

Accuracy of Analysts' Earnings Estimates



UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Bachelor of Science (B.Sc.) in Finance

Spring of 2019

Supervisor: Marcin Zamojski

Authors: Joakim Nilsson & Philip Svensson

Abstract

This thesis investigates consensus and individual analyst firm accuracy in forecasts of earnings per share (EPS) for U.S. stocks in 2009–2018. Moreover, we investigate if the analysts' forecasting predictiveness is affected by the size of the company which is observed. Finally, we examine if differently weighted models can beat an equally weighted consensus forecast. We find statistical evidence that analysts' forecasts of EPS have predictive power. Furthermore, we find that the size of a company impacts the predictive ability of analysts. Analysts of larger companies, included in S&P 500, are more accurate in their forecasts, relative to analysts of smaller companies, included in Russell 2000. Finally, after categorizing the analyst firms by predictiveness, we create models to explore the possibility of beating the average with weighted combinations. Our results show that none of the suggested models are statistically significantly different from the consensus and therefore we cannot conclude that differently weighted models outperform an equally weighted consensus.

Keywords: Financial Forecasting, Analyst Accuracy, Consensus Estimates

JEL Classification: G17

Acknowledgements

We would like to express our great appreciation to our supervisor Marcin Zamojski for his guidance and useful support during the writing process of this thesis. Furthermore, we would like to offer our thanks to the Finance Department at the School of Business, Economics and Law at the University of Gothenburg for providing us access to financial databases such as Bloomberg and Thomson Reuters.

Table of Contents

1. Introduction	5
1.1 Background	5
1.2 Purpose	6
1.3 Hypotheses	6
1.4 Summary of Results	7
2. Literature Review	8
3. Theoretical Framework	11
3.1 Earnings Per Share	11
3.2 Fixed Effects Panel Regression	11
3.3 Diebold-Mariano tests	12
4. Data	14
5. Results	16
5.1 Analyst Accuracy	16
5.2 Company Size Effect on Analyst Accuracy	19
5.3 Individual Analyst Firm Accuracy	21
5.4 Alternative Consensus Models	25
6. Conclusions	28
References	29
Appendix	31

1. Introduction

1.1 Background

Analysts are in the business of selling their opinions, but are their forecasts worth the price? Predicting the future in all fields of finance has been proven to be a difficult task (Timmermann, 2018), and has for several reasons been questioned if it is valuable at all. A financial analyst is expected to produce reliable forecasts of earnings, growth, and company performance. Financial analysts are an important part of the financial system and their analysis affects stock prices, at least in the short term (Hilary and Hsu, 2013).

Numerous researchers have conducted studies on topics related to the challenges and difficulties in financial forecasting. Bias in financial forecasting has been observed for instance by Sedor (2002), and Hilary and Hsu (2013). However, these studies are somewhat contradictory, mainly derived from the effects of intentional- and unintentional bias. Whilst Sedor suggests that analysts tend to give overly optimistic forecasts, Hilary and Hsu observe the tendency for sell-side analysts to intentionally produce estimates below the outcomes, i.e., lowballing.

Furthermore, analysts are incentivized to do a good job since they are evaluated on their projections and thus being accurate may lead to a successful career, as examined by Hong and Kubik (2003). However, this may lead to analysts' hesitation to disagree with the consensus because it could risk their reputation and, in the worst-case scenario, their employment. The effect that forecasts has on career can, by this reasoning instead, be regarded as a disadvantage due to potential management bias and fear of contrarian projections, which in turn might lead to inaccurate estimations.

It is arguably easy to evaluate analyst performance in hindsight, which makes them vulnerable to criticism. Questions about analyst performance have been recurring throughout time (Dreman and Berry, 1995) and are still as relevant as ever. In this paper we investigate whether financial forecasts are valuable at all.

1.2 Purpose

We study the financial forecasting landscape and evaluate analyst accuracy for U.S. stocks in 2009–2018. As previous studies by Timmerman (2018), and Dreman and Berry (1995) conclude, there are several difficulties and challenges in financial forecasting. Analysts forecasts are hardly a new aspect in the field of finance and their work has a strong impact on the financial markets. It is important to examine historical forecasts to evaluate if in fact they are worth their price.

More specifically this paper aims to determine if forecasts of financial earnings are valuable at all. Furthermore, it investigates how the predictiveness of the forecasted earnings is affected by company size. Finally, individual analyst firms are compared against one another and ranked by predictiveness with the ambition to create a more predictive model than an equally weighted consensus. This set of problems leads us to our main research question: are analysts' forecasts valuable in predicting earnings per share?

1.3 Hypotheses

Dreman and Berry (1995) suggest that forecasting errors have been frequent in the past and are consistent throughout business cycles. We evaluate analysts' earnings per share projections and assess whether it has been something to attribute weight to during the last ten years. Our first hypothesis is to investigate if analysts' forecasts of EPS for U.S. stocks were predictive in 2009–2018.

Consensus forecasts are generally more predictive than individual forecasts (Clements, 2015). Smaller companies in our sample are usually covered by only a handful of analysts whilst larger companies are more intensively covered (Appendix IV). Therefore, the consensus of the larger companies will contain more individual estimates which mean that it is less affected by potential outliers amongst the estimates. Furthermore, Dichev and Tang (2009) suggest that earnings predictability and earnings volatility are negatively correlated, which in this case would further highlight the difficulty in predicting the earnings of smaller companies. The proposed difficulties in forecasting companies of different size leads to our second hypothesis, which is that the predictive power of larger U.S. companies EPS is greater than for smaller companies.

As Andrade and Le Bihan (2013) suggest, analysts tend to come to different conclusions even though they use the same public information. Analyst interpretation of information is therefore directly decisive in the accuracy of individual analyst firms. For our third hypothesis we investigate whether some analyst firms are more accurate in predicting EPS than others and should be given more consideration when making investment decisions.

The results from the third hypothesis gives us the means necessary to create differently weighted models with the ambition to outperform the equally weighted consensus.

1.4 Summary of Results

We find statistical evidence that analysts' forecasts of EPS have predictive power. Furthermore, we find that the predictive power is greater in forecasts for larger companies, included in S&P 500, relative to smaller companies, included in Russell 2000. Finally, after creating differently weighted models based on the analyst firms predictiveness, we could not conclude that differently weighted models outperform an equally weighted consensus with statistical significance.

2. Literature Review

Timmermann (2018) highlights several of the key challenges in financial forecasting. Some of these difficulties in establishing predictability are low signal-to-noise ratio, persistent predictors, and model instability. These all stem from an information overload in one way or another. However, these challenges do not only portray the difficulties in financial forecasting but also demonstrate why analysts come to different conclusions in forecasting even though they use the same information. The emphasis should, therefore, be on interpretation of information. Furthermore, Timmermann discusses forecasting methods with the ambition to overcome these predictability challenges. On the topic of forecasting methods, Genre, Kenny et al. (2012), model the ECB Survey of Professional Forecasters (SPF) to see if there are any benefits in combining methods to outperform the equally weighted consensus estimate. Methods such as principal components and trimmed means, performance-based weighting, least square estimates of optimal weights and Bayesian shrinkage. They find that no single model dominates throughout the sample and that the combined models were not statistically significant. This leads them to conclude that there is no case to replace equal weighting in preference to the combination models on the forecasts of the ECB SPF. Similarly, Clements (2015) researches if survey forecasters can outperform a simple time-series model in which variables move monotonically towards the long run expectation.¹ For some economic variables, he finds that the survey consensus forecasts are superior to the time-series model. The consensus forecast is particularly accurate for CPI inflation, unemployment rate, and the Treasury bill rate. Finally, his research suggests that for most individual forecasters it would be beneficial to make simple mechanical adjustments to improve their accuracy.

Accurate forecasting in finance is arguably difficult, which has been demonstrated in several studies. Dreman and Berry (1995) look at 66,100 consensus estimates in 1974–1991 to evaluate potential analyst forecasting errors. Their results are consistent with other research showing that analysts tend to be too optimistic when issuing forecasts (Sedor, 2002). Furthermore, they find the forecasting errors to be consistent through business cycles and across industries during the

¹ The long-term target is divided by the number of periods to get the expected change within each period.

time period, which indicates that analyst errors are in fact present independently of circumstances. They also find that forecasting errors are more frequent than previously anticipated as well as increasing throughout the observed period. This confirms the belief that forecasting errors are frequent and highlights the difficulty in attributing significance to analysts' projections. Dichev and Tang (2009) suggest how company size will affect the analyst firms forecasting accuracy. They conclude that there is a negative relationship between earnings' volatility and predictability. Smaller companies tend to have more fluctuations in their earnings whilst larger companies are more stable and reliable. The earnings' volatility will, by this reasoning, naturally impact the predictability of earnings. Furthermore, coverage depends on the size of a company, with a positive correlation between analyses issued and market value (Appendix IV). Whilst bigger companies are intensively covered, relatively small companies tend to have less analyst coverage which could make the consensus estimate less accurate.

However, difficulty is not the only factor that complicates forecasting. In many cases an analyst's bias—either unintentional or intentional, plays an important role. Sedor (2002) gives reasons for unintentional bias in financial forecasting. She argues that financial analysts tend to be optimistic rather than pessimistic, especially when it comes to companies with recent losses. More specifically she tests whether forecast optimism is a consequence of an analyst's reaction to the structure of information about managers' future plans. She concludes that when an analyst is offered thoroughly explained information from the managers, rather than as written lists, they are more inclined to give an optimistic estimate over the coming two years. Moreover, the results indicate that the effect was stronger when the firm had suffered a recent loss, which could have implications on forecasting for smaller companies which are more prone to losses. These results encourage the view that there might be some harm in too much interaction with executive officers or other managerial staff. Managers are often skilled at portraying their company's future plans as exciting, which corresponds with the thoroughly explained information versus list argument, and in turn mislead analysts. Finally, she argues that analysts might add an intentional bias to maintain a sound relationship with the management, which could be damaged by a pessimistic forecast. Furthermore, Andrade and Le Bihan (2013) suggest that forecasters are consistently unintentionally biased. By researching the ECB Survey of Professional Forecasters, they find that even when using the same inputs forecasters come to different conclusions. Furthermore, they propose that the disagreements and bias stem in part

from inattentiveness,² where they find that twenty percent of the observed forecasters do not incorporate new information released each quarter.

Intentional bias, where the analyst themselves add a bias in the forecast, is widely covered in research. Hong and Kubik (2003) provide possible explanations for this phenomenon. They conclude that accurate financial forecasters are more likely to have relatively better career opportunities. They suggest that this leads to a conservative approach where analysts tend to forecast earnings close to what they expect the consensus to be, contrary to the earnings that they truly expect. An inaccurate forecast far from the consensus is worse than an equally inaccurate forecast in line with the consensus because it is, in the worst case, as inaccurate as the majority. Moreover, Hong and Kubik observe a positive relationship between a sell-side analyst being more positive than consensus and more favourable career opportunities. This is explained by the benefits it has to the investment bank in generating banking business and brokerage fees. These economic advantages give support to their findings that the positive relationship is stronger when observing analysts who covered stocks underwritten by their own analyst firms because it has a direct effect on their own revenue. Hilary and Hsu (2013) suggest that analysts who deliver more consistent forecasting errors have a greater ability to move prices, i.e., bigger market impact. Additionally, in line with Hong and Kubik, they show that the analyst performance has consequences for the analysts' careers. In this case, however, in forecasting error consistency. They claim that it is better to be consistently wrong than inconsistently inaccurate. A more consistent analyst is less likely to be demoted to less prestigious brokerage houses and more likely to be nominated to the All-Star Analysts list by Institutional Investor magazine. Furthermore, they also conclude that accurate analysts have a greater impact on prices than inaccurate analysts. Aligned with Sedor (2002), Hilary and Hsu (2013) also explain why forecasts might be biased. They claim that analysts might intentionally add a downward bias, known as lowballing, to help the managers beat their estimates. This could in the long run have a positive impact on the relationship between the analyst and the manager, which could make the analysts' job easier because the managers are more accommodating towards them. Consequently, the biased forecasters will be more consistent and more informative than unbiased forecasters, rewarding the lowballing analysts for their foul play.

² Failure to revise forecasts when new information is available.

3. Theoretical Framework

3.1 Earnings Per Share

We base our study on Earnings Per Share (EPS) ratio. EPS is a simple measure that facilitates comparisons of companies even though the size may differ. However, the EPS ratio can be viewed differently depending on one's perspective. There is the realized EPS, which as earlier mentioned is simply the company's earnings divided by the number of shares. From an analyst's perspective the model is extended to capture the fact that it is a forecast of expected events. The forecasted EPS proposed by Keane and Runkle (1998) reads as following:

$$EPS_{n,t+1}^j = E(EPS_{t+1}^j | I_{n,t}) \quad (1)$$

where, t denotes the year, j denotes the firm, n denotes the analyst, and $I_{n,t}$ is the information set that is available to the analyst n at time t . The emphasis is on the information set because it is the base of the projections. Furthermore, the information set will be subject to the covering analyst interpretation and inattention as discussed by Andrade and Le Bihan (2013). Naturally, some information of a company will be available to the public whilst some depends on the relationship between the analyst and the forecasted company (Hillary and Hsu, 2013). Whatever approach the analyst utilizes to arrive at the forecasted earnings—it will be dependent on available information, which tends to depend on the size of the analysed company. Information that could consist of current projects, competitor analysis, interaction with managers, etc. The varying information but most importantly interpretation of the individual analyst leads to different EPS forecasts.

3.2 Fixed Effects Panel Regression

A fixed effects model is often used with panel data and is a model in which the group means are fixed. Compared to a standard regression model the fixed effects model enables a causal effect to be observed under weaker assumptions. Fixed effects regressions provide unbiased estimates if unobserved confounders are present, in contrast to a standard regression which provides biased estimates of causal effect. By this reasoning, a model with fixed effects is particularly appropriate in the context of causal inference (Brüderl and Ludwig, 2014). The

model on which we base part of our study is a panel regression with analyst and time fixed effects as described by the following equation:

$$EPS_{t,i} = \beta_0 + \beta_1 EPS_{t-1,i} + \beta_2 \text{Forecasted } EPS_{t-1,i} + \varepsilon_{it} + \eta_t + \xi_i \quad (2)$$

where $EPS_{t,i}$ denotes the reported EPS, $EPS_{t-1,i}$ denotes the previous quarters EPS, $\text{Forecasted } EPS_{t-1,i}$ denotes the consensus forecast issued by the analysts, ε_{it} denotes the error term, η_t denotes time fixed effects, and ξ_i denotes firm fixed effects. The main goal of the model is to find if $\text{Forecasted } EPS_{t-1,i}$ is predictive of $EPS_{t,i}$. The $EPS_{t-1,i}$ is included to facilitate a comparison between the analyst consensus estimate as a predictor and observing the previous quarter's earnings and expecting earnings based on this. The inclusion of time fixed effects is to control for time variations for the variable $EPS_{t-1,i}$, while firm fixed effects aims to control for differences across individual analyst firms.

3.3 Diebold-Mariano tests

A Diebold-Mariano test is commonly used when two or more forecasting models on the same variable of interest is available. In this thesis the DM test is utilized to determine whether forecasts issued by competing analyst firms differ with statistical significance. The model, presented by Diebold and Mariano (1995), is based directly on predictive performance and can be tailored to fit different settings. The DM test builds on the residuals of forecasting errors where the residual is defined as:

$$e_{it} = \hat{y}_{it} - y_t \quad (3)$$

where \hat{y}_{it} denotes the forecast, y_t denotes the outcome, and e_{it} denotes the residual. The complete equation for calculating the DM-statistic reads as following:

$$DM = \frac{\bar{d}}{\sqrt{\frac{[\gamma_0 + 2\sum_{k=1}^{h-1} \gamma_k]}{n}}} \quad (4)$$

where, d is the difference of the forecasts' residuals squared, \bar{d} denotes the average of d throughout the time period of observation, γ is autocovariance, h is equal to $n^{1/3} + 1$, k is lag,

and n is number of observations. The DM test allows for other accuracy methods to be used compared to previously developed tests. Forecast errors can have a mean not equal to zero, and be serially correlated, contemporaneously correlated and non-Gaussian (Diebold and Mariano, 1995).

4. Data

Our sample consists of one hundred randomly selected U.S. companies, divided into four parts containing 25 companies each taken from different size categories (Appendix I). The categories are: (i) the one hundred largest companies in the S&P 500 index, (ii) the one hundred smallest companies in the S&P 500 index, (iii) the range of 900–1000 in the Russell 2000 index, and (iv) the range of 1900–2000 in the Russell 2000 index. This sample gives us a good variety of differently sized companies with different coverage and the means necessary to compare larger to smaller companies.

For each company we obtain consensus earnings per share estimates on a quarterly basis, reported earnings per share on a quarterly basis (ADJ+ comparable), and number of analysts for each quarter (Appendix IV), for 2009–2018. This data comes from Bloomberg (retrieved on March 27, 2019). This data is used to test Hypothesis I and II.

The individual company's analyst coverage for each quarter in 2009–2018, in Appendix IV, is summarized in Table 1.

Table 1: Summary Statistics of Analysts Coverage

	Top 100 from S&P 500 (i)	400–500 from S&P 500 (ii)	900–1000 from Russell 2000 (iii)	1900–2000 from Russell 2000 (iv)
25 Percentile	18.25	8.60	2.35	2.25
75 Percentile	23.95	14.95	6.51	4.75
Median	20.78	12.92	4.13	3.81
Mean	21.45	13.51	5.58	3.84
Standard Deviation	5.11	6.51	4.59	1.87

Table containing summary statistics for analyst coverage per consensus estimate across S&P 500 and Russell 2000 indices.

From Table 1 it is visible that there is a positive correlation between analyses issued and market value. Companies from the Russell 2000 index are considerably less covered than the companies from the S&P 500 index. Furthermore, the observed relationship seems not only to

be viable for comparison across the two indices but also within the indices themselves. Companies from category (i) are more intensively covered than companies from category (ii), and companies from category (iii) are more intensively covered than companies from category (iv).

In addition to the data from Bloomberg, we obtain individual analyst firms quarterly EPS estimates from Reuters for 2016–2018 (retrieved on April 4, 2019). Detailed historical forecasts are only available for 2016–2018 and, therefore, the analysis of Hypothesis III uses a shorter sample period than Hypothesis I and II. This data is used to test Hypothesis III and serves as a base for the weighted models in Subsection 5.4.

5. Results

This part of the thesis presents our results and discussions related to the hypotheses outlined in the introduction. It is presented with tables summarizing the regressions for each individual hypothesis followed by an econometric analysis, and a discussion of the findings. We use a significance level of five percent to determine if our results are significant.

5.1 Analyst Accuracy

To answer Hypothesis I (predictiveness of analysts' forecasts of EPS for U.S. stocks in 2009–2018) we use consensus EPS projections by analysts and compare them to actual outcomes in the financial statements. We measure accuracy of the consensus estimate by looking at its forecast error. To investigate the efficiency of the analyst's forecasts, we run a panel regression for which the sample period is 2009–2018. The panel dataset is unbalanced with gaps where there is no available data. The reason there is no data is because the companies were either not listed or lacked coverage for the time period. To test our first hypothesis, we run six regressions. First, we run a panel regression without fixed effects (i). Furthermore, we run two regressions, testing each of the separate variables with fixed effects. In (ii) we have time fixed effects and in (iii) we have analyst fixed effects. Finally, we run the panel regression with time and analyst fixed effects to conclude whether the forecasts are on average predictive (iv). To test the robustness of the results from regressions (i)–(iv) we run two additional univariate regressions. The regressions are run with clustered standard errors for the individual companies.

Table 2: Analyst Accuracy

	<i>Coefficient</i>	<i>SE</i>	<i>t-value</i>	<i>p-value</i>	<i>R²</i>	<i>Observations</i>
$EPS_{t,i} = \beta_0 + \beta_1 EPS_{t-1,i} + \beta_2 \text{Forecasted } EPS_{t-1,i} + \varepsilon_{it}$ (i)						
<i>EPS_{t-1}</i>	0.197	0.101	19.32	0.000	0.864	3313
<i>Forecasted EPS_{t-1}</i>	0.799	0.100	79.58	0.000		
$EPS_{t,i} = \beta_0 + \beta_1 EPS_{t-1,i} + \beta_2 \text{Forecasted } EPS_{t-1,i} + \varepsilon_{it} + \eta_t$ (ii)						
<i>EPS_{t-1}</i>	0.197	0.099	1.97	0.049	0.866	3313
<i>Forecasted EPS_{t-1}</i>	0.799	0.155	5.15	0.000		
$EPS_{t,i} = \beta_0 + \beta_1 EPS_{t-1,i} + \beta_2 \text{Forecasted } EPS_{t-1,i} + \varepsilon_{it} + \xi_i$ (iii)						
<i>EPS_{t-1}</i>	0.184	0.087	2.11	0.035	0.867	3313
<i>Forecasted EPS_{t-1}</i>	0.793	0.170	4.66	0.000		
$EPS_{t,i} = \beta_0 + \beta_1 EPS_{t-1,i} + \beta_2 \text{Forecasted } EPS_{t-1,i} + \varepsilon_{it} + \eta_t + \xi_i$ (iv)						
<i>EPS_{t-1}</i>	0.183	0.086	2.12	0.034	0.869	3313
<i>Forecasted EPS_{t-1}</i>	0.792	0.169	4.68	0.000		

Regression (i): Panel regression without fixed effects. Regression (ii) Panel regression with time fixed effects but not analyst fixed effects. Regression (iii): Panel regression with analyst fixed effects but not time fixed effects. Regression (iv): Panel regression with time and analyst fixed effects.

In table 2 we can see that the four regressions all have high t-values and low p-values, below our threshold of five percent significance level. Furthermore, the high R^2 for the regressions leads us to believe that the models have some explanatory power. Even though we test for all different combinations of time and analyst fixed, all four regressions generate very similar coefficients of the lagged EPS and Forecasted EPS. We can therefore conclude that the

estimated relationship between EPS, lagged EPS and Forecasted EPS is not affected by omitted variable bias due to factors that are constant over time. This conclusion, in combination with the similar coefficients which serve as a robustness test, strengthens our observation. The low p-values, in conjunction with the high coefficients on Forecasted EPS in relation to lagged EPS, allows us to confirm that analysts' forecasts are more predictive of future EPS than relying on last quarter's EPS.

In addition to the regressions with the independent variables lagged EPS and Forecasted EPS we run two univariate regressions with one independent variable in each of the models. In model (v), with lagged EPS as the independent variable, we use time fixed effects. In model (vi), with Forecasted EPS as the independent variable, we use analyst fixed effects.

Table 3: Univariate Regressions

	<i>Coefficient</i>	<i>SE</i>	<i>t-value</i>	<i>p-value</i>	<i>R²</i>	<i>Observations</i>
$EPS_{t,i} = \beta_0 + \beta_1 EPS_{t-1,i} + \varepsilon_{it} + \eta_t$ (v)						
<i>EPS_{t-1}</i>	0.795	0.089	8.86	0.000	0.61	3482
$EPS_{t,i} = \beta_0 + \beta_1 Forecasted\ EPS_{t-1,i} + \varepsilon_{it} + \xi_i$ (vi)						
<i>Forecasted EPS_{t-1}</i>	0.909	0.096	9.38	0.000	0.856	3415

Regression (v): Panel regression with time fixed effects, EPS lagged as the independent variable and EPS as the dependent variable. Regression (vi): Panel regression with analyst fixed effects, Forecasted EPS as the independent variable and EPS as the dependent variable.

The coefficient for lagged EPS in the univariate regression (v) exceeds the coefficients from table 2 with a wide margin. When only including lagged EPS as an independent variable it shows a clear relationship with the actual outcome of EPS. Furthermore, the t-value is high enough to indicate that the relationship is statistically significant. However, the Forecasted EPS still has a coefficient closer to one. This coupled with the high t-value and R^2 strengthens our belief that analysts' forecasts have more predictive value than lagged EPS.

Table 2 and 3 clearly display the relationship between the higher Forecasted EPS relative to the lagged EPS. However, when analyzing the results, the timing of the data must be considered. Lagged EPS is, as earlier mentioned, based on the previous quarters EPS, which means that the

data is approximately four months old. On the other side, forecasted EPS is an average of the forecasts from the individual analyst firms. These are generally released within the last weeks before the company reports its earnings and are therefore the results of close to four months' worth of information. Also, they are often associated with costs as opposed to the lagged EPS which is free. An argument could therefore be made that these factors make the comparison unviable. However, this argument overlooks the key aspect and purpose of financial analysts and their forecasts—gathering, interpreting, and summarizing information. The comparison between the Forecasted EPS and lagged EPS is therefore valid if one considers the potential costs that can be associated with the analysts forecast, but also the aspect of time.

It is important that investors are informed about analysts' predictive power in forecasting and if it is worthwhile to follow their advice. Our findings lead us to conclude that analysts are predictive of future earnings per share. Analysts' forecasts have more predictive value than the last quarter's EPS in forecasting, which is observed by the high difference between the two explanatory variables in regression (iv). The relationship suggests that analysts' forecasts on EPS should be considered when making investments. However, as mentioned previously, forecasts tend to be associated with costs which must be taken into consideration. The predictiveness of the consensus estimate in our thesis contradicts the previous findings of Dreman and Berry (1995) who find that the average forecasting errors is high enough to question the usefulness of analysts and suggest that investors should not rely on consensus estimates.

5.2 Company Size Effect on Analyst Accuracy

To answer Hypothesis II (predictiveness of analysts' forecasts for smaller versus larger U.S. companies), we use the same data as for Hypothesis I. We extend our model to include a size dummy where an inclusion in S&P 500 indicates a large company and an inclusion in Russell 2000 indicates a small company. We research the potential difficulty in forecasting reliable consensus estimates for small companies versus for large ones as measured by market value, i.e., what influence company size has on the analysts predictiveness. This is completed by a panel regression with time and analyst fixed effects, and with interaction and dummy variables representing large and small companies. The regression is run with clustered standard errors for the individual companies.

Table 4: Company Size Effect on Analyst Accuracy

	<i>Coefficient</i>	<i>SE</i>	<i>t-value</i>	<i>p-value</i>	<i>R²</i>	<i>Observations</i>
$EPS_t = \beta_0 + \beta_1 EPS_{t-1} + \beta_2 FEPS_{t-1} \times DSmall + \beta_3 FEPS_{t-1} \times DLarge + \varepsilon_t + \eta_t + \xi_i$ (vii)						
<i>EPS_{t-1}</i>	0.186	0.094	1.99	0.046	0.805	3313
<i>Forecasted EPS Small</i>	0.766	0.187	4.09	0.000		
<i>Forecasted EPS Large</i>	0.899	0.055	16.35	0.000		

Panel regression with time and analyst fixed effects, and with interaction between Forecasted EPS and a dummy variable representing small and large companies.

The model (vii) allows us to compare the relationship and significance of forecasting on small relative to large companies. The t- and p-values indicate that all explanatory variables are significant at a five percent level. The R^2 confirms that the model has some explanatory power. With the inclusion of an interaction term and a dummy variable for size we are able to incorporate the effect size has on the coefficients. The coefficient of the Forecasted EPS on larger companies is considerably higher than the coefficient of the Forecasted EPS on smaller companies. Interpreting the coefficients of the estimated EPS leads us to conclude that analysts were more accurate in their forecasts on large companies relative to small companies in 2009–2018. Thus, we can verify that company size has an effect on analyst forecasting accuracy.

The results that analysts are more accurate in their forecasts for larger than smaller companies correspond with expectations based on previous studies such as Dichev and Tang (2009). They state that smaller companies tend to have more volatile earnings, which in turn affects the predictability of earnings. Naturally, these fluctuations in earnings for smaller companies makes it difficult to forecast with precision. Vice versa, larger companies are generally more stable in terms of growth, earnings and cash flows, which simplifies forecasting. Furthermore, regulation differs across indices. In order to be included in S&P 500 the company must report profits for five consecutive quarters (S&P U.S. Indices Methodology, 2019). Russell 2000 does not have the same requirements (Russell U.S. Equity Indexes). This leads to exclusively profitable companies in S&P 500 index, whilst companies in the Russell 2000 index possibly could have

recent losses. Sedor (2002) suggests, that there is a tendency for analysts to be overly optimistic in forecasting companies with recent losses. This unintentional bias increases the potential for more frequent errors for companies in the Russell 2000. The general difference in regulation for the two indices could therefore possibly have implications on the accuracy of the forecasts.

An additional aspect that could contribute to the discrepancy between large and small companies is the more intense analyst coverage for the larger ones. More analysts should reasonably lead to a more reliable consensus estimate since it incorporates many different opinions and generates a more balanced average. For instance, Facebook, included in the S&P 500 index, averages 28.7 analyst estimates per quarter whilst Remark Holdings Inc, included in the Russell 2000 index, averages one analyst estimate per quarter. This vast difference of analyst coverage between the two indices is apparent throughout our sample where the companies from the S&P 500 average 17.5 analysts whilst the companies in the Russell 2000 average 4.7 (Appendix IV).

5.3 Individual Analyst Firm Accuracy

To answer Hypothesis III (predictiveness of individual analyst firms), we use the individual analyst firm estimates from Reuters for 2016–2018. We investigate the performance of individual analyst firms to gauge their relative predictiveness. This gives us the means necessary for comparing the firms on their forecasting accuracy. This comparison is accomplished with the use of the Diebold-Mariano test to observe which forecasters are the most accurate and inaccurate.

When calculating the DM-statistics we produce a matrix with the 123 analyst firms in our sample on the axes. Two analyst firms only generate DM-statistics if they have forecasted earnings for the same company and during the same quarter as another analyst firm in more than one quarter. If a cell in the matrix is blank it means that there are not more than one comparative quarter between two analyst firms. Which, for instance, is the case between Avondale Partners and Obsidian Research Group in Figure 1. The maximum possible number

of comparable DM-statistics for our sample would be 7503.³ After excluding the quarters where n was not higher than one we arrive at the complete matrix which contains 1092 DM-statistics.⁴

Figure 1: Subset of DM-Statistics

	AVONDALE PARTNERS, LLC	RAYMOND JAMES	WILLIAM BLAIR & COMPANY
AVONDALE PARTNERS, LLC		0.452	0.098
RAYMOND JAMES	-0.452		-0.095
WILLIAM BLAIR & COMPANY	-0.098	0.095	
OBSIDIAN RESEARCH GROUP			
CANTOR FITZGERALD		0.215	0.354
ATLANTIC EQUITIES		-0.088	-0.167
CREDIT SUISSE		0.070	-0.126
EVERCORE ISI		0.022	-0.139
GUGGENHEIM SECURITIES LLC		0.122	-0.196
PIPER JAFFRAY		0.088	-0.018
SUNTRUST ROBINSON HUMPHREY CAPITAL MARKETS		0.121	-0.002
SVB LEERINK		-0.216	-0.028
WOLFE RESEARCH		-0.231	-0.386
BTIG		0.140	0.151

Excerpt from the full matrix containing all analyst firms on the axes. The matrix contains all DM-statistics for when analyst firms has issued forecasts for the same company and time-period in more than one quarter.

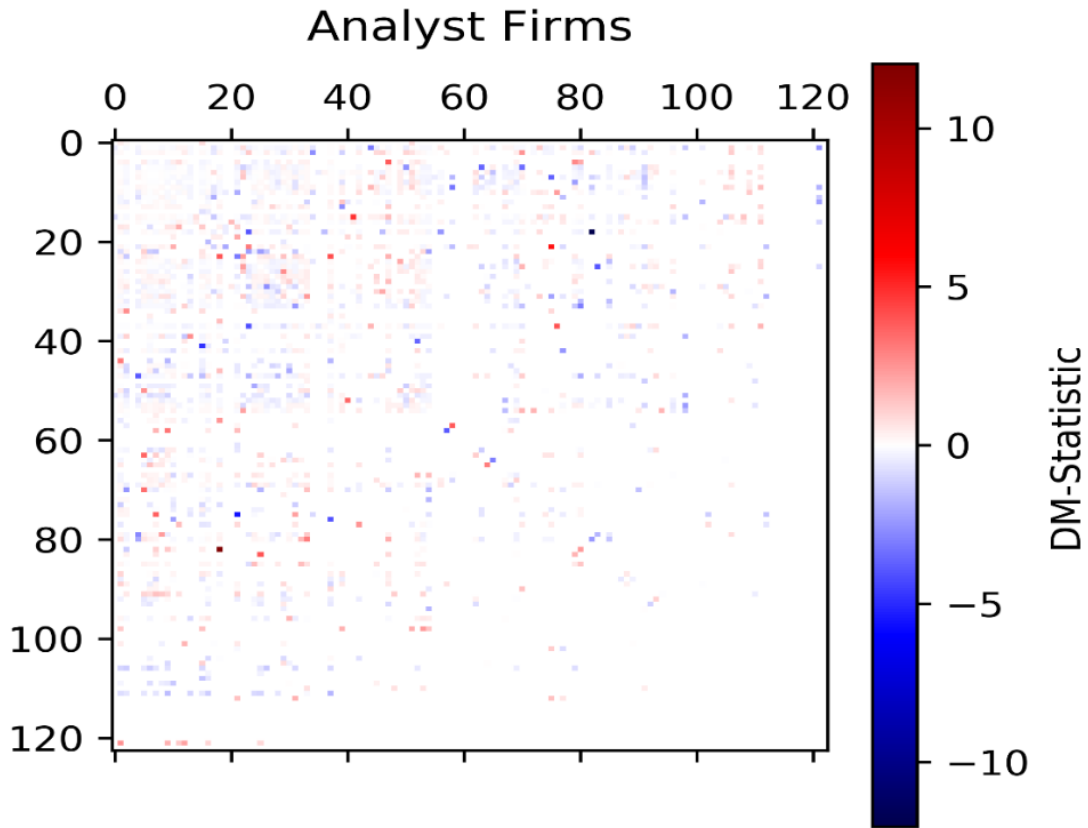
Figure 1 shows a subset of the complete matrix that contains the DM-statistics for the analyst firms. If the value is positive/negative the firm in the column produced smaller/larger forecast squared errors than the company in the row. For instance, this suggests that Avondale Partners have performed worse than Raymond James and William Blair & Company. A DM-statistic larger than 1.96 in absolute values is significant at a five percent level. This implies that Avondale Partners comparative values are not statistically significant.

After running the DM-test, the different significance levels of the complete matrix are illustrated in Figure 2. The DM-statistics are mainly centred around zero, which is depicted by the dimly coloured squares. As observable, most forecasters do not produce significantly better/worse forecasts and are excluded with a typical statistical threshold. Accordingly, different significance levels do not have big implications for the analysis in this case.

³ Actually, there are 15006 DM-statistics. However, each one is mirrored one time as can be observed in the comparative values between Avondale Partners, LLC and Raymond James. Where the DM-statistic is -0.452 and 0.452 respectively. $15006/2=7503$.

⁴ $2184/2=1092$.

Figure 2: DM-Statistic



The figure represents the full matrix with DM-statistics but with colours instead of values, with the analyst firms on both axes. The boxes are the analyst firms' colour-coded DM-statistics where a dimly shaded box represents a DM-statistic close to zero.

To better depict the significant values at a five percent level we filter out the insignificant values and remove them from the analysis, illustrated in Appendix III. There are not that many occasions where we can observe different analyst firms' forecasts for the same company and time-period that are statistically significant at a five percent level. This means that from the original matrix containing 1092 comparative occasions we are left with a matrix with only 42 DM-statistics (Appendix III).⁵ The small sample size is a contributing factor to the high p-values and insignificant DM-statistics. Our uncertainty about whether the analyst firms outperformed each other or not would decrease with a larger sample size. Since there are not that many comparable occasions for the analyst firms only very large differences in performance yields a small p-value.

⁵ 84/2=42.

Figure 3: Selected Performers

	RAYMOND JAMES	WILLIAM BLAIR & COMPANY	CANTOR FITZGERALD	ATLANTIC EQUITIES	EVERCORE ISI	GUGGENHEIM SECURITIES LLC	PIPER JAFFRAY	SUNTRUST ROBINSON HUMPHREY	WOLFE RESEARCH	BTIG	STIFEL NICOLAUS AND COMPANY	KEYBANC CAPITAL MARKETS INC.	BWS FINANCIAL	B. RILEY FBR, INC.	DOUGHERTY & COMPANY LLC
RAYMOND JAMES															
WILLIAM BLAIR & COMPANY															
CANTOR FITZGERALD															
ATLANTIC EQUITIES															
EVERCORE ISI															
GUGGENHEIM SECURITIES LLC															
PIPER JAFFRAY															
SUNTRUST ROBINSON HUMPHREY														-1.963	
WOLFE RESEARCH															
BTIG															
STIFEL NICOLAUS AND COMPANY															
KEYBANC CAPITAL MARKETS INC.															
BWS FINANCIAL														2.039	
B. RILEY FBR, INC.								1.964					-2.038		
DOUGHERTY & COMPANY LLC															
GRIFFIN SECURITIES, INC.												3.968		-3.118	2.293
JMP SECURITIES															
OPPENHEIMER & CO., INC.															2.293
PACIFIC CREST SECURITIES-KBCM															
STEPHENS INC.															
SUSQUEHANNA FINANCIAL GROUP LLLP															
WUNDERLICH SECURITIES															
WEDBUSH SECURITIES INC.															
JANNEY MONTGOMERY SCOTT LLC		2.725													
MACQUARIE RESEARCH															
LONGBOW RESEARCH										2.679					
SEAPORT GLOBAL SECURITIES LLC															
IMPERIAL CAPITAL, LLC											-4.871				
THE BUCKINGHAM RESEARCH GROUP															
BERENBERG	3.142														
CRT CAPITAL															
NOMURA SECURITIES INTL (AMERICA)			-3.868												
ASCENDANT CAPITAL MARKETS															
MILLMAN RESEARCH ASSOCIATES				2.544											
MKM PARTNERS															
NEEDHAM & COMPANY INC.															
D.A. DAVIDSON & COMPANY															2.187
SCOTIA HOWARD WEIL												2.464			
CJS SECURITIES															
ALEMBIC GLOBAL ADVISORS							3.272								
FELTL & COMPANY				3.662											
SCOTIABANK GBM															
TD SECURITIES															
B. RILEY & CO.		-2.4822		3.422											
NORTHLAND SECURITIES					3.708										-5.565
HUBER RESEARCH PARTNERS									-2.345						
CRAIG HALLUM															
FBN SECURITIES			-2.793			2.276									
MONNESS, CRESPI, HARDT, & CO INC.			-1.994												
AEGIS CAPITAL												12.027			
CRISPIDEA															
SUMMIT INSIGHTS GROUP															
NEPHRON RESEARCH	2.592								2.214						

Excerpt from Appendix III which is a matrix where all insignificant DM-statistics are excluded.

As observed in Figure 3, which is a subset of Appendix III, there are some analyst firms that have been outperforming their peers in forecasting EPS, whilst some analyst firms have been underperforming in their predictions. As seen in the green coloured columns, there are three analyst firms that outperform their competitors in all three tests with statistical significance. These are: Atlantic Equities, Keybank Capital Market Inc, and Dougherty & Company LLC. Furthermore, there is one analyst firm that underperformed in all three tests with statistical significance: Cantor Fitzgerald.

However, it is difficult to draw too drastic conclusions from the three most accurate analyst firms or the poor performing Cantor Fitzgerald. To fully determine if they are better or worse at forecasting than their competitors, a larger sample size would be needed. To further the analysis and determine if their accuracy stems from being more prone to bias than the others, we would have to take a more qualitative approach.

Even though a larger sample size would be preferable, the results bring some insight into the difference in accuracy of analyst firms. The results from the Diebold-Mariano test are in line with Hypothesis III and suggests that some analyst firms are more accurate in predicting EPS than others and should be given more consideration when making investment decisions.

5.4 Alternative Consensus Models

To expand on the outcome of the Diebold-Mariano test we create our own consensus models that attribute weight based on past performance in forecasting EPS. A historically more accurate analyst firm receives a higher weight than a less accurate firm. The differently weighted models are then compared with the equally weighted consensus to illustrate the accuracy of the models.

To fairly observe the accuracy and test the performance of the models we divide the dataset into two parts—training and test set. The first two years (2016–2017) determines the composition of the models whilst the third year (2018) serves as an evaluation for the performance of the models. This division helps us to avoid look-ahead bias.⁶ ‘Consensus’ gives equal weight to all analyst firms in the data set. The ‘Performance Based’ model only includes

⁶ A bias that occurs when a study relies on data that was not yet available during the time period of study.

analyst firms that have outperformed other analyst firms with statistical significance and weigh these based on the number of firms which they have outperformed. For instance, the outperforming Atlantic Equities, Keybank Capital Market Inc, and Dougherty & Company LLC receives a weight of three since they all outperformed three competitors with statistical significance. The 'Performance Based +1' model is similar to the 'Performance Based' model but includes analyst firms that have neither outperformed nor underperformed (others). Furthermore, in the 'Excluding' model all analyst firms not underperforming are equally weighted i.e., outperformers and others. The 'Performance Based ²' model skews weights more towards the outperforming firms. For example, the three outperforming analyst firms receive a weight of nine instead of previously three. The DM-statistics of the models with different weights are summarized in Table 5. The values indicate how accurate the models are at predicting EPS in 2018. As in the previous figures, the models in the columns that have a positive value indicate that they are more predictive than the corresponding model in the rows, and vice versa.

Table 5: Weighted Models

	<i>Consensus</i>	<i>Performance Based</i>	<i>Performance Based +1</i>	<i>Excluding</i>	<i>Performance Based²</i>
<i>Consensus</i>		0.108	0.089	0.083	-0.100
<i>Performance Based</i>	-0.108		-0.044	-0.117	-0.113
<i>Performance Based + 1</i>	-0.089	0.044		-0.097	-0.093
<i>Excluding</i>	-0.083	0.017	0.097		-0.088
<i>Performance Based²</i>	0.100	0.113	0.093	0.088	

Table showcasing the DM-statistics of the weighted models. A positive value in the cell means the model in the column produces a lower squared forecast error than the model in the row.

Of the models, 'Performance Based', model produces the lowest squared forecast errors, outperforming the remaining models. On the other hand, 'Performance Based ²'

produces the highest squared forecast errors of the five models, which indicates that it is not beneficial to attribute too much weight to the top analyst firms from 2016–2017. This means that they could not continue with their previous predictiveness. The consensus model performs slightly worse than all the remaining models except for ‘Performance Based ²’, which suggests that it may be possible to beat the consensus with weighted models.

However, the DM-statistics are all close to zero, which corresponds with high p-values and insignificant differences among the models. Therefore, in line with Genre, Kenny et. al (2012), we cannot conclude that differently weighted models outperform the equally weighted consensus.

6. Conclusions

This thesis investigates if financial analysts' forecasts are valuable at all. Analysts forecasts have been questioned (Dreman and Berry, 1995) throughout time because the general difficulty in forecasting (Timmermann, 2018), but also due to factors related to bias. Analysts are an integral part of the financial system and their work affects the stock market, at least in the short term (Hilary and Hsu, 2013). It is therefore of great importance to examine historical forecasts to evaluate if in fact they are worth their price.

This paper finds statistically significant evidence that analysts are predictive when forecasting EPS. Analyst consensus estimates are far more predictive than relying on the last quarters EPS in forecasting. Furthermore, we conclude that company size has an effect on the predictive power of the consensus estimate. The consensus estimates for the companies included in the S&P 500 index are more accurate than for the companies included in the Russell 2000 index. Moreover, when exploring the accuracy of individual analyst firms, we find that some analyst firms outperformed their peers, whilst others underperformed. However, it is hard to draw any drastic conclusion from the limited data sample and a non-qualitative approach. Finally, we explore the possibility of outperforming the equally weighted consensus with differently weighted models. We observe a tendency for the consensus model to produce higher squared forecast errors than our own models, which implies that the created models are more predictive in forecasting. However, the DM-statistics are all low which indicates insignificant differences among the models. Therefore, we cannot draw the conclusion that differently weighted models outperform the equally weighted consensus.

References

- Andrade, Philippe, and Le Bihan, Hervé, (2013), Inattentive Professional Forecasters, *Journal of Monetary Economics*, Vol. 60 (8), pp. 967-982
- Brüderl, Josef, and Ludwig, Volker, (2014), The SAGE Handbook of Regression Analysis and Causal Inference, pp. 327-354, SAGE Publications Ltd
- Clements, P. Michael, (2015), Are Professional Macroeconomic Forecasters Able to do Better than Forecasting Trends, *Journal of Money, Credit and Banking*, Vol. 47 (2-3), pp. 349-381
- Dichev, D. Ilia, and Tang, W. Vicki, (2009), Earnings Volatility and Earnings Predictability, *Journal of Accounting and Economics*, Vol. 47 (1-2), pp. 160-181
- Diebold, Francis. X, and Mariano, Roberto, (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, Vol. 13, pp. 253-265
- Dreman, N. David, and Berry, A, Michael, (1995), Analyst Forecasting Errors and Their Implications for Security Analysis, *Financial Analysts Journal*, Vol. 51 (3), pp. 30-41
- FTSE Russell, 2019. *Russell U.S Equity Indexes*. [Report]. Available at: <https://www.ftse.com/products/downloads/Russell-US-indexes.pdf?fbclid=IwAR0AbkiBVV5NFbE0EHulHZQ5IY43wl7klVo8tcTNdpqJspzDSxMQkuiH6r4> [Last visited 15-05-2019]
- Genre, Véronique, Kenny, Geoff, Meyler, Aidan, Timmermann, Allan, (2012), Combining Expert Forecasts: Can Anything beat the Simple Average?, *International Journal of Forecasting*, Vol. 29 (1), pp. 108-121
- Hilary, Gilles, and Hsu, Charles, (2013), Analyst Forecast Consistency, Vol. 68 (1), pp. 271-297
- Hong, Harrison, and Kubik, D. Jeffrey, (2003), Analyzing the Analysts': Career Concerns and Biased Earnings Forecasts, *The Journal of Finance*, Vol. 58 (1), pp. 313-35
- Keane, P. Michael, and Runkle, E. David, (1998), Are Financial Analysts' Forecasts of Corporate Profits Rational?, *Journal of Political Economy*, Vol. 106 (4), pp. 768-805

Sedor, M. Lisa, (2002), An Explanation for Unintentional Optimism in Analysts' Earnings Forecasts, *The Accounting Review*, Vol. 77 (4), pp. 731-753

S&P Dow Jones, 2019. *S&P U.S. Indices Methodology*. [Report]. Available at:
https://us.spindices.com/documents/methodologies/methodology-sp-us-indices.pdf?force_download=true [Last visited 15-05-2019]

Timmermann, Allan, (2018), Forecasting Methods in Finance, *Annual Review of Financial Economics*, Vol. 10, pp. 449-479

Appendix

A.I: Sample Companies

Top 100 from S&P 500	400–500 from S&P 500	900–1000 from Russell 2000	1900–2000 from Russell 2000
CVX	FL	TCX	MARK
FB	XRX	ETM	BRS
CMCSA	ARNC	ACCO	CBFV
AMT	IPGP	SXI	AQUA
ABT	FBHS	QUOT	INAP
SBUX	NWL	TPC	QHC
INTU	TRIP	OMER	ROX
ORCL	RJF	CETV	SUP
PNC	LB	UTL	ICBK
UNH	MHK	MLAB	SURF
PM	AIV	AKS	SELB
CHTR	ROL	DNR	VERI
BKNG	WHR	GFF	PFSW
NKE	ALLE	TDW	XOMA
ADBE	HSIC	ARCB	OPBK
COP	CF	TYPE	AAC
ABBV	WRK	EXTR	NH
MDLZ	JNPR	WAIR	AVEO
MSFT	COTY	OIS	KIRK
GS	CBOE	QADA	BW
JNJ	LKQ	JBSS	CSU
DUK	NLSN	UBA	CMRX
PG	HBI	ECPG	CBIO
GILD	MOS	CLDT	HOV
AMGN	WU	HMHC	IDRA

A.II: Literature Summary

Author	Summary
Andrade and Le Bihan (2013)	Forecasters are consistently biased and come to different conclusions even though they use the same information
Clements (2015)	Difficulties of individual forecasts
Dichev and Tang (2009)	Negative relationship between earnings' volatility and predictability
Dreman and Berry (1995)	Empirical research which concludes that forecasting errors are consistent and tend to be overly optimistic
Hilary and Hsu (2013)	Forecasting error consistency leads to better career opportunities than inconsistent errors
Hong and Kubik (2003)	Analysts' tend to forecast earnings close to consensus. Positive relationship between positive forecasters and favourable career opportunities
Sedor (2002)	Analysts' tend to be more optimistic than pessimistic. Unintentional bias due to relationship with company's executives
Timmermann (2018)	Highlights several difficulties and challenges in financial forecasting
Genre, Kenny, et al. (2012)	Examines if models can outperform the equally weighted average

A.IV: Average Number of Analysts 2009–2018

Top 100 from S&P 500	400–500 from S&P 500	900–1000 from Russell 2000	1900–2000 from Russell 2000
CVX: 17.7	FL: 16.3	TCX: 1.7	MARK: 1
FB: 28.7	XRX: 9.9	ETM: 2.2	BRS: 7.5
CMCSA: 22.3	ARNC: 9	ACCO: 4.6	CBFV: 1.5
AMT: 17.5	IPGP: 8.1	SXI: 2.5	AQUA: 7
ABT: 18.8	FBHS: 15.3	QUOT: 3.6	INAP: 3.9
SBUX: 23.3	NWL: 14	TPC: 4.7	QHC: 3.3
INTU: 16.6	TRIP: 20.1	OMER: 4.8	ROX: 1.2
ORCL: 35.3	RJF: 8.2	CETV: 5.5	SUP: 3.2
PNC: 20.7	LB: 25.3	UTL: 2.8	ICBK: 3.8
UNH: 20.6	MHK: 12.9	MLAB: 1.5	SURF: 2
PM: 16.4	AIV: 2.9	AKS: 13.2	SELB: 4.7
CHTR: 10.6	ROL: 2.8	DNR: 16.1	VERI: 2.5
BKNG: 20.8	WHR: 7.1	GFF: 1.8	PFSW: 2.4
NKE: 26.1	ALLE: 7.6	TDW: 4	XOMA: 4.6
ADBE: 24.6	HSIC: 14.4	ARCB: 14.7	OPBK: 2
COP: 19	CF: 13.8	TYPE: 4.1	AAC: 3.8
ABBV: 15.3	WRK: 14	EXTR: 4.2	NH: 3.7
MDLZ: 19.2	JNPR: 30.3	WAIR: 8.7	AVEO: 3.9
MSFT: 29.4	COTY: 12.7	OIS: 16.2	KIRK: 4.3
GS: 22.5	CBOE: 14.6	QADA: 1.5	BW: 4.8
JNJ: 20.7	LKQ: 12.4	JBSS: 1.2	CSU: 7.6
DUK: 20.2	NLSN: 12.8	UBA: 3.4	CMRX: 6.5
PG: 20.8	HBI: 11.4	ECPG: 6.3	CBIO: 2
GILD: 26.3	MOS: 17.4	CLDT: 3.7	HOV: 5.3
AMGN: 23	WU: 24.6	HMHC: 6.7	IDRA: 3.7
S&P 500 average: 17.5		Russell 2000 average: 4.7	