Since many companies have yet to report annual financial information on the fiscal year of 2019, this report will use financial information from 2018 to predict firm performances in 2019 in terms of annual Earning Per Share, using the model developed in the above analysis. The table below summarizes the top ten performing companies and their corresponding annual Earning Per Share predictions.

Company	Dividend	Negative EPS	Negative Earning	Negative Accrual	2019 Predicted EPS
Booking Holdings Inc	0	0	0	1	$117.22
Amazon.com Inc	0	0	0	1	$93.23
Capital Financial Holdings Inc	0.062	1	0	0	$76.51
Alphabet Inc	0	0	0	1	$69.25
Constellation Software Inc	115.713	0	0	1	$55.75
AutoZone Inc	0	0	0	1	$53.61
ATRION Corp	9.473	0	0	1	$46.76
Graham Holdings Co	28.617	0	0	1	$45.83
Mettler-Toledo International Inc	0	0	0	1	$36.43
Intuitive Surgical Inc	0	0	0	0	$30.16

Based on our analysis, we found several commonalities among the ten most profitable stocks. Most of the companies neither have negative earnings nor negative earnings per share. We also find that most of the top ten companies have negative accrual. Upon first look, it indicates a potential ‘cookie jar’ behavior that companies use to boost revenue, but further investigation shows that many of the above companies run paid upfront services that prompt them to recognize revenue over time.

Although the 2019 annual Earning Per Share data is yet to be available to assess the prediction accuracy from our model and analysis made above, our prediction of 2018 annual EPS based on 2017 indicates limited model predictive power and therefore we expect the significant discrepancy to persist for the fiscal year of 2019 as well. We believe the limited predictive power is due to the fact that earning prediction is a complex process and subjected to the influence of a range of variables. Hence, the model we have developed for the purpose of this analysis will be inadequate to predict the actual earning since it only contains a few predictors. For example, we did not include variables such as annual GDP to reflect the macroeconomic condition, which will have a market-wide impact. Although the model accuracy is subject to further improvement by taking into account more variables, we believe it is more informative to interpret the model variable coefficients to identify the influence of various financial information on predicted annual Earning Per Share.

Conclusion

Earnings prediction remains to be a very difficult task. In our analyses, we recreate models developed in past researches in an attempt to predict future earnings. Building on the best model we found, we attempt to find and add new variables to improve model accuracy. As seen from our analysis, our final and best model did not provide a high test accuracy. Moreover, we do not expect our model's accuracy on actual future earnings prediction to be highly accurate as well. There could be many reasons why earnings prediction is difficult through simply just modelling. A potential reason could be the fact that investment analysts and professionals often have more private information that is not available for the public which can provide much more insightful information that contributes very well to earnings prediction. Although the model we developed were not as successful in earnings prediction as we had hoped, we were able to derive insightful information on financial variables that have a significant impact on earnings, as well as those that do not.

References

Malkiel, Burton Gordon. (2003). *A random walk down Wall Street: the time-tested strategy for successful investing*. New York :W.W. Norton,

Rotman School of Management
UNIVERSITY OF TORONTO

Rotman

Final Page

Grade: _____