Predicting stock market responses by debiasing the financial analyst consensus

Thijs Koenraadt (484951)

September 17, 2018

Master Thesis submitted to obtain the degree of Master of Science in Behavioural Economics

Supervisor: Prof. dr. Aurélien Baillon

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Department of Behavioural Economics

ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Abstract

Few researchers have explored the use of crowd aggregation techniques for aggregating financial analyst forecasts. Therefore, this thesis investigates a relatively new technique in the financial analyst environment: *The Contribution Weighted Model (CWM)*, proposed by Budescu and Chen (2014). It is common knowledge that the average of all analyst forecasts, *the analyst consensus*, is often biased. This model attempts to mitigate the effect of biases by identifying expertise amongst the analysts and assigning weights to forecasts based on a total contribution score. The question *'To what extent is it possible to more accurately forecast earnings than the analyst consensus using the contribution weighted model of Budescu and Chen (2014) and to use the difference to predict stock market responses?'* will be answered. This research extracts forecast data from the Institutional Brokers' Estimate System (I/B/E/S) for analyst forecasts in the sample period of January 2011 to December 2017. Firstly, accuracy of the CWM is compared to the analyst consensus and adjusted versions of the CWM. Results of statistical analyses indicate the CWM is more accurate than the analyst consensus, but not more accurate than the adjusted versions. Secondly, the difference between the CWM and the analyst consensus is used to proxy for earnings surprises. Accuracy of the CWM does not seem to be sufficiently high to function as a relevant proxy, as no relation seem to exist between predicted earnings surprises and actual earnings surprises using the full sample. Thirdly, there does not seem to be a relation between predicted earnings surprises and stock market responses. Also, imposing multiple benchmarks does not seem to have an effect on the predictability of market responses. Further research could focus on refining the input of the CWM and investigating the drivers of market responses.

Keywords: contribution weighted model, earnings announcement, earnings surprise, market expectation, analyst consensus, financial analyst, earnings forecast, stock market response, wisdom of the crowd

Contents

1 Introduction

During quarterly earnings announcements of companies large stock price deviations are sometimes observed. Share prices are observed to change by up to 20%. Investors may see their money evaporate over the course of a day. Some stakeholders leverage the opportunity these fluctuations present, taking actions informed by estimates of share price movements. This thesis uses a wisdom of the crowd aggregation technique to predict stock price movements.

Stock price movements of companies arise when actual company earnings, announced during earnings announcements, are higher or lower than the expectated by the market. The difference between actual earnings and the expectation of the market is known as an *earnings surprise.* When actual earnings of a company are higher, the market is positively surprised and share prices typically increase, and when actual earnings are lower the market is negatively surprised and share prices are usually observed to decrease. The *expectation of the market* is equal to the *analyst consensus,* which is the average analyst forecast for a company's earnings announcement at a point in time. Therefore, the financial analyst is an important player. Analysts function as information intermediary between companies and investors. Analysts' most salient piece of information is the forecast of companies' earnings (Ramnath et al., 2008).

For many years investors have used the analyst consensus as relevant input for investment decisions. Many (amateur) investors do not have the time nor the skills to technically analyse companies and therefore rely on financial analysts (Koller et al., 2010). However, the analyst consensus is often incorrect, meaning, the expectation of the market, in many cases, differs from actual earnings of a company. A direct consequence is that large share price movements are observed during earnings announcements. An earnings surprise can be a severe financial hit for some because they see their assets diminish, and to the gain of others when the share price movement is anticipated (Ronen and Yaari, 2008).

Early research in this area focussed on finding a proxy for the markets' expectation of future company earnings. It was found that the analyst consensus was more efficiently employable to identify unanticipated earnings and subsequent market responses than time-series models [e.g. Fried and Givoly (1982), and O'Brien (1988)]. Nonetheless, the analyst consensus is not a perfect proxy for the markets' expectations. Schipper (1991) argued that an association exists between earnings surprises and market responses, but that much of this relation has yet to be discovered. This marked the beginning of two streams of research that are relevant to this thesis, namely, increasing the accuracy of the analyst consensus and finding variables that influence stock price movements.

It has focussed on forecast characteristics, Brown (1991) for example, found that recent forecasts are, on average, more accurate than older forecasts. He found increases in accuracy by aggregating forecasts of no older than 30 days before the earnings announcement. Also, research covers the environment analysts work in (Bradshaw, 2011). Analysts have found to be prone to conflicts of interest and biases. For instance, they work for financial institutions that require them to prioritise on other things than accuracy of the forecast. In terms of biases, Friesen and Weller (2006) find that overconfidence is present amongst analysts. They rely too heavily on private information. Furthermore, analyst characteristics, such as experience and expertise, affect forecast accuracy (Jacob et al. 1999). Lastly, company specific factors are important. Especially, the quality of the earnings and the richness of the information environment (Brown et al. 2015).

In terms of stock price movements, research has focused on the effect of missing, meeting, or beating the analyst consensus benchmark (Bartov et al. 2002). Especially, the effect seems to be more profound for larger earnings surprises, and less predictable for just meeting the analyst consensus (Kinney et al., 2002). Furthermore, the effect is found to be greater when additional benchmarks are met. For instance, meeting additional price targets or growth targets (Rees and Sivaramakrishnan, 2007).

Until now, no accurate measure for predicting actual earnings has been found, whereas the stock market responses to earnings surprises seem to be pronounced. Few researchers have tried to use wisdom of the crowd aggregation techniques to predict earnings surprises and stock market responses. To fill this gap, the present investigation explored the extent to which a wisdom of the crowd aggregation technique can be used to forecast earning more accurately than the analyst consensus and, thereby, be used to predict deviations in stock prices. Wisdom of the crowd entails that in the long run, the average of a group of forecasters is more accurate than each of the individual forecasts. Hence, the analyst consensus, the average of analyst forecasts, should, in theory, be an accurate measure. However, according to Surowiecky (2004), the average of a group of forecasters is only accurate when certain conditions are satisfied. For financial analysts these conditions do not hold, as analysts work in an environment where conflicts of interests and biases prevail. For instance, herding is common among inexperienced analysts. Herding is the phenomenon where analysts follow or copy each others' forecast. In the case of inexperienced analysts, they tend to give forecasts closer to the analyst consensus. Budescu and Chen (2014) designed a crowd aggregation technique which gets around biases and conflicts of interests by identifying experts amongst the analysts. Budescu and Chen (2014) call this aggregation technique, the contribution weighted model (CWM). To date, two studies have provided support for the accuracy of this model. Especially, when expertise amongst the crowd is high and information sharing possible.

In the present case, expertise is ostensibly present as analysts working for large financial institutions are hired for their abilities. Also, information sharing is to a lesser extent available, as they can see each other's research reports and gain knowledge from multiple information sources such as company management. Therefore, it is expected that the CWM has significant value for financial analyst forecasts.

An example is used to further clarify what this thesis will investigate. The analyst consensus forecasts $10 Earnings Per Share (EPS) for company X, the CWM forecasts $13 EPS. Based on the CWM forecast a positive earnings surprise of $3 is expected. Actual EPS turns out to be $15. Therefore, actual earnings surprise is $5, and it is expected that that the market responds positively by which share price will increase. The predicted earnings surprise of $3 could have been used to predict a share price increase. Therefore, the more accurate the CWM, the more precisely earnings surprises and likely stock market responses can be predicted.

2 Literature review

2.1 Background information

This paragraph is provided to give background information for readers unfamiliar with the topic. Specifically, information is given on the financial environment analysts work in. The following concepts are explained: earnings announcements, earnings per share, financial analyst, analyst consensus, conflicts of interest, and biases.

In most countries, quarterly earnings announcements are mandatory for publicly traded firms. Throughout the year several information sources, including management announcements, press releases and pre-announcements, inform the investor on companies' earnings and have an effect on stock return variance. From these information sources, only earnings announcements are disclosed in fixed quarterly disclosing periods. Therefore, it is easy for the market to anticipate new information becoming available. Furthermore, regulation requires earnings announcement to satisfy certain conditions, making them a reliable information source (Shon and Zhou, 2011).

During an earnings announcement companies must disclose financial information such as company financial fitness, management analysis, market risk disclosures, industry analysis, and internal controls. Also included is the earnings per share (EPS) metric, generally regarded as the most salient piece of information (Koller et al., 2010). EPS is the profit of a company divided by the number of shares. Buying a share in a company gives a return that is equal to EPS (conditional on whether all profits are passed on to shareholders). EPS is a widely used metric because it is easy to interpret and communicate (Henry et al., 2010).

In the period before the earnings announcements, financial analysts convey their projection on the financial performance of companies by drawing up research reports. They forecast what financials will be announced during the next earnings announcement. The most salient parts of the reports are investment advice (i.e., sell, buy, or hold), EPS forecast, and share price targets. Financial analysts work for brokerage houses, investment banks or research institutes and provide forecasts and recommendations on firms accompanied by a clarifying research report. Hence, they are also called sell-side analysts as their research reports influence stock sales volume or establish lucrative deals with analysed firms from which their employers benefit (Ramnath, 2008).

Usually analysts specialise in an industry and a subset of companies within that industry (Dunn and Nathan, 2005). Analysts base their analysis on vast amounts of private and public information sources such as financial statements, industry prospects, and communication with management. People have limited time or ability and therefore rely on the work of analysts to interpret, and communicate this information. Communication happens through informal and formal channels with clients, investors, industry leaders, company management and other market participants (Bradshaw, 2011).

Ever since it was discovered that the analyst consensus is more accurate than time-series forecasting model in forecasting quarterly earnings, and more closely associated with stock returns, it has been used by researchers as a proxy for the markets' expectation on future company earnings (Kothari, 2001). Furthermore, it is used to quantify earnings surprises.

Notwithstanding that analysts function as information intermediaries and their aggregate forecast is used to proxy for the markets' expectation on future earnings, their forecasts are often inaccurate.Literature attributes this to the financial environment analysts work in, which makes them prone to conflicts of interest and biases.

Bradshaw (2011) summarised the most common conflicts of interest concerning analysts.

1. Investment banking fees: Analysts work at investment banks which do everything in their power to lure in lucrative deals. For instance, they demand analysts draw up overly optimistic reports to be selected to do the underwriting for a client (Lin and McNichols, 1998).
2. Management relationships: Analysts get important relevant information directly from management. To uphold a good relationship and not be excluded from this information source, analysts are believed to draw up optimistic reports (Francis and Philbrick, 1993).
3. Increase trading volume: Another income source for investment banks is investor trading in stock. When analyst reports convey an optimistic message regarding stocks, investors are more inclined to buy them, leading to more revenues for the investment bank firm.
4. Institutional investor relationships: Institutional investors are on the receiving end of investment advice. Sometimes institutional ownership in a stock is large due to analyst research, which makes analysts less inclined to draw up negative reports, as this could affect stock return variance, and be to the detriment of the relationship with the institution.
5. Paying for research: Firms covered by analysts receive more investor attention and have easier access to capital. Nonetheless, only half of public firms are followed. A recent finding is that companies pay analysts to follow their firm. Therefore, analysts are inclined to convey an optimistic message.
6. Affinity with firm: Analysts usually cover firms long-term. It is argued that analysts develop affinity with the firm or management. This could lead to optimistic reports.

Furthermore, analysts are prone to biases. Hong et al. (2000) find that less experienced analysts herd more and report forecasts later in time. This can be attributed to career concerns as bold forecasts are associated with increases in termination. Also, analysts with large portfolios tend to herd because they have limited time to analyse companies. Furthermore, Friesen and Weller (2006) find that, on average, analysts rely too much on private information due to overconfidence. They also find that cognitive dissonance is present amongst analysts. When analysts draw up optimistic reports, they tend to acquire information that backs up this initial report, rather than weighting information against and in favour equally. Lastly, Easterwood and Nutt (1999) found that analysts do not incorporate new information effectively. They respond differently to positive information, to which they overreact, than to negative information, to which they underreact.

2.2 Aggregating analyst forecasts and accuracy

This is an important paragraph as the contribution weighted model (CWM) will be explained. The CWM is the foundation of this thesis. It is used to aggregate analyst forecasts more effectively than the analyst consensus. It is expected that the model will be more accurate for most earnings announcements. However, that accuracy will be dependent on certain model specific factors and drivers defined by financial literature. The forecast difference between the analyst consensus and the model will be used to predict earnings surprises and stock market responses. The latter two definitions will be elaborated on in paragraph 1.3: Earnings surprises and market responses.

First, the principles and conditions of wisdom of the crowd will be explained. Then, a high-level overview of the CWM will be provided (Budescu and Chen, 2014). Also, accuracy drivers specific to the CWM are covered in this sub-paragraph. Further, factors from financial literature that make analyst forecasts more accurate, and the scope of this thesis will be covered. Lastly, a short summarising sub-paragraph will be provided.

2.2.1 Wisdom of the crowd

In 1906, Francis Galton found that a simple crowd estimated the weight of an ox very accurately. He simply averaged all forecasts on the ox's weight. He found that a simple mean of the group usually outperforms the most expert (or accurate?) individual of the group. He called this phenomenon Vox Populi, wisdom of the crowd. This has been validated in many experiments that followed, e.g., Gordon (1924) and Treynor (1987). The analyst consensus is similarly a simple average of all analyst forecasts for an earnings announcement. According to wisdom of the crowd, this should be very accurate. However, this is not always the case. Surowiecki (2004) states that the crowd is only accurate when four conditions are satisfied. The conditions are diversity, independence, decentralisation, and aggregation:

Diversity is mainly referred to in the cognitive sense. Cognitive diversity is the diversity in people's ideas, views, and solutions. As such, when every individual analyst brings out a forecast, and has a different perspective on a company's' earnings than others, they are said to be diverse. The more diverse these perspectives the better. Also, different perspectives can come about from, for instance, having had different schooling or being raised from a different value perspective.

The independence condition postulates that analysts should not be influenced by other analysts. Analysts may base their perspective on similar information, however, they should not be influenced by each other during the formation of their perspectives. The idea is that when analysts base their perspective on others, it is possible that their judgement becomes biased towards a wrong group judgement. Thus, when individuals follow each other, it is possible that the individual with the wrong solution is followed. An example of not satisfying the independence condition is herding, which was seen amongst inexperienced analysts.

Decentralisation is the distribution of power among local authorities rather than having one central authority. The idea is, when power is decentralised, local authorities or individuals, guided by their own beliefs, views and knowledge, will develop different ways of tackling problems than other local authorities or individuals. Thus, local specialisations and tacit knowledge will be enhanced. Analysts work for various firms and are therefore quasi-decentralised. Analysts working for the same firm might form similar ideas regarding the future earnings of a company based on similar information. Whereas analysts working for different firms probably forecast on the merits of different views and information.

Aggregation of the information is the last condition. In stock markets, the stock price reflects quite accurately the aggregation of private information every investor has. Therefore, the aggregation of analyst forecasts is the expectation of the market of future earnings of a company.

2.2.2 Contribution weighted model (Budescu and Chen, 2014)

Budescu and Chen designed a crowd aggregation model that intents to get rid of biases at the aggregate level. The CWM intents to this by identifying expertise amongst the crowd rather than aggregating the whole crowd. Specifically, it weights judges' (analysts) forecasts relative to other judges. Each judge conveys an opinion or forecast on a matter. In doing so, this judge changes the consensus up to that point. It is argued that the more a judge changes the group consensus by deviating from the consensus, the newer information this judge holds. Therefore, a difference between the average of the group before and after the judges' additional forecast can be seen, according to Budescu and Chen (2014), as the judges' expertise or misjudgement. As mentioned above, judges' forecasts are weighted. The model (CWM), assigns weights to judges based on the contribution of a forecast to the group average. Each forecast constitutes a positive contribution when it changes the group consensus in the direction of the actual opinion or outcome, or negative contribution, when it changes away from it.

For example, Microsoft reveals its quarterly financials and states that EPS is $10. Analyst consensus (the average of all analyst forecasts) was $11. On the one hand, some analysts predicted EPS of above $11, moving the average away from actual EPS of $10, negatively contributing to the group average. On the other hand, some analysts predicted EPS of below $11, positively contributing to the group average by moving it closer to actual EPS. In this instance, the latter group of analysts would be assigned positive weights, and the former negative weights.

Evidently, one forecast is not enough to identify real expertise. Budescu and Chen (2014) found that as the number of forecasts per analyst grows, the model becomes more accurate. Therefore, by repeating the above process, each analyst will be assigned a total contribution score computed by taking the average contribution of all his forecasts. Analysts with a positive total contribution score (i.e. score > 0) are considered positive contributors, and with a negative score negative contributors. Budescu and Chen (2014) argue that aggregating only positive contributors often leads to highly accurate forecasts.

As the concept of the model can be difficult to grasp at first, it is shortly summarised again. Beware of the terms used, as they will be used throughout the thesis:

1) Analysts are assigned a contribution score per forecast. This score is positive when the forecast moves the mean closer to actual earnings, and negative when it moves the mean further away.
2) For each analyst individually, the score of all his forecasts is averaged. This average is the total contribution score.
3) For each earnings announcement, the CWM aggregates only forecasts of analysts who have a positive total contribution score. Analysts with a positive total contribution score are called positive contributors.

Technicalities of the model will be explained at the methodology section. Below, previous environments in which the CWM is tested, and the drivers identified by Budescu and Chen (2014) and Chen et al. (2016), that make the model more accurate are outlined. These drivers will be used in the formation of hypotheses. Namely, they will be used to explain why the CWM is more accurate for some earnings announcements than for others.

The model has been tested in three environments. Budescu and Chen (2014) first applied it on a Forecasting ACE project website. On the website judges could forecast events on several domains, for instance, law, politics, or sports events. Here the forecasters focussed on binary questions only (e.g. Will Trump win the elections?). Budescu and Chen (2014) compared the model to various other aggregation methods and found the CWM to be most accurate. The second environment is more closely related to the topic of this thesis. Budescu and Chen (2014) acquired a dataset from the European Central Bank's Survey of Professional Forecasters. In this dataset, forecasters predict Europe's future inflation and GDP. In contrast to the previous study, judges forecast continuous values. Again, superior performance of the CWM compared to other models was found (Budescu and Chen, 2014). Lastly, the CWM was tested in a special forecasting study (Good Judgment Project). Questions with binary outcomes were answered by judges from all over the world. This project was introduced as a contest between various crowd aggregation techniques. Again, the CWM was most accurate (Chen et al. 2016). Budescu and Chen (2014) and Chen et al. (2016) identified a few drivers that specifically lead to higher performance. These are:

- *Increasing the level of total contribution score.* Chen et al. (2016) found that judges with a higher total contribution score make the CWM more accurate. Specifically, they investigated this by increasing the threshold of the total contribution score from 0 to a newly defined level. Only analysts with a total average contribution of above the newly defined threshold are considered positive contributors and are aggregated. The reason for increased accuracy is intuitive. Analysts with higher scores are believed to hold superior information. When analysts are consistent and calibrated, their total contribution score will be higher than analysts who hold less information or are less calibrated. The former analyst will be assigned mostly positive contribution scores per forecast, and the latter analyst will be assigned more negative contribution scores per forecast decreasing the average total contribution score.
- *Decreasing the level of past absolute errors.* Budescu and Chen (2014) found that judges with lower past mistakes are an indicator of higher future performance. They identify top performers by looking for judges with low past absolute errors in conjunction with a positive total contribution score. Therefore, when positive contributors have lower past absolute mistakes, accuracy of the CWM is expected to be higher than when they have higher past absolute mistakes.
- *The number of forecasts.* Budescu and Chen (2014) found that the total contribution score becomes more calibrated when the number of forecasts per judge increases. When the number of forecasts per judge increases, the total contribution score is based on an increasing number of forecasts and becomes more reliable. As a result, the CWM will be based on a more calibrated measure, increasing the model's accuracy.
- *Number of positive contributors per earnings announcement.* Budescu and Chen (2014) found a positive relation between the number of positive contributors and accuracy of the CWM. This is very intuitive. In general, due to the workings of the wisdom of the crowd, when the number of people increases, the accuracy of their aggregate forecast increases. Therefore, when the number of positive contributors per event increases, the accuracy of the CWM is likely to increase.
- *Ratio of positive contributions per analyst.* Budescu and Chen (2014) found that judges with higher percentages of positive contributions were stronger performers than analysts with lower percentages. The ratio of positive contributions per analyst is computed by dividing the number of positive forecasts by the total number of forecasts. Higher ratios indicate that the judge more consistently moves the mean closer to the actual outcome. Therefore, when positive contributors have higher ratios, the CWM is expected to be more accurate than when they have lower ratios.
- *Increasing the level of expertise and training.* Chen et al. (2016) found that an increase in expertise and training increases accuracy of the CWM. Training increases a crowd's expertise. The mean forecast is expected to become more accurate when expertise grows amongst the crowd because each individual judge will be more able to accurately forecast. Hence, it is expected that only top performers will consistently move the mean closer to the actual outcome and will have a positive total contribution score. As the CWM is based on positive contributors, accuracy is expected to increase. Furthermore, they found that training positively enhanced a crowd's expertise.
- *Increasing the level of information available.* Chen et al. (2016) found that when the information environment is richer, the CWM is more accurate. Specifically, they measure information available by pooling judges in teams and allowing them to discuss the problem.

The judges were able to share news, interpretations, and individual forecasts. Furthermore, they found that forecasts closer to the time of resolution are more accurate because more information is available at later stages.

2.2.3 Accuracy drivers from financial literature

Brown et al. (2015) use an extensive survey to investigate how analysts draw up their research reports. According to the analysts' answers, forecasting difficulty varies per company. Some companies have straightforward financials, for instance, consistent earnings backed by operating cash flows. Others have caveats such as weak corporate governance or lacking management communication. Therefore, accuracy of the CWM is expected to vary per company.

The drivers that will be used in this thesis are chosen because they align with the drivers of Budescu and Chen (2014) and Chen et al. (2016). Specifically, most are from Das et al. (1998), Clement (1999), Jacob et al. (1999), and Brown (2001). These highly cited papers look for determinants of forecast accuracy of analysts. Most papers that follow use the same determinants, and, as such, may be considered as generally accepted. These drivers are outlined below.

- *Past performance:* Brown (2001) found that past performance explains future performance well. He compares a model including various analyst characteristics with a model including only past performance. He found both models to perform equally as well. Hence, this supports the use of the CWM, which is based on relative past performance.
- *The level of expertise and training*: Like Chen et al. (2016), financial literature acknowledges the importance of expertise and training for forecast accuracy. Expertise and training can not be measured directly but are proxied for by specialisation and experience. Specialisation is measured by counting the number of companies and industries an analyst follows. When an analyst specialises in a few companies or industries, he is expected to perform better than when he follows a multitude of companies and industries. Larger portfolios are more complex. Also, larger portfolios require more time, making time constraints more likely and thorough analysis less likely (Clement, 1999). Hence, when analysts specialise in industries and companies they are expected to hold higher expertise in these industries and companies. Experience is the improvement of forecast accuracy after affiliation with a company or industry. There are three types of experience. 1) General experience, which is the improvement of forecasting accuracy in general is associated with the knowhow or basic skills to navigate the financial environment. With experience, it is assumed that analysts will learn their way around the environment. 2) Company experience - the improvement of forecasting accuracy for a company. With experience, it is assumed an analyst learns about a company's operations and information sources, and so gains forecast accuracy. 3) Previous forecast errors due to biases and poor judgement can be used to learn what to do better across subsequent forecasts (Jacob et al., 1999). Therefore, experience is like training, learning by practicing.
- *The richness of the information environment*: Analysts acquire information from various sources. The more they can obtain, the more accurate their forecasts will be. According to Brown et al. (2015), the following are most important (in chronological order): knowledge of the industry, private relationship with management, Q&A session of earnings announcements, and management's projection on future earnings. This is like the findings of Chen et al. (2016), who also state that the richness of the information environment is important. It is impossible to measure these drivers directly, therefore, proxies are used in

financial literature. Specifically, size, analyst following, and forecast age. Usually, the bigger firms are, the more accurate analyst forecasts. It is believed that bigger firms tend to generate more public information, as bigger firms are often more closely monitored by investors, news outlets, and the media (Brown, 1998; Das et al., 1998). Furthermore, analyst following is measured by the counting number of analysts following a firm. Analysts seem to complement each other with the provision of private information. The more analysts there are following a firm, the more information is complemented and the more forecast accuracy is achieved (Das et al., 1998; Alford and Berger, 1999). Lastly, forecast age is measured by the number of days between the earnings announcement and the day the forecast was released. With time, more information becomes available. As older forecasts are based on old information and newer forecasts likely on new information, a decrease in forecast age is likely to increase forecast accuracy (Clement, 1999;, Jacob et al., 1999). Furthermore, the veracity of the information available is important too. According to Brown et al. (2015), the quality of reported financials depends on a number of factors such as the use of one-time items or special items (which indicate fraudulent activities) or high operating cash flows (which indicate sustainable business). When quality is higher, forecasts tend to be more accurate.

2.2.4 New information vs analyst biases

There are a few reasons why a forecast can be different from actual earnings. Firstly, analysts are biased. For instance, analysts who are overconfident might draw up optimistic forecasts. This makes their forecasts less accurate. Another reason is new information becoming available during the earnings announcement. For these earnings announcements, the forecasts of analysts were not based on this new information but on old information, presumably making them less accurate. It is impossible to anticipate this. Therefore, this thesis will focus mitigating biases, and not on anticipating new information.

2.2.5 In sum

The CWM will be used to forecast earnings. This is a weighted average of specifically selected forecasts for an earnings announcement based on relative past performance of analysts. By only aggregating forecasts of positive contributors, the model is aimed at reducing biases at the aggregate level. It is expected that it will be more accurate than the analyst consensus, which is based on a simple average of all forecasts for an earnings announcement. The drivers that increase accuracy of the CWM and the drivers that increase accuracy in financial literature will be used to explain the difference between the CWM and analyst consensus. It is important to know if the CWM is more accurate, by what magnitude, and why, because it will be used to predict earnings surprises. Earnings surprises and associated market responses will be explained in paragraph 1.3.

2.3 Earnings surprises and stock market responses

Earnings surprises are measured by actual earnings minus the analyst consensus. Positive earnings surprises denote earnings that are higher than expected and are generally followed by increases in share price, and negative earnings surprises by decreases because earnings are lower than expected. When the CWM is more accurate than the analyst consensus, its forecast will be closer to actual earnings than the analyst consensus. Hence, when the CWM forecast is closer to actual, it will be possible to predict the direction of an earnings surprise by the CWM's forecast minus analyst consensus.

In this paragraph, this will be elaborated on. Furthermore, the market does not always respond in the expected direction. Hence, additional benchmarks are required. These are also explained and will be used in the thesis to predict the stock market response.

2.3.1 Missing, meeting, or beating the expectation of the market and additional benchmarks

The analyst consensus is used throughout to proxy for the market's expectation. Therefore, the analyst consensus could be considered an earnings benchmark that companies should meet to satisfy the market. Note, the expectation of the market, the analyst consensus, and earnings benchmark are used interchangeably. Kasznik and McNichols (2002) investigate the relationship between meeting the earnings benchmark and company share value. They find higher returns for companies that meet the benchmark. Also, meeting the benchmark consistently for consecutive earnings announcements, are rewarded with even higher returns. The market seems to reward companies that are expected to generate higher future earnings. They call this extra return a market premium. On the other hand, companies with inconsistent earnings seem to be penalised. Inconsistency is seen when a company met expectations in the previous month, and fails to so in the subsequent month. Note that meeting the earnings benchmark is the same as a small positive earnings surprise, because the analyst consensus is the benchmark and earnings surprises are measured by differences between actual earnings and the analyst consensus. Likewise, missing the benchmark is a negative earnings surprise. In sum, both meeting the analyst consensus and meeting it consecutively seem to be rewarded with a market premium.

Bartov et al. (2002) confirm this as they found that meeting or beating the earnings benchmark results in firms trading at a market premium. Another finding is that the number of companies meeting the benchmark has increased. This is a sign of earnings management and analyst management. Earnings management is manipulating the books (e.g. taking up large accruals) to meet earnings. Analyst management is tempering the analysts' expectations by carefully selecting what information to give to the analysts. In other words, some firms that trade at a market premium do so because of manipulation rather than expected future performance. Bartov et al. (2002) refer to this as the earnings announcement game, where meeting or beating the consensus is the goal. Despite evidence of manipulation and controlling for it, firms that met the analyst consensus had higher returns compared to firms that did not.

Furthermore, Myers et al. (2007) state that companies with consecutive positive earnings changes are rewarded with higher returns. They argue that companies have incentives to manipulate earnings because they are rewarded with a higher share price. They find that some companies have long positive earnings strings which are impossible if earnings were not manipulated. Furthermore, when the string of positive earnings changes ends, the market responds extra negatively and sharp declines in share price are observed. They measured strings of positive earnings by comparing earnings announced in the current quarter to earnings announced in previous quarters. Similarly, Rees (2005) put forward that when firms meet two thresholds, a positive earnings response is more likely. The first threshold is the common analyst consensus, the second is looking at the earnings change compared to the previous quarter. When current quarter's EPS overstates last quarter's EPS, the market seems to respond positively. In other words, having higher earnings compared to previous periods consecutively, in conjunction with meeting the earnings benchmark, seems to be rewarded with a positive market response.

Furthermore, Bhojraj et al. (2009), find that there are short-term benefits for meeting the analyst consensus versus not meeting it. However, it can be to the detriment of long term growth when firms are willing to give up long-term projects to meet short-term market demands. Still, they find that the market does not always recognise it when earnings are manipulated and even responds positively to positive earnings surprises when there are signs of manipulation. They test this by using proxies for earnings quality such as, for instance, accruals. When a company has high accruals, earnings manipulation is likely.

On the contrary, Kinney et al. (2002) find that in 43% of earnings announcements, meeting the analyst consensus results in negative share price returns. This means that for 43% of earnings announcements, the market responds in the unexpected direction. This is especially so for firms that just meet the earnings benchmark (i.e. small positive earnings surprises). Kinney et al. (2002) argue that meeting the earnings benchmark by a small margin is a sign of earnings manipulation and therefore is rewarded less. Also, Gleason and Mills (2008) state that meeting the expectation of the market with a small margin is a sign of low earnings quality.

In other words, besides meeting the analyst consensus, meeting it consistently, and having persistent earnings growth, and despite that earnings manipulation is not always recognised, earnings quality does seem to have an impact on the response of the market.

2.3.3 In sum

The markets' response to missing, meeting, or beating the analyst consensus. Sometimes, the market response to these earnings surprises is opposite to expected. Therefore, additional benchmarks have been imposed. These are: 1) consecutive meeting or beating of the benchmark, 2) consecutive increases in earnings. It is expected that when multiple conditions are met, the earnings surprise and market response will be in the same direction more often than when only the expectation of the market is met. Still, when earnings quality measured by the ERC is low, the markets' response is expected to be low.

2.4 Other factors influencing stock market responses

There are many more factors that influence stock market responses, however, it goes beyond the scope of this thesis to include them. Therefore, in no means a 100% correlation is expected between meeting the benchmarks mentioned in paragraph 1.3 and market responses. Some of these other factors are:

- Investor inattention: Sometimes the markets' response to earnings announcements is slow in terms of trading volume and market price reaction during earnings announcements. The response follows in subsequent weeks and is named the post-announcement drift. Hirshleifer et al. (2009) state that the market underreacts due to inattention.
- Investor sentiment: Theories of investor sentiment explain the phenomenom that some companies' stock is trading above fundamental value. For instance, Mian and Sankaraguruswamy (2012), find that the market responds to earnings announcement more pronounced during periods with higher sentiment.

- Meeting revenue targets: Rees and Sivaramakrishnan (2007) find that meeting revenue targets is more important than earnings targets for high growth firms.

3 Hypotheses

The introduction and the theoretical framework have set the stage of this thesis. In this paragraph, the exact problem, purpose, and hypotheses will be further delineated.

3.1 The problem

The problem is that analysts have a prominent role in the financial environment, but their forecasts of quarterly company earnings are often incorrect. Many stakeholders rely on analyst reports and specifically on the average of all analysts' earnings forecast. The average of analyst earnings forecasts is also called the analyst consensus, and is, among others, used by investors and researchers. The wisdom of the crowd theory postulates that the average of analyst forecasts is more accurate than each individual forecast over a large number of forecasts. Therefore, the aggregate should be a reliable measure. However, the analyst consensus is, just as the individual analyst forecasts, often incorrect. This can, among other things, be attributed to conflicts of interest and biases amongst analysts, which make the conditions on which wisdom of the crowd thrives unsatisfied. One such condition is that of independence. It is found, that analysts who are constrained by time, herd. Therefore, some analyst forecasts are dependent on others. This makes the individual forecast more accurate, but the consensus forecast less accurate.

There are many stakeholders that are influenced. Investors use analyst reports and the analyst consensus as guide to form expectations of future quarterly earnings. When the analyst consensus differs a lot from actual earnings, the market usually responds by increased buying or selling shares. This results in increases and decreases of share prices after company earnings announcements. For some investors this is a problem because a large shift in share price diminishes their returns. For others this is a blessing because they were able to foresee the markets' response. Consequently, companies are affected as well. When companies need extra financial resources they often issue new shares. Hence, share price is very important to them. Large decreases in share price might mean that they are not able to issue shares and have to delay investment projects. Conversely, large increases might mean they are able to pursue more projects. Therefore, companies manage their earnings to meet and satisfy earnings targets which are formed by the analyst consensus. Lastly, researchers use the analyst consensus as a proxy for the market's expectation on a company's future earnings. Research involves the information content of earnings announcements and market inefficiencies. A poor benchmark leads to biased results.

3.2 The purpose

Undeniably, all stakeholders are in the position to be financially hit by analyst consensus misses. However, opportunities arise as well. In the Netherlands we have a saying, every death is someone else's bread. By which is meant, when one person loses another wins. This thesis will focus on how to win by overcoming the problem of a biased analyst consensus. Specifically, by coming up with a more accurate and consistent measure than the analyst consensus and so predict earnings surprises, which can then be used to predict stock market responses.

A new crowd aggregation technique proposed by Budescu and Chen (2014) will be used to create a more accurate and persistent measure than the analyst consensus. It is chosen because it is very adaptable to the dataset of this thesis and for the reasons put forward above. Other techniques were found less adaptable or less appropriate. As in previous paragraphs, Budescu and Chen (2014) call the new technique the contribution weighted model (CWM). The CWM is a new model and has only been tested in a few environments of which the financial analyst environment is not one. Therefore, the objectives of this thesis are: 1) to test the model of Budescu and Chen (2014) in a new environment, and 2) determine whether the model can be used to predict stock market responses.

The following research question will be answered:

To what extent is it possible to more accurately forecast earnings than the analyst consensus using the contribution weighted model of Budescu and Chen (2014) and to use the difference to predict stock market responses?

To answer the research question, two hypotheses are tested.

3.3 Hypothesis 1

The objective of the first hypothesis is to test the CWM of Budescu and Chen (2014) in a new environment. It focusses on if the model is more accurate than the analyst consensus and why it is more accurate. Especially, H1a is the fundament of this thesis. The answer to this hypothesis needs to be statistically significant and practically significant. Statistical significance will validate the model in a new environment. Practical significance will validate the model to be used in practice by investors. It has relevance in practice when it predicts the size and direction of earnings surprises correctly and consistently.

The analyst consensus is based on a simple average and is often biased. Contrary, the contribution weighted model of Budescu and Chen (2014) is expected to provide debiased forecasts, as it attempts to identify superior analysts by their total contribution score. They state that when only forecasts of these 'experts' are aggregated, it is possible to get rid of biases (e.g. herding) at the aggregate level, and thus obtain more accurate results. Therefore, it is expected that the CWM is more accurate than a biased analyst consensus.

H1a: The CWM of Budescu and Chen (2014) more accurately forecasts future earnings than the analyst consensus.

To test the validity and robustness of the original model, it is compared to two other versions. One is based on a different weighting scheme and is from Budescu and Chen (2014). It uses a simple average to aggregate forecasts of positive contributors, rather than weighting forecasts of positive contributors by their total contribution score. By testing the difference between this contribution average model (CAM) and the CWM, it will be possible to identify the ability of the CWM to weight analyst forecasts. The other version is more novel as it has not directly been tested by Budescu and Chen (2014). Budescu and Chen (2014) find that among the positive contributors, some are most superior. They identify these analysts by two requirements: a positive total contribution score and a lower level of absolute mistake than the analyst consensus. Analysts that satisfy these requirements are superior performers. Essentially, they are the experts amongst the experts. Rather than just identifying these superior performers as in Budescu and Chen (2014), I will test whether aggregating these superior performers increases accuracy compared to the CWM. The model including the extra condition will be called CWM(X). By testing for this difference, the ability of the CWM to be refined, will be investigated.

H1b: The CWM of Budescu and Chen (2014) more accurately forecasts future earnings than the CAM.

H1c: The CWM(X) more accurately forecasts future earnings than the CWM.

Next, to gain understanding of why the CWM is more accurate, hypothesis 1d to 1j are imposed. The focus is on accuracy drivers of the CWM that have been identified by Budescu and Chen (2014) and Chen et al. (2016). For reference of the accuracy drivers, see paragraph 2.2.2 and 2.2.3 of the literature review. Understanding what drives accuracy and when the CWM is more accurate than the analyst consensus, is valuable information for investors and validates the use of the CWM in practice. The novelty is that the drivers will be tested in conjunction, rather than in isolation as in Budescu and Chen (2014) and Chen et al. (2016). The benefits of testing the accuracy drivers simultaneously is intuitive. The drivers are likely to interact dynamically, hence, changes in variables can be investigated while keeping others constant.

Each of the accuracy drivers identified by Budescu and Chen (2014) and Chen et al. (2016) is expected to have an effect on the accuracy of the CWM compared to the analyst consensus. The goal is to characterise the difference in accuracy in detail. Every driver that is significantly related, increases the level of detail by which the difference can be described. In other words, the aim of this hypothesis is to identify which model characteristics attribute most to the accuracy of the CWM in the financial analyst environment. Each hypothesis resembles a characteristic and they will all be tested in conjunction.

H1d: The level of total contribution score is positively related to accuracy of the contribution weighted model compared to the analyst consensus, ceteris paribus.

H1e: The level of past absolute mistakes is negatively related to accuracy of the contribution weighted model compared to the analyst consensus, ceteris paribus.

H1f: The number of forecasts is positively related to accuracy of the CWM compared to the analyst consensus, ceteris paribus.

H1g: The number of positive and negative contributors per earnings announcement is positively related to accuracy of the CWM compared to the analyst consensus, ceteris paribus.

H1h: The ratio of positive contributions per analyst is positively related to accuracy of the CWM compared to the analyst consensus, ceteris paribus.

H1i: The level of expertise and training is positively related to accuracy of the CWM compared to the analyst consensus, ceteris paribus.

H1j: The level of information available is positively related to accuracy of the CWM compared to the analyst consensus, ceteris paribus.

To reiterate, hypothesis 1a is the most important hypothesis. It tests if and by how much the CWM is more accurate than the analyst consensus. It is the fundament of this thesis. If no significant difference in accuracy is found, it won’t be relevant to test the other hypotheses. The other hypotheses are established for robustness and practical significance of the model.

3.4 Hypothesis 2

Stock prices tend to increase or decrease after positive or negative earnings surprises. Hypothesis 2 is established to test whether the difference between the forecast of the CWM and analyst consensus (i.e. predicted earnings surprise) can be used to predict stock market responses. Two things are important: 1) how well does a predicted earnings surprise proxy for actual earnings surprises, and 2) how well does the market respond in the same direction as predicted earnings surprises.

A predicted earnings surprises is a good proxy when it is strongly and positively related to actual earnings surprises. This is very intuitive. For example, when for company X at a point in time, the predicted earnings surprise is negative, a negative market response is expected. When the actual earnings surprise for the same company at the same point in time is positive, it is likely that the market will respond positively, which is opposite of predicted. In other words, when the predicted earnings surprise is not accurately and consistently in the same direction as actual earnings surprises, it won't be a strong proxy.

H2a: A positive relationship exists between predicted earnings surprises, using the CWM of Budescu and Chen (2014), and actual earnings surprises.

Next, once the proxy is established, it is important to test how well the market responds in the same direction as the proxy. For example, the CWM has more accurate forecasts than the analyst consensus for 80% of earnings announcements of company X. Future earnings of company Y will be announced in a week. The CWM's forecast is \$15, and of the analyst consensus is 10\$. In this case, a positive earnings surprise is expected (\$15 - \$10 = \$5) for 80% of the earnings announcements. Hence, when a strong and positive relation exists between predicted earnings surprises and stock market responses, a positive stock market response and increase in share price is expected for about 80% of the earnings announcements.

H2b: A positive relationship exists between predicted earnings surprises, using the CWM of Budescu and Chen (2014), and stock market responses.

To conclude, hypothesis 2a is included to see whether the proxy does what it is supposed to do; predict an actual earnings surprise. Regardless of whether the market will respond in the hypothesised direction, a significant result would validate the use of the CWM. Next, the intention is to find market responses after predicted earnings surprises. This depends on how well an actual earnings surprise is followed by a stock market response in the same direction.

4. Methodology and results

In this paragraph, methodology and results of the hypotheses are covered. Firstly, the technicalities of the contribution weighted model (CWM) are given. Here, formulas and definitions are stated that are used throughout the thesis. Secondly, an explanation of the sample is provided. It entails things such as where the data is from, conditions imposed, and descriptive statistics. Next, the methodology of hypotheses 1 are covered, and immediately following the results of hypothesis 1. Subsequently, the methodology and results of hypothesis 2 are stated. For both hypotheses, the following are provided in chronological order: the method of testing, descriptive statistics, and results.

4.1 The contribution weighted model (CWM)

In this subparagraph the model will be elaborated on in detail. For ease, a legend is created and can be found in appendix A. In all the calculations earnings will be used rather than EPS. Companies are of varying sizes with different numbers of shares outstanding. Therefore, EPS is normalised by the number of shares outstanding.

4.1.1 Assigning contribution scores to forecasts

Analyst (j)'s earnings forecast (p_{jit}) either moves the mean forecast of all analysts (i.e. analyst consensus) closer to or further away from actual earnings. A score is assigned to the forecast based on how contributing it is and is derived via the following steps.

First, the absolute difference between actual earnings (O_{it}) and the analyst consensus (M_{it}) including analyst (j) for earnings announcement (i) at time (t) is computed. Time (t) is measured in quarters. Note, notwithstanding that companies have quarterly announcements, they do not all have the same quarter-end months. For instance, Microsoft's quarters end in March, June, September, December. Finisar's are in January, April, July, October. To sum, the absolute difference (S_{it}) is computed by:

$$S_{it} = |O_{it} - M_{it}|$$

Secondly, the absolute difference (${S_{it}}^{-j}$) between actual earnings (O_{it}) and the analyst consensus without analyst (j) (${M_{it}}^{-j}$) is calculated by:

$${S_{it}}^{-j} = \left|O_{it} - {M_{it}}^{-j}\right|$$

Lastly, by subtracting both absolute differences, S_{it} and ${S_{it}}^{-j}$, the contribution score (C_{jit}) of analyst (j)'s forecast for earnings announcement (i) at time (t) is derived:

$$C_{jit} = {S_{it}}^{-j} - S_{it}$$

When C_{jit} is positive, the analyst consensus (S_{it}) was closer to actual earnings with analyst (j)'s forecast (i.e. smaller) than the analyst consensus ${S_{it}}^{-j}$ without analyst (j)'s forecast (i.e. bigger). And when negative, was further away from actual earnings with analyst (j) than without.

4.1.2 Computing total contribution scores for analysts

Subsequently, the total contribution score ($\overline{C_{Jt}}$) is calculated. This is the mean contribution score of all forecasts of analyst (j) up to time (t). Analyst (j) does not give a forecast for every earnings announcement. Only earnings announcements for which analyst (j) gave a forecast needs to be included in the calculations. Therefore:

$N_t = \{earnings\ announcements\ (i)\ up\ to\ time\ (t)\}$

$N_{jt} = \{earnings\ announcements\ (i)\ \in\ N_t : analyst\ (j)\ gave\ a\ forecast\}$

$$\overline{C_{Jt}} = \frac{1}{N_{jt}} \sum_{i\,\in\,N_{jt}} C_{jit}$$

$\overline{C_{Jt}}$ is computed in such a way that it dynamically changes through time for each individual analyst. For example, analyst (j) has made 100 forecasts for 100 different companies' earnings announcements up to 31 December 2015 (N_{jt}), and the sum of analyst (j)'s contributions (C_{jit}) is a positive 5000. Then $\overline{C_{Jt}}$ is 50 (5000 / 100). However, when the same analyst has made 120 forecasts for earning announcements up to 31 March 2016 (N_{jt}), and the sum of his contributions (C_{jit}) is a negative -6000. Then $\overline{C_{Jt}}$ is -50 (-6000/120).

4.1.3 Aggregating positive contributors

For each earnings announcement, the total contribution scores ($\overline{C_{Jt}}$) of analysts decide which forecasts are aggregated. Budescu and Chen (2014) argue to only aggregate forecasts of analysts when their $\overline{C_{Jt}} > 0$. Hence, this will be adopted. Whenever, $\overline{C_{Jt}}$ exceeds 0, analysts are considered positive contributors. And when it is lower than 0, analysts are considered negative contributors. By aggregating the forecasts of positive contributors, a new group forecast is formed (P_{it}). P_{it} is the consensus of positive contributors of earnings announcement (i) at time (t). Budescu and Chen (2014) impose two methods to aggregate positive forecasters by assigning weights (ω_{jit}) to analysts (j) for earnings announcement (i) at time (t):

$J_i = set\ of\ analysts\ (j)\ who\ gave\ a\ forecast\ on\ earnings\ announcement\ (i)$

$J_i^+ = subset\ of\ J_i\ such\ that\ \overline{C_{Jt}} > 0$

The first method is a simple mean with equal weights (ω_{jit}):

$$\omega_{jit} = \frac{1}{size(J_i^+)}$$

The second method is a weighted average based on $\overline{C_{Jt}}$ of each positive contributor:

$$\omega_{jit} = \frac{\overline{C_{Jt}}}{\overline{\sum_{h \in J_i^+} C_{Jt}}} \qquad if \quad j \in J_i^+$$

P_{it} of next earnings announcement is derived by:

$$P_{i(t+1)} = \sum_{j\,\in\,J_i} p_{ji(t+1)} * \omega_{jit}$$

Where $\omega_{jit} = 0 \qquad if \quad j \notin J_i^+$

The model with equal weights (ω_{jit}) will be referred to as the contribution average model (CAM), and the model with varying weights (ω_{jit}) as the contribution weighted model (CWM). In sections below, the accuracy of CAM and CWM will be compared with the accuracy of the analyst consensus (M_{it}). The analyst consensus will be referred to as AC.

4.1.4 Measuring accuracy

Accuracy is measured by the size of mistakes of each model. When CWM's mistake is smaller than AC's, CWM is more accurate than AC. The mistake (X_{it}) of the CWM and CAM is derived by the absolute difference between P_{it} and actual earnings:

$$X_{it} = |O_{it} - P_{it}|$$

Lower values of X_{it} indicate higher accuracy. Throughout this thesis the same notation for mistakes (X_{it}) will be used for CWM and CAM but specified to each model. Hence, X_{it} becomes:

$X_{it(CWM)} = \left|O_{it} - P_{it(CWM)}\right|$ and $X_{it(CAM)} = \left|O_{it} - P_{it(CAM)}\right|$.

Note that an analysts' individual absolute mistake is measured by:

$$X_{jit} = \left|O_{it} - p_{jit}\right|$$

Furthermore, Budescu and Chen (2014), identify top performers by looking at analyst (j)'s level of $\overline{C_{Jt}}$, and level of past average absolute mistake ($\overline{X_{Jt}}$):

$$\overline{X_{Jt}} = \frac{1}{N_{jt}} \sum_{i \in N_{jt}} p_{jit}$$

Budescu and Chen (2014) state that analysts with $\overline{X_{Jt}}$ below a baseline are considered top performers. In this thesis, the analyst consensus will be used as baseline. Specifically, analysts that have a lower $\overline{X_{Jt}}$ than the mean of all analysts for earnings announcement (i) at time (t), in conjunction with a positive $\overline{C_{Jt}}$ will be considered top performers. The model where this condition is imposed will be called CWM(X).

4.1.5 Measuring earnings surprises and market responses

Lastly, the goal of this thesis is to predict earnings surprises and market responses using the forecast of the CWM. Actual earnings surprises are measured by the difference between actual earnings and the analyst consensus:

$$S_{it} = O_{it} - M_{it}.$$

CWM's predicted earnings surprises will be measured by the difference between the group forecast of CWM and the analyst consensus on next periods' earnings:

$$ES_{i(t+1)} = P_{i(t+1)} - M_{i(t+1)}.$$

The measures are the same as used in other literature [e.g. Bartov et al. (2002), Skinner and Sloan (2002)].

The response of the market to predicted earnings surprises is measured in terms of abnormal return AR_{fit} of company (f) for earnings announcement (i) at time (t). To calculate it, the share price of today (d), the share price of yesterday (d-1), and market return (MR_d) of today (d) are used:

$$AR_{fit} = \left[\left(\frac{Share\ price_{fd}}{Share\ price_{f(d-1)}}\right) - 1\right] - MR_d$$

A difference with most other literature is the use of MR_d. Other literature generaly use carefully selected industry portfolios to calculate AR_{fit}. It goes beyond the scope of this thesis to create such portfolios. I found that some other literature used the market return (Athanasakou and Walker, 2011), therefore, it was used for this thesis as well. Next, the cumulative abnormal return (CAR_{fit}) of a three day period is calculated. This is related to a day before the earnings announcement (day-1), the day of the earnings announcement (day0), and the day after the earnings announcement (day+1). It is in line with other literature (e.g. Alford and Berger (1999), and Lim (2001)). The cumulative abnormal return (CAR_{fit}) denotes the markets' response to earnings surprises:

$$CAR_{fit} = AR_{fit(day-1)} + AR_{fit(day0)} + AR_{fit(day+1)}.$$

Furthermore, some earnings announcements are done on a Monday or Friday, which means that day-1 or day+1 do not have a value (market is closed on weekends). For these earnings announcements, the previous business day is used. When earnings are announced on a Friday $AR_{fit(day+3)}$ is used and earnings announcements on a Monday $AR_{fit(day-3)}$. If also day -3 and day +3 did not have a value (e.g. public holiday), day -4 and day +4 were used (Alford and Berger (1999)).

4.2 Sample

4.2.1 Data

To test the hypotheses, data is extracted from the Institutional Brokers' Estimate System (IBES), The Center for Research in Security Prices (CRSP) and Compustat. These databases are provided by the Wharton Research Data Services (WRDS) of Wharton University of Pennsylvania.

IBES is used to gather data concerning earnings announcements, including: analyst EPS forecasts per earnings announcement on an individual forecast level, analyst EPS consensus per earnings announcement, actual EPS per earnings announcement, and total number of shares outstanding. CRSP is used to obtain company share prices of the day before an earnings announcement, of the day itself, and the day after. Compustat is used to assign Global Industry Classification Standard (CIGS) to companies in the sample.

The original sample consists of 1082683 forecasts of quarterly earnings announcements from January 2011 to December 2017, and only companies of the United States of America (United States) are included. United States has the most sophisticated accounting regulation and information sharing system, making the data more reliable and combining data sources more available. Also, this rules out cross-country specific factors (e.g. accounting regulation) which potentially influence the results.

Not all observations could be used due to missing information, an inoperable date on which the forecast was given, analysts having multiple forecasts for the same earnings announcement, missing identifiers to link data sources, and analysts having too few forecasts. Dropping these observations reduced the sample size to 289814 forecasts. All drops are extensively explained in appendix B.

Furthermore, two conditions are imposed that hold throughout the whole thesis. They do not cause for drops in the number of forecasts but do determine whether a forecast is included in the calculations of the total contribution score ($\overline{C_{Jt}}$) and whether an earnings announcement is used in the comparison between the CWM, CAM, and AC. Condition 1 is that each analyst up to time (t), should have a minimum of 10 forecasts. Condition 2 is that a group forecast [$P_{i(t+1)}$] for earnings announcement (i) at time (t) should be based on at least 5 positive contributors. Over the whole sample period, after imposing these conditions, 1605 earnings announcements are not used in comparisons between the CWM, CAM, and AC. In appendix C, both conditions are elaborated on.

4.2.2 Descriptive statistics sample

After making the above changes, the remaining sample is used to compute CWM. In total, there are 289814 forecasts remaining, 17019 earnings announcements, 1146 companies, and 4510 analysts. In table 1 descriptive statistics are given for the total period, January 2011 to December 2017. Note that the total sample holds 17019 earnings announcements, and after imposing the conditions it holds 15414. The conditions do not drop forecasts from the database, and only have an effect on whether forecasts are used in the calculation of the total contribution score $\overline{C_{Jt}}$. Therefore, I do not refrain to 15414 earnings announcements, because relevant information, for instance the total number of forecasts of analysts, would then be excluded. Similarly, the fact that the minimum number of forecasts for one analyst is 1 is stated, following the same line of reasoning.

The number of forecasts per analysts differs. For instance, one analyst has 1 forecast, whereas another has 698 forecasts. On average, analysts have 64 forecasts. The main reason for this difference is that some analysts started forecasting later in the sample period. Furthermore, forecasts are given, on average, 62 days before the earnings announcement with a standard deviation of 34 days.

The number of forecasts is spread evenly across quarters, with a difference between the first and third quartile of 946 (Q3: 10864 – Q1: 9918). Accordingly, also the number of earnings announcements, the number of companies with a forecast, and the number of analysts are well balanced across time. In total, there are 29 quarters.

On the contrary, differences in the number of forecasts per industry exist. This is mainly because the number of companies per industry differs. For example, one industry includes 1 company, whereas another has 117. As the number of companies differs, earnings announcements and analysts also differ across industries. In total, there are 68 industries.

Analysts, on average, follow 29 companies throughout the sample period. Also, on average, there are 17 analysts per earnings announcement, with a minimum of 10 and a maximum of 59. This difference is mainly driven by firm size.

Lastly, earnings announcement per quarter and companies per quarter are the same because companies have one earnings announcement per quarter.

Table 1. Descriptive statistics full sample, January 2011 - December 2017

	Obs	Mean	Std. Dev.	Min	Quartile 1	Median	Quartile 3	Max
Forecasts								
Forecasts	289814							
Forecasts per EA	289814	17	6	10	12	15	20	59
Forecasts per analyst	289814	64	78	1	7	29	99	698
Forecasts per company	289814	253	218	10	67	195	405	1199
Forecasts per industry	289814	3331	3872	69	389	1041	8309	10839
Forecasts per quarter	289814	10316	770	8831	9918	10330	10864	12367
Forecast age	289814	62	34	1	25	77	90	120
EA								
Earnings announcements (EA)	17019							
Earnings announcements per industry	17019	250	270	3	63	160	371	1443
Earnings announcements per quarter	17019	606	37	526	581	614	635	667
Company								
Companies	1146							
Companies per industry	1146	17	19	1	6	11	23	117
Companies per quarter	1146	606	37	526	581	614	635	667
Market capitalisation (x1000)	1146	\$25091	\$68301	\$22	\$2204	\$5948	\$21017	\$1083402
Analysts								
Analysts	4510							
Analysts per EA	4510	17	6	10	12	15	20	59
Analysts per company	4510	29	15	8	17	26	37	103
Analysts per industry	4510	142	117	15	53	103	200	553
Analysts per quarter	4510	2806	122	2431	2782	2851	2884	2966
Industry & periods								
Industries	68							
Quarters	29							

4.2.3 Descriptive statistics C_{jit} and $\overline{C_{Jt}}$

The contribution score per forecast (C_{jit}) and total contribution score per analyst ($\overline{C_{Jt}}$) are important concepts because they determine on which forecasts the CWM is based. Therefore, it is important to understand the distribution of their datapoints. In table 2, descriptive statistics are given. For each forecast C_{jit} is computed and for each analyst $\overline{C_{Jt}}$. Hence, C_{jit} has 289814 observations and $\overline{C_{Jt}}$ 4510. Also, the percentage of positive C_{jit} per analyst and the percentage of positive $\overline{C_{Jt}}$ per earnings announcement are depicted.

Both C_{jit} and $\overline{C_{Jt}}$ are shifted to the right as their skewness and kurtosis are both highly positive. On both sides of the spectrum, positive and negative, outliers exist. However, the positive outliers are up to 7 times larger. This is further exemplified by the boxplots in figure 2, in which C denotes C_{jit} and total C is $\overline{C_{Jt}}$. Also, half of the analysts are assigned a positive C_{jit} for 55% of their forecasts. A few analysts had 100% positive C_{jit}'s. The reason for this high percentage is that it is based on one or a few forecasts, these analysts started giving forecasts during the sample period.

Furthermore, for each earnings announcement the median of analysts that have a positive $\overline{C_{Jt}}$ is 82% (i.e. positive contributors). In the CWM and CAM, analyst forecasts are aggregated when $\overline{C_{Jt}}$ is positive. Hence, this means about 82% of all forecasts per earnings announcements are aggregated. This could be caused by the high, especially positive, outliers for C_{jit} and could give serious implications for the accuracy of the CWM and CAM. If all analyst forecasts for earnings announcement (i) were to be aggregated, the CWM is very similar, and CAM identical to the AC. In other words, throughout this thesis the accuracy of the CWM and CAM model might be very similar to the accuracy of the AC.

Likely, one of the main drivers of these outliers of C_{jit} is the difference in size of companies. For instance, C_{jit} is higher when analyst (j) forecasts 5.000.000.000 earnings, the analyst consensus is 6.000.000.000 for company K, and actual earnings is 6.500.000.000, then when analyst (j) forecasts 5.000.000 earnings, the analyst consensus is 6.000.000, and actual earnings is 6.500.000 for company L.

The outliers for C_{jit} also seem to induce another implication. A large variation in $\overline{C_{Jt}}$. Forecasters with a high $\overline{C_{Jt}}$ compared to other forecasters for earnings announcement (i) at time (t), are given excessive weights (ω_{jit}). For these earnings announcements, the forecast is for a large part based on one forecast. This is likely to be to the detriment of accuracy.

Different truncation levels have been tested to see if mitigating the outliers of C_{jit} and $\overline{C_{Jt}}$ helps to reduce the high median of positive analysts per earnings announcements of 82%, and to see what the effect is of outliers on accuracy of the CWM and CAM. Truncating did not have an effect. The 82% did not change much. Similarly, accuracy of the CWM and CAM did not differ. However, that does not alleviate the issue. An elaborate depiction of the truncation levels can be found in appendix D.

Table 2. Descriptive statistics C_{jit} and $\overline{C_{jt}}$: outliers and high median positive contributors

	Obs	Mean	Std. Dev.	Min	Quartile 1	Median	Quartile 3	Max	Skewness	Kurtosis
C_{jit} per analyst	289814	723	14335	**-566253**	-93	37	483	**4177551**	166	45879
$\overline{C_{jt}}$ per analyst	4510	693	6777	**-79573**	0	162	672	**414581**	50	3091
Percentage $C_{jit} > 0$ per analyst	4510	0.54	0.21	0.00	0.45	0.55	0.65	1.00	-0.20	1.10
Percentage $\overline{C_{jt}} > 0$ per EA	17019	0.78	0.19	0.00	0.69	**0.82**	0.92	1.00	-1.61	3.52

C_{jit} denotes contribution per forecast. $\overline{C_{jt}}$ denotes total positive contribution. The percentage $C_{jit} > 0$ per analyst derived by: Number of forecasts of analyst (j) for which ($C_{jit} > 0$) / Number of all forecasts of analyst (j). The percentage $\overline{C_{jt}} > 0$ per EA derived by: Number of analysts for which [$\overline{C_{jt}} > 0$] for earnings announcement (i) / Number of all analysts (j) for earnings announcement (i).

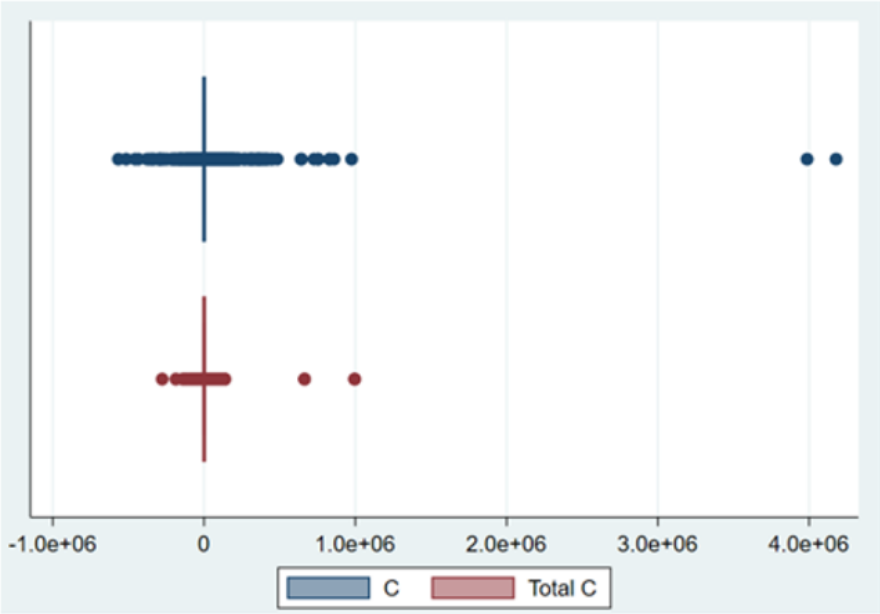

Figure 1. Boxplots of C_{jit} and $\overline{C_{Jt}}$

4.3 Methodology hypothesis 1

Hypothesis 1 is established to test if the contribution weighted model is more accurate than the analyst consensus, and why it is more accurate. First, for hypothesis 1a, 1b, and 1c (i.e. if), the test methods used, and descriptive statistics are given. Then, for hypothesis 1d (i.e. why), also the methods used, and descriptive statistics are stated.

4.3.1 Testing hypotheses 1a, 1b, and 1c

Hypothesis 1a is that the CWM more accurately forecasts future earnings than the AC. And hypothesis 1b and 1c that the CWM is more accurate than the CAM and CWM(X).

Accuracy is compared by the level of mistake per model for earnings announcement (i) at time (t). Mistake of the analyst consensus is defined by: $S_{it} = |O_{it} - M_{it}|$, and mistake of the other models by: $X_{it} = |O_{it} - P_{it}|$. Therefore, comparison of the CWM with the AC entails: $X_{it(CWM)} - S_{it}$, and comparison of the CWM with the CAM and CWM(X) entails: $X_{it(CWM)} - X_{it(CAM)}$ and $X_{it(CWM)} - X_{it[CWM(X)]}$. Note that negative values indicate that the CWM is more accurate than the other models.

It will be tested with a Paired t-test, Wilcoxon signed-rank test, and sign test. The Paired t-test is a parametric test and is most suitable when all parametric assumptions are satisfied. However, given that C_{jit} and $\overline{C_{Jt}}$ showcased non-normal behaviour, the paired difference in mistakes between the CWM and AC, CAM, and CWM(X) might also be non-normally distributed. The boxplot and histogram in figure 2 depict the distribution of the paired difference in mistakes between the CWM and AC. In appendix E, boxplots of the difference between the CWM and CAM and CWM(X) are given, but do not look different. For this test, I assume that the normal distribution requirement is satisfied. In any case, the Wilcoxon signed-rank test will be used for robustness, in case non-normalisation is not satisfied. Similarly, the Wilcoxon signed-rank test requires that, under the null hypothesis, the distribution of the paired difference in mistakes between the models is symmetric. Therefore, the sign test will be used for further robustness.

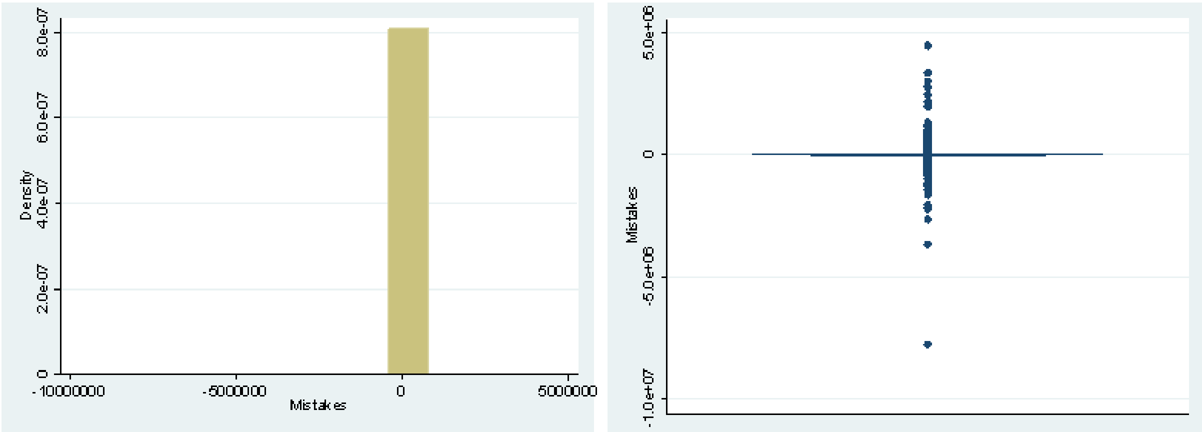

Figure 2. Histogram and boxplot of $X_{it(CWM)} - S_{it}$

4.3.2 Descriptive statistics hypotheses 1a, 1b, and 1c

In table 6 the CWM, and the AC, CAM, and CWM(X) are compared on level of absolute mistakes. This gives an indication of the potential significant difference between the models using the Paired t-test. Also, the number of earnings announcements the CWM is more accurate than the other models is given, the number of times the other model is more accurate than the CWM, and the number of earnings announcements they have an equal level of mistakes. This gives an indication for the Wilcoxon signed-rank test and sign-test. Lastly, absolute mistakes are in dollars. For instance, on average, the absolute difference between the forecast of the CWM and actual earnings per earnings announcement (i) at time (t) is $52005. Notice that CWM(X) has a lower number of earnings announcements of 12815. This is due to the extra imposed condition (i.e. requiring analysts to have an $\overline{X_{jt}}$ below the mean of all analysts for earnings announcement (i) at time (t)).

The CWM seems more accurate than the AC, which can be derived from the smaller absolute mistake per earnings announcement, and the higher number of earnings announcements that the CWM is more accurate. To be more precise, 10.65% [(52005–58201 / 58201)*100] more accurate in terms of mean difference in absolute mistake and 11.06% [(8553–6848 / 15414)*100] more accurate in terms of number of earnings announcements.

The CWM and CAM do not seem to differ much. However, the CAM is based on the same composition of analysts as the CWM and only uses a different weighting scheme. Therefore, it is possible to compute how much of the accuracy increase of the CWM over the AC is due to the CWM's ability to pick experts and to the ability to weight analyst forecasts. -0.42% of 10.65% [(51784–52005 / 52005)*100] and 1.44% of 11.06% [(7810–7588 / 15414)*100] seems to be due to weighting. It seems that the CAM is slightly more efficient in weighting analyst forecasts in terms of mean absolute mistake. Moreover, the rest of the increase of 10.65% and 11.06% of accuracy is likely due to the ability of the CWM to pick experts.

Lastly, the CWM(X) has a larger absolute mistake than the CWM. This could be due to the difference in subset of earnings announcements between the CWM and CWM(X), because other than that, the

CWM and CWM(X) seem to very similar. Imposing the extra condition even seems to slightly decrease the accuracy (i.e. 6355 > 6215 number of earnings announcements).

Table 3. Comparison of mistakes of CWM and AC, CAM, and CAM(X)

	CWM	AC	CAM	CWM(X)
Mean absolute mistake	$ 52,005	$ 58,201	$ 51,784	$ 55,413
Median absolute mistake	$ 7,680	$ 8,666	$ 7,646	$ 8,205
SD absolute mistake	$ 481,423	$ 547,176	$ 487,098	$ 499,098
Minimum absolute mistake	$ -	$ -	$ -	$ -
Maximum absolute mistake	$ 52,144,789	$ 59,957,696	$ 53,509,171	$ 49,674,746
Number of earnings announcements	$ 15,414	$ 15,414	$ 15,414	$ 12,815
N of EA: CWM more accurate	N/A	8553	7810	6355
N of EA: Other more accurate	N/A	6848	7588	6215
N of EA: CWM and Other equal	N/A	13	16	245

4.3.3 Testing hypotheses 1d to 1j

Hypothesis 1d is established to test if the various accuracy drivers of the CWM and financial literature explain the difference in accuracy between the CWM and AC. Understanding why the model works will increase the validation of the model in the financial analyst environment. For reference of the accuracy drivers, see paragraph 2.2.2 and 2.2.3 of the literature review. The hypothesis entails an ordinary least squares (OLS) model:

$$\begin{aligned} CWM - AC_{it} = {} & \beta_0 + \beta_1 MeanC_{it} + \beta_2 PastAccuracy_{it} + \beta_3 NForecast_{it} + \beta_4 PosConEA_{it} \\ & + \beta_5 NegConEA_{it} + \beta_6 ContributionsJ_{it} + \beta_7 GenExperience_{it} \\ & + \beta_8 ComExperience_{it} + \beta_9 IndExpertise_{it} + \beta_{10} ComExpertise_{it} + \beta_{11} Size_{it} \\ & + \beta_{12} Following_{it} + \beta_{13} Age_{it} + u_{it} \end{aligned}$$

4.3.3.1 Dependent variable

The dependent variable measures the difference in accuracy between the CWM and AC, it will be called 'CWM-AC' and will be computed by:

- $X_{it(CWM)} - S_{it}$

Negative values indicate that the CWM is more accurate than the AC for earnings announcement (i) at time (t).

4.3.3.2 Explanatory variables

Analysts with higher values of $\overline{C_{Jt}}$ are expected to be more accurate than analysts with lower values of $\overline{C_{Jt}}$. The reason being, analysts with higher past performance compared to their peers are expected to have higher values of $\overline{C_{Jt}}$ (i.e. more contributing) and higher future performance than analysts with lower past performance relative to peers. This is likely to make the CWM more accurate compared to the AC for earnings announcement (i) at time (t). The variable ('MeanC') will be calculated by:

- *Mean of* $\overline{C_{Jt}}$ *of positive contributors – Mean of* $\overline{C_{Jt}}$ *of all analysts for earnings announcement (i) at time (t).*

Analysts with lower values of $\overline{X_{Jt}}$ are expected to be more accurate than analysts with higher values of $\overline{X_{Jt}}$. The reason for this is that higher past performance expectedly results in higher future performance. The variable ('PastAccuracy') will be computed by:

- *Mean* $\overline{X_{Jt}}$ *of positive contributors – Mean* $\overline{X_{Jt}}$ *of all analysts for earnings announcement (i) at time (t).*

When the number of forecasts per analyst increases, $\overline{C_{Jt}}$ becomes more reliable (i.e. when it is based on increasing number of forecasts it becomes more calibrated). Hence, CWM is expected to become more accurate for earnings announcement (i) at time (t). 'NForecasts' will be computed by:

- *Mean number of forecasts of all analysts for earnings announcement (i) at time (t).*

When the number of positive contributors per earnings announcement (i) at time (t) becomes larger and the number of negative contributors stays constant, accuracy of the CWM is expected to become greater compared to the AC. When the number of negative contributors for earnings announcement (i) at time (t) increases and the number of positive contributors stays constant, AC will be based on analysts with less past relative performance, decreasing accuracy of the AC compared to the CWM. An alternative interpretation for 'NegConEA' is that when the number of negative contributors increases, in accordance with wisdom of the crowd, accuracy of the AC for earnings announcement (i) at time (t) becomes greater because it will be based on more forecasts.

- 'PosConEA' will be computed by: *counting the number of positive contributors per earnings announcement (i) at time (t).*
- 'NegConEA' will be computed by: *counting the number of negative contributors per earnings (i) at time (t).*

When for positive contributors, the ratio of forecasts with a positive contribution score increase, they become more consistently contributing. Therefore, accuracy of the CWM is expected to increase.

- 'ContributionsJ' will be computed in two steps:
  - *Analyst (j)'s number of forecasts that were assigned a positive* C_{jit} */ analyst (j)'s total number of forecasts up to time (t). Which is the ratio of positive contributions of analyst (j).*
  - *Mean ratio of positive contributions of positive contributors – Mean ratio of positive contributions of all analysts for earnings announcement (i) at time (t).*

Expertise is likely to increase accuracy of the CWM compared to the NA, it is proxied for by specialisation.

- Industry specialisation ('IndExpertise') will be computed by: *The mean of number of industries for which the positive contributors supplied at least one forecast during the year (i.e. t-5 to t-1) – The mean number of industres for which all analysts supplied at least one forecast during the year for earnings announcement (i) at time (t).*
- Company specialisation ('ComExpertise') will be computed by: *The mean of number of companies for which the positive contributors supplied at least one forecast during the year*

(i.e. t-5 to t-1) – The mean number of companies for which all analysts supplied at least one forecast during the year for earnings announcement (i) at time (t).

Training is likely to increase expertise amongst the crowd. It is proxied for by experience.

- General experience ('GenExperience') is computed will be by: *The mean number of quarters in which the positive contributors have made at least one forecast – The mean number of quarters in which all analysts have made at least one forecast for earnings announcement (i) at time (t).*
- Company experience ('ComExperience') will be computed by: *The mean number of quarters in which the positive contributors have made at least one forecast for company (f) – The mean number of quarters in which all analysts have made at least one forecast for earnings announcement (i) at time (t).*

Company size, analyst following, and forecast age are proxies for the richness of the information environment. Each of them influences how much information is available for companies and affects the accuracy of the CWM.

- Company size ('Size') will be computed by: *the number of shares outstanding x share price for company (f) at the forecast-end period of earnings announcement (i) at time (t). In other words, the market capitalisation.*
- Analyst following ('Following') will be computed by: *counting the number of analysts per earnings announcement at time (t).*
- Forecast age ('Age') is computed by: *mean forecast age of positive contributors – mean forecast age of all analysts for earnings announcement (i) at time (t).*

4.3.3.3 Expected effects

A negative relation is expected for: 'MeanC', 'NForecasts', 'ContributionsJ', 'PosConEA', 'GenExperience', 'ComExperience', 'Size' and 'Following'. When these variables increase, 'CWM-AC' is expected to decrease. Contrary, a positive relation is expected for: 'PastAccuracy', 'IndExpertise', 'ComExpertise', and 'Age'. When these variables increase, 'CWM-AC' is expected to increase. Lastly, for 'NegConEA' it is uncertain', both increases and decreases can increase or decrease 'CWM-AC'. For a more thorough explanation of the expected effects, see appendix F.

4.3.4 Descriptive statistics hypotheses 1d to 1j

In table 9, descriptive statistics of the variables are given, including the expected signs of the coefficients. A minus sign stands for an expected negative relation with 'CWM-AC', and a positive sign for an expected positive relation. Furthermore, in table 10 correlations between the variables is depicted. Correlation of the independent variables with 'CWM-AC' is unexpectedly low. Therefore, the independent variables do not seem to be linearly related to the 'CWM-AC'. This is likely to have consequences for the explanatory power of the model.

From hypotheses 1a, 1b, and 1c it became apparent that, on average, the CWM is only slightly better than the AC. Therefore, it could have been expected that effects were going to be small, but not this small. The model only includes variables that capture the effect of characteristics of the CWM. Since these variables do not seem to be linearly related to accuracy of the CWM compared to the AC, other variables should capture the difference in accuracy. In literature, many alternative factors are put forward such as the quality of reported earnings that induce accuracy. However, these would not be a reason to result in a difference in accuracy between the CWM and AC. Therefore, including multiple alternative variables would be a matter of datamining rather than theorisation. This would go beyond the scope of this thesis.

The most obvious reason for the low linear association is that the model does not satisfy the assumptions of OLS. For instance, the model is missing important non-linearities. Also, the range of the 'CWM-NA' is very large and the standard deviation is out of proportion. Hence, it could also be an outlier issue.

Table 4. Descriptive statistics variables and expected signs coefficients

Variable	Obs	Mean	Std. Dev.	Min	Max	Expected sign
CWM-NA	15,414	-6196.12	107596.20	-7812907.00	3311482.00	
MeanC	15,414	261.35	544.42	0.00	14625.41	-
PastAccuracy	15,414	-3452.06	20840.24	-678776.60	581927.10	+
NForecast	15,414	79.02	46.75	5.80	265.19	-
PosConEA	15,414	13.25	5.96	5.00	57.00	-
NegConEA	15,414	3.99	3.15	0.00	32.00	?
ContributionsJ	15,414	0.01	0.03	-0.14	0.19	-
GenExperience	15,414	1.31	1.89	-8.80	23.42	-
ComExperience	15,414	0.41	0.68	-2.53	6.44	-
IndExpertise	15,414	0.11	0.22	-1.04	1.79	+
ComExpertise	15,414	0.68	0.77	-2.92	5.67	+
Following	15,414	16.95	6.44	1.00	50.00	-
Size	15,414	40077046.93	217578204.76	4689.85	11713354980.60	-
Age	15,414	0.00	4.51	-32.60	31.55	+

CWM-NA: $X_{it(CWM)} - S_{it}$. **MeanC**: *Mean of* $\overline{C_{Jt}}$ *of positive contributors – Mean of* $\overline{C_{Jt}}$ *of all analysts for earnings announcement (i) at time (t).* ***PastAccuracy****: Mean* $\overline{X_{Jt}}$ *of positive contributors – Mean* $\overline{X_{Jt}}$ *of all analysts for earnings announcement (i) at time (t).* ***NForecast****: Mean number of forecasts of all analysts for earnings announcement (i) at time (t).* **PosConEA**: *Count of the number of positive contributors per earnings announcement (i) at time (t).* **NegConEA**: Count of *the number of negative contributors per earnings (i) at time (t).* **ContributionsJ**: *Analyst (j)'s number of forecasts that were assigned a positive* C_{jit} */ analyst (j)'s total number of forecasts up to time (t). Which is the ratio of positive contributions of analyst (j). Next, mean ratio of positive contributions of positive contributors – Mean ratio of positive contributions of all analysts for earnings announcement (i) at time (t).* **IndExpertise**: *The mean of number of industries for which the positive contributors supplied at least one forecast during the year (i.e. t-5 to t-1) – The mean number of industries for which all analysts supplied at least one forecast during the year for earnings announcement (i) at time (t).* **ComExpertise**: *The mean of number of companies for which the positive contributors supplied at least one forecast during the year (i.e. t-5 to t-1) – The mean number of companies for which all analysts supplied at least one forecast during the year for earnings announcement (i) at time (t).* **GenExperience**: *The mean number of quarters in which the positive contributors have made at least one forecast – The mean number of quarters in which all analysts have made at least one forecast for earnings announcement (i) at time (t).* **ComExperience:** *The mean number of quarters in which the positive contributors have made at least one forecast for company (f) – The mean number of quarters in which all analysts have made at least one forecast for earnings announcement (i) at time (t).* **Size**: *the number of shares outstanding x share price for company (f) at the forecast-end period of earnings announcement (i) at time (t).* **Following**: Count of *the number of analysts per earnings announcement at time (t).* **Age**: *Mean forecast age of positive contributors – mean forecast age of all analysts for earnings announcement (i) at time (t).*

Table 5. Correlation variables

	CWM-NA	MeanC	PastAccuracy	NForecast	PosConEA	NegConEA	ContributionsJ	GenExperience	ComExperience	IndExpertise	ComExpertise	Size	Following	Age
CWM-NA	1													
MeanC	-0.0313	1												
PastAccuracy	-0.0108	0.0331	1											
NForecast	0.0147	-0.1881	0.0563	1										
PosConEA	-0.0177	-0.1689	0.0257	0.3399	1									
NegConEA	-0.0143	0.3804	-0.1847	-0.3738	-0.105	1								
ContributionsJ	-0.01	0.1766	-0.0269	-0.149	-0.1529	0.3185	1							
GenExperience	0.0112	0.0566	0.0518	0.0445	-0.198	0.1452	0.0087	1						
ComExperience	-0.0038	0.0598	0.0208	0.0495	-0.0156	0.0878	0.0226	0.4736	1					
IndExpertise	0.0026	0.0803	-0.0041	-0.1718	-0.2348	0.28	0.0476	0.4192	0.1684	1				
ComExpertise	-0.0039	0.0793	-0.0111	-0.2613	-0.2814	0.3909	0.047	0.5166	0.2277	0.6064	1			
Following	-0.0103	0.012	-0.0671	0.1563	0.7625	0.344	0.0359	-0.1223	0.0033	-0.0944	-0.0999	1		
Size	-0.079	0.0568	-0.0265	-0.0231	0.0993	0.0848	0.0238	-0.0304	-0.004	-0.0235	-0.03	0.1698	1	
Age	0.0009	-0.0069	-0.0083	-0.0215	-0.0338	-0.0149	-0.0539	0.1046	0.0896	0.1027	0.0837	-0.052	0.0029	1

4.4 Results hypothesis 1

4.4.1 Hypotheses 1a, 1b, and 1c

Each hypothesis will be covered separately in the following order: 1) The Paired t-test that the mean paired difference in mistakes between the CWM and the other models equals 0. 2) The Wilcoxon Signed-rank test, that the magnitude of paired rank differences is symmetrically distributed about zero. 3) The sign test that the median paired difference in mistakes between the CWM and other models equals 0.

Hypothesis 1a looks at the difference between the CWM and the NA, and had the following test results:

1. Paired t-test: The null hypothesis was rejected, t(15413)= -7.1496, p < 0.001.
2. Wilcoxon signed-rank test: the sum of positive ranks was 48678281, of negative ranks 70125034, and zero differences 91. The null hypothesis was rejected, z = -19.410, p = 0.0000.
3. Sign test: The null hypothesis was rejected, p = 0.0000, at a significance level of 1%.

Therefore, the CWM was statistically significantly more accurate than the NA.

Hypothesis 1b is used for the difference between the CWM and CAM, and had the following test results:

1. Paired t-test: The null hypothesis was not rejected, t(15413)= 0.6950, p < 0.7565.
2. Wilcoxon signed-rank test: the sum of positive ranks was 57656445, of negative ranks 61146824, and zero differences 136. The null hypothesis was rejected, z = -3.159, p = 0.0016.
3. Sign test: The null hypothesis was rejected, p = 0.0749, at a significance level of 10%.

Therefore, the CWM was statistically significantly more accurate than the CAM using the Wilcoxon signed-rank test and sign test, but not the Paired t-test.

Hypothesis 1c is established to test the differences between the CWM and CWM(X), and had the following test results:

1. Paired t-test: The null hypothesis was not rejected, t(12814)= 1.7943, p < 0.9636.
2. Wilcoxon signed-rank test: the sum of positive ranks was 40923134, of negative ranks 41165251, and zero differences 30135. The null hypothesis was not rejected, z = -0.289, p = 0.7725.
3. Sign test: The null hypothesis was not rejected, p = 0.1075, at a significance level of 10%.

Therefore, the CWM(X) was not statistically significantly more accurate than the CWM at a reasonable significance level.

4.4.2 Hypotheses 1d to 1j

Table 6 depicts the regression results of the accuracy drivers on 'CWM-AC'. The R^2 is 0.009 and the adjusted R^2 is 0.008, hence 99.2% of the total variation in 'CWM-AC' is unexplained. The F-test (10.23) is significant with a p-value of 0.000 at 1% significance level. This means that the null hypothesis, that at none of the independent variables is linearly related to 'CWM-AC', is rejected. Therefore, the model does in theory have some validity, but this is mostly due to the large sample size of 15414. Furthermore, the model is missing non-linearities, heteroskedasticity is present, and the normality assumption of the residuals is not satisfied. See appendix G for tests.

Therefore, out of the 13 independent variables, 5 are significant. However, even the statistically significant variables are not valid because of the low explanatory power and violation of the assumptions. Therefore, no conclusion can be drawn for hypothesis 1d to 1j. In appendix G some alternative models are discussed, with the purpose of amending the data.

Table 6. Regression Results CWM-AC

	Expected sign	Coef.	P-value
MeanC	-	-5.642	***0.001
PastAccuracy	+	-0.056	0.188
NForecast	-	23.149	0.290
PosConEA	-	-1151.482	***.000
NegConEA	?	-486.794	0.278
ContributionsJ	-	-30617.580	0.335
GenExperience	-	1030.167	*0.096
ComExperience	-	-1211.069	0.406
IndExpertise	+	1950.756	0.698
ComExpertise	+	-2520.046	0.125
Following	-	934.759	***0.001
Size	-	-0.00004	***0.000
Age	+	23.947	0.902
F-test		10.230	***0.000
R-squared		0.009	
Adjusted R-squared		0.008	
Observations		15414	

* denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

4.5 Methodology hypothesis 2

Hypothesis 2 is established to test whether earnings surprises of the contribution weighted model are a good proxy for actual earnings surprises, and whether this proxy (i.e. predicted earnings surprises) can be used to predict stock market responses. Firstly, the method of testing of the hypotheses is described and afterwards the descriptive statistics are given.

4.5.1 Testing hypotheses 2a and 2b

Two correlations will be used for the hypotheses: The Pearson correlation coefficient to test for a linear relation. The Spearman rank correlation coefficient to test for a monotonic relationship. Specifically, tested is whether the predicted earnings surprise of the contribution weighted model ($P_{it} - M_{it}$) is correlated with actual earnings surprises ($O_{it} - M_{it}$), and whether $P_{it} - M_{it}$ is correlated with the response of the stock market (CAR_{fit}).

4.5.2 Descriptive statistics hypotheses 2a and 2b

The relationship between the predicted earnings surprise and actual earnings surprise is most important (i.e. H2a). This validates the use of the predicted earnings surprise to proxy for actual earnings surprises. Also, it validates the contribution weighted model regardless of whether a relation with stock market responses is found. From hypothesis 1 is known that the CWM is significantly more accurate than the AC. Specifically for 8553 over 6848 earnings announcements (see table 3). In other words, for a little bit over half of the earnings announcements, the direction of $P_{it} - M_{it}$ is the same as $O_{it} - M_{it}$, and for all the others it is in the opposite direction.

Figure 3 depicts a scatterplot of the relationship between actual earnings surprises and predicted earnings surprises. It shows signs of outliers, which presumably leads to the large difference between the Pearson and Spearman correlation coefficient depicted in table 7. In terms of the Pearson correlation coefficient, $P_{it} - M_{it}$ is a reasonable proxy for $O_{it} - M_{it}$ with 0.6559. However, in terms of the Spearman correlation coefficient it is not, with 0.2851. The linear relationship seems to be biased. The Spearman is likely to be more correct because it is less affected by outliers as it is based on ranks. A reason for this difference between the correlation coefficients could be, that for only about half of the earnings surprises, the predicted and actual earnings surprise have the same sign (i.e. low Spearman coefficient), and that predicted and actual earnings surprises have large outliers (i.e. high Pearson coefficient).

Another striking feature depicted in table 7 is the low correlation between CAR_{fit} and $O_{it} - M_{it}$. This means that there is only a weak association between hypothesised stock market responses and actual earnings surprises. Theorised and widely used as benchmark for expectation of the market is M_{it} (i.e. Analyst consensus) seems be followed by a different response than expected. Missing or beating it, seems to only limitedly explains stock movements. Therefore, even when a more accurate proxy is found, it will likely limitedly explain stock movements.

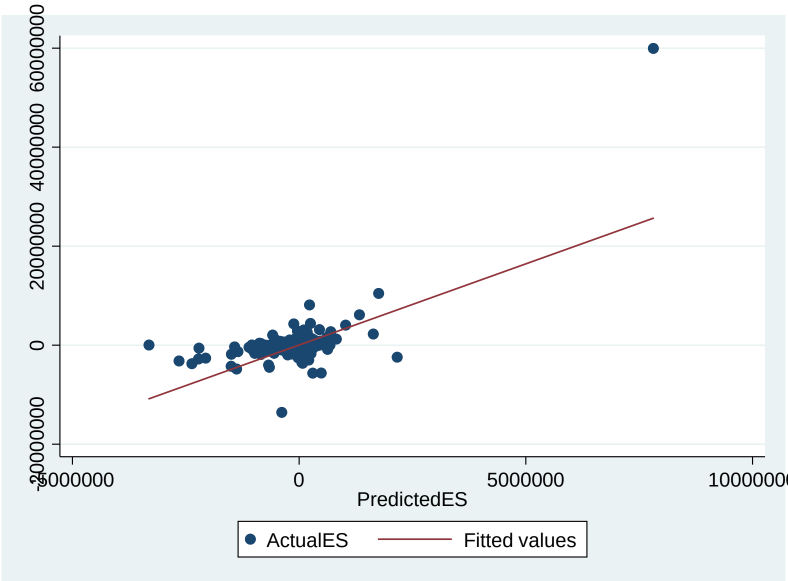

Figure 3. Scatterplot of relationship actual and predicted earnings surprises

Table 7. Pearson correlation (left) and Spearman correlation (right), Full sample

	CAR	Actual ES	Predicted ES
CAR	1		
Actual ES	**0.1339**	1	
Predicted ES	-0.0402	0.6559	1

	CAR sign	Actual ES sign	Predicted ES sign
CAR sign	1		
Actual ES sign	**0.1714**	1	
Predicted ES sign	-0.0124	0.2851	1

CAR denotes CAR_{fit}. ES denotes earnings surprise. Sign denotes positive/ negative.

As, $P_{it} - M_{it}$ does not seem to be a useful proxy, a subset of companies is identified for which the CWM performed superior. Superior is defined using two conditions (explained in detail in appendix H). The first condition is that the CWM is more accurate for 70% of a company's earnings announcements. The second condition is that a minimum of 8 earnings announcements of the company are included in the sample. After imposing these conditions, a subset of 90 companies and 1692 earnings announcements remains. Table 8 depicts descriptive statistics. For instance, there are 18 companies for which the CWM was more accurate for 75% to 79.99% of earnings announcements. Also, the average number of earnings announcements and total number of earnings announcements are depicted. Caution should be taken when interpreting these numbers. Despite that of each company there are, on average, 19 earnings announcements in the sample, these results could still be random.

Furthermore, table 9 states the Pearson and Spearman correlation coefficients. Note that the Pearson and Spearman correlation coefficients of predicted and actual earnings surprises are similar and strong now. Indicating that both the linear and monotonic relationship are roughly the same. Therefore, it is a much more reasonable proxy in terms of association. Nonetheless, correlation between predicted earnings surprises and CAR_{fit} is low. It is slightly better than the small negative relation of -0.0402 and -0.0124, using the full sample seen in table 7, but it seems not enough to predict stock market responses.

Table 8. Companies for which the CWM performed ...% of EA better than AC

	Companies	Avg. EA comp.	Total number EA
70% - 74.99%	31	20	611
75% - 79.99%	18	17	309
80% - 84.99%	21	21	439
85% - 89.99%	14	16	229
90% - 94.99%	6	17	103
95% - 100%	0	0	0
	90	19	1692

Table 9. Pearson correlation (left) and Spearman correlation (right), 70% top performers

	CAR	Actual ES	Predicted ES
CAR	1		
Actual ES	0.1687	1	
Predicted ES	0.0361	**0.7116**	1

	CAR sign	Actual ES sign	Predicted ES sign
CAR sign	1		
Actual ES sign	0.1321	1	
Predicted ES sign	0.007	**0.6719**	1

CAR denotes CAR_{fit}. ES denotes earnings surprise. Sign denotes positive/ negative.

To see what the correlations mean in actual values, descriptive statistics are given in table 10 and 11. Table 10 depicts descriptive statistics of stock market responses after positive and negative earnings surprises for the full sample, and table 11 for the 70% top performers. Earnings surprises are grouped by magnitude. Note that they are given in EPS rather than in earnings. This is done due to differences in size between companies. For instance, an earnings surprise of company X of -50000 in earnings is only -0.05 in EPS, whereas for company Y an earnings surprise of -50000 in earnings is a surprise of -0.50 in EPS. Also, investors do not look at earnings but at EPS. They would in theory respond more severely to the earnings announcement of company Y than to the earnings announcement of company X. Moreover, it is in line with literature such as Bartov et al. (2002) and Kasznik and McNichols (2002) to look at earnings surprises in EPS. Furthermore, an earnings surprise of less than 0 means missing the expectation of the market, of 0 meeting, and of more than 0 beating. Lastly, 'Same direction' denotes whether the direction of an earnings surprise is equal to the direction of the market response for earnings announcement (i) at time (t).

Like in the correlation tables, the predicted earnings surprise does not seem to predict stock market responses, as on average 49% (i.e. full sample) and 50% (i.e. 70% top performers) are in the same direction. The size of the earnings surprise does not seem to matter, counter to literature (e.g. Kinney et al., 2002). The ratio under 'Same Direction' seems to randomly increase and decrease for various sizes of predicted earnings surprises. Also, actual earnings surprises seem to limitedly predict stock market responses. Contrary, a relationship between the size of actual earnings surprises and the ratio under 'Same Direction' does seem to be present. This is more in line with literature. Especially, just meeting the analyst consensus is sometimes (i.e. not always) seen as a sign of earnings management and therefore, the market response is difficult to predict. This can be observed in table 10, around 0, the ratio under 'Same Direction' is lower than for greater actual earnings surprises.

To investigate if the strength of the relation between earnings surprises and stock market responses can increase, the effect of two additional benchmarks is investigated. These are depicted in appendix I. Results are given for missing, meeting, or beating the analyst consensus in combination with the benchmark of consecutive beating the market and consecutive earnings growth. No striking differences are found with the results stated in this paragraph.

Table 10. Response market after predicted (left) and actual (right) earnings announcement, full sample

Predicted ES			EA	Same direction
x	<	-0.141	689	0.52
-0.14	-	-0.061	915	0.48
-0.06	-	-0.021	1845	0.50
-0.02	-	-0.001	4998	0.48
0			20	0.55
0.001	-	0.02	4602	0.50
0.021	-	0.06	1290	0.50
0.061	-	0.14	474	0.48
x	>	0.141	312	0.54
Total			**15145**	**0.49**

Actual ES			EA	Same direction
x	<	-0.141	1188	0.63
-0.14	-	-0.061	1129	0.65
-0.06	-	-0.021	1392	0.63
-0.02	-	-0.001	1173	0.57
0			1189	0.42
0.001	-	0.02	1648	0.50
0.021	-	0.06	3255	0.57
0.061	-	0.14	2274	0.62
x	>	0.141	1897	0.63
Total			**15145**	**0.58**

Table 11. Response market after predicted (left) and actual (right) earnings announcement, 70% top performers

Predicted ES			EA	Same direction
x	<	-0.141	190	0.49
-0.14	-	-0.061	160	0.41
-0.06	-	-0.021	243	0.52
-0.02	-	-0.001	287	0.51
0			1	1.00
0.001	-	0.02	342	0.54
0.021	-	0.06	234	0.52
0.061	-	0.14	115	0.45
x	>	0.141	120	0.53
Grand Total			**1692**	**0.50**

Actual ES			EA	Same direction
x	<	-0.141	257	0.58
-0.14	-	-0.061	185	0.61
-0.06	-	-0.021	165	0.52
-0.02	-	-0.001	102	0.51
0			43	0.35
0.001	-	0.02	102	0.63
0.021	-	0.06	235	0.55
0.061	-	0.14	250	0.58
x	>	0.141	353	0.58
Grand Total			**1692**	**0.57**

4.6 Results hypotheses 2a and 2b

In this paragraph the significance of the correlations is stated.

Firstly, the significance of the proxy, the relationship between predicted and actual earnings surprises, was tested:

- Pearson correlation for the full sample: ρ = 0.6559, p-value = 0.0000, 15145 observations. The null hypothesis was rejected at a significance level of 1%. This relationship was likely to be biased and therefore this result is presumably not valid.

- Spearman correlation for the full sample: ρ_S= 0.2851, p-value = 0.0000, 15415 observations. The null hypothesis was rejected at a significance level of 1%.

- Pearson correlation for the subset of the sample: ρ = 0.7116, p-value = 0.0000, 1692 observations. The null hypothesis was rejected at a significance level of 1%.

- Spearman correlation for subset of the sample: ρ_S= 0.6719, p-value = 0.0000, 1692 observations. The null hypothesis was rejected at a significance level of 1%.

In sum, a monotonic and linear relationship is found for the full sample and subset of the sample. Therefore, in theory enough evidence is found for hypothesis 2a. However, the relationship for the full sample is likely to be invalid, and the relationship for the subsample might be based on results that are random. In other words, in practice not enough evidence is found for hypothesis 2a.

Secondly, the significance of the relation between predicted earnings surprises and stock market movements was tested:

- Pearson correlation for the full sample: ρ = 0.-0.0402, p-value = 0.0000, 15145 observations. Hence, the null hypothesis was rejected at a significance level of 1%.

- Spearman correlation for the full sample: ρ_S= -0.0124, p-value = 0.129, 15415 observations. Hence, the null hypothesis was not rejected at a significance level of 1%.

- Pearson correlation for the subset of the sample: ρ = 0.0361, p-value = 0.138, 1692 observations. Hence, the null hypothesis was not rejected at a significance level of 1%.

- Spearman correlation for subset of the sample: ρ_S= 0.007, p-value = 0.773, 1692 observations. Hence, the null hypothesis was not rejected at a significance level of 1%.

In sum, a negative linear relationship exists for the full sample. All other values are insignificant. Therefore, not enough evidence is found for hypothesis 2b.

For comparison, the correlation between actual earnings surprises and stock market responses were all significant at the 1% significance level.

5 Discussion and interpretation

This thesis examined the validity of the contribution weighted model of Budescu and Chen (2014) in the financial analyst environment (i.e. hypothesis 1) and its applicability to predict stock market responses (i.e. hypothesis 2).

5.1 Limitations and assumptions

A limitation of the research in this thesis was the dependence on finding a significant result for hypothesis 1a. Insignificant results would greatly reduce chances of finding significant (i.e. statistical and practical) results for the regression of hypothesis 1 and for hypothesis 2. For hypothesis 2, it would mean that assumptions would have to be made, that if it would occur, extra conditions had to be imposed. Imposing extra conditions would greatly reduce the external validity.

Also, a response period of 3 days was used to measure stock market responses. One caveat of using a short response window of three days is that companies with, especially negative, earnings surprises tend to disclose information before the actual earnings announcement to mitigate the effect on share price. The response to these information disclosures are likely not included in the cumulative abnormal return measure (Skinner and Sloan, 2002). There were a couple of reasons for choosing a short response period. Firstly, a trade-off had to be made between the response period and latest forecast. In this thesis, forecasts of up to 2 days before the earnings announcement were used in the calculations of the CWM. Therefore, the longest possible response period was of 1 day before the earnings announcement. If the response period increased, only older forecasts were to be used, losing the more recent forecasts closer to earnings announcements. It was decided to keep all forecasts and go for the short response period. Especially, because a short response period is widely used in literature.

Lastly, it was assumed that after earnings surprises, the market would respond in the hypothesised direction. In the literature (e.g. Bartov et al., 2002; Kasznik and McNichols, 2002; Kinney et al., 2002; and Myers et al., 2007), it is implied that imposing certain benchmarks such as consecutive meeting the analyst consensus, increase the markets' response. Therefore, the line of reasoning used for this assumption was to impose multiple benchmarks so that the response would become more pronounced. It turned out that imposing the benchmarks did not have an effect. Regardless of finding a market response, the validation of the CWM did not depend on it, because hypothesis 2a was included.

5.2 Hypothesis 1

Starting off with hypothesis 1, the contribution weighted model (CWM) was tested for accuracy against the analyst consensus (AC), the contribution average model (CAM), and the contribution weighted model with an extra condition (CWM(X)). For the CWM to be of value, it needed to be statistically significantly more accurate than the other models, and practically significantly more accurate. Firstly, the comparison between the CWM and AC will be examined. Secondly, the comparison between the CWM and CAM and CWM(X) will be discussed. Thirdly, the investigation of the accuracy drivers will be expounded.

5.2.1 hypothesis 1a

The comparison between the CWM and AC was made in terms of accuracy. Specifically, by testing the difference in mean mistake per earnings announcement, and by the number of earnings announcements the CWM was found to be more accurate. The CWM was on average $6196 per earnings announcement and for 1705 earnings announcements more accurate than the analyst consensus, which was about 10.65% more accurate in terms of magnitude and 11.06% in terms of earnings announcements. This seems to be on the lower end of what Budescu and Chen (2014) and Chen et al. (2016) found. In one study they found an increase in accuracy of about 28% compared to a simple average (i.e. the same as the analyst consensus), and in another an increase of 7.38% compared to a baseline. This baseline did not resemble the crowd, so caution should be taken comparing the results of this thesis with that baseline. Nonetheless, the study of the 7.38% most resembles the financial analyst environment, as this result was found for predictions on inflation and real GDP growth. In their latest study, they found increases in accuracy over other models ranging from 27.25% to 64.23% (Chen et al., 2016). These other models did not include a simple average, which makes comparison with results found in this thesis less suitable. However, these other models were based on carefully picked forecasters, which one can assume were more accurate than a simple average. Therefore, in this thesis, the CWM proved its value in statistical terms in a new environment, as also in the financial analyst environment increases in accuracy compared to a simple average were found. However, the results were less overwhelming, which made the practical significance of the model in this environment likely insignificant. In practice, the model is only valid when it has value to investors. Investors would benefit when the forecast of the CWM could be used to predict earnings surprises. To realise this, the CWM should predict the size and direction of earnings surprises correctly and consistently. This was not the case. The model was more accurate than the analyst consensus for 8553 of 15414 earnings announcements, which was only about 55%. Therefore, the advantage of employing this model in practice seems limited.

A reason that the accuracy increase of the CWM was on the lower end, compared to Budescu and Chen (2014) and Chen et al. (2016), could be that the mean and median number of positive contributors per earnings announcement were 78% and 82%. This indicates that forecasts of the CWM and AC were for a large part based on the same composition of analysts. Budescu and Chen (2014) and Chen et al. (2016) do not mention this issue. Therefore, it is not certain how or if they dealt with it. In this thesis, truncating the datapoints was tested, but it did not change the high percentage of positive contributors.

5.2.2 Hypotheses 1b and 1c

Moreover, the CWM was compared to the CAM and CWM(X). Accuracy between these models were not statistically different. This strengthened the robustness of the CWM, as different weighting schemes and an extra condition did not seem to decrease accuracy. However, it also did not give future prospects that increases in accuracy are realisable. Nonetheless, with regards to the comparison between the CAM and CWM two things became evident after the comparison between the models. The increase in accuracy of the CWM over the AC is likely:

1) due to the ability of the CWM to identify superior performers for 10.22% (i.e. of total of 10.65%) and 9.62% in terms of earnings announcements (i.e. of total of 11.06%), and

2) less due to the ability of the CWM to weight analysts based on their level of expertise for about -0.43% in terms of magnitude (i.e. of total of 10.65%) and 1.44% in terms of earnings announcements (i.e. of total of 11.06%).

This is similar to the results of Budescu and Chen (2014) who also found minimal improvement of the CWM over the CAM in general, and due to weighting.

With regards to the CWM(X), based on the findings, for this dataset one can say, that the CWM efficiently identifies top performers based on the total contribution score. Imposing an extra condition on which to identify top performers does not lead to extra accuracy. However, that does not mean other conditions could not increase accuracy, or that in a different environment it would not work. It was a novel model as it was not tested by Budescu and Chen (2014). They only mention how they used the condition to identify top performers. Nonetheless, no increased accuracy was found and the number of earnings announcements for which it could be used was lower than the CWM. Therefore, in the financial analyst environment no evidence was found to use the CWM(X) over the CWM.

5.2.3 Hypotheses 1d to 1j

To test why the CWM was more accurate than the AC, an OLS regression was performed. The goal was to see what parts of the model explained accuracy most, and under what circumstances it was most accurate, and thereby validate the practical use of the model. Overall the model did not do a very good job in explaining a difference between the CWM and analyst consensus. With an $R^2$of 0.0086 and adjusted R^2 of 0.0078, very little of the variance of the difference was explained by the independent variables (i.e. accuracy drivers identified by Budescu and Chen (2014) and Chen et al. (2016)). Also, OLS assumptions were not met, and attempts to amend the data did not work. Based on these results, no conclusions could be drawn.

5.3 Hypothesis 2

Hypothesis 2 was used to test whether predicted earnings surprises were a proxy for actual earnings surprises and to test whether this proxy can be used to predict stock market responses. In other words, whether investors can use the contribution weighted model in their investment strategy was investigated. It was found that the earnings surprise using the full sample did not proxy well for actual earnings surprises. Therefore, extra conditions had to be imposed, limiting the validity of the results. For a subset (i.e. 70% top performers), predicted earnings surprises were strongly and positively related to actual earnings surprises with Pearson and Spearman correlation coefficients of 0.7116 and 0.6719. Hence, the proxy did what it was supposed to do, in that it found a correlation between predicted and actual earnings surprises. Nonetheless, the relationship with stock market responses was weak and not significant. Earnings surprises were followed by stock market responses in the expected direction 50% of the earnings announcements. Therefore, one could be better of just flipping a coin, and save time. In fact, even actual surprises were weakly correlated with stock market responses. Actual earnings surprises were followed by market responses in the expected direction only about 58% of the earnings announcements. A reason for the insignificant relation between earnings surprises and market responses could be that a response period of 3 days was used (Skinner and Sloan, 2002).

However, it was similar to the results of Kinney et al. (2002), who found that on average 56% of earnings surprises are followed by market responses in the same direction. Kinney et al. (2002) also tested for magnitude, and found that there is a more linear relation for greater earnings surprises. In this thesis, a more linear relation seemed to be present for greater actual earnings surprises but not very profoundly. And for predicted earnings surprises it was not present, as for varying earnings surprise sizes, the percentage that stock market responses were in the same direction, randomly increased or decreased, for both the full sample and the subset of the sample. It was expected that the response in the same direction could be increased by imposing extra benchmarks. As Kasznik and McNichols (2002) found a stronger relation for consecutive meeting the analyst consensus, and Myers et al. (2007) for subsequent growth of earnings, these were imposed. However, there were some signs that these extra benchmarks had an effect for actual earnings surprises, but this was not very profound. And for predicted earnings surprises no effect was found. The results of predicted earnings surprises seemed to be guided by randomness. In other words, there seems to be some correlation between actual earnings surprises and market responses, however, a large part of the stock market movement was not reliably related to the size of earnings surprises, consecutive beating of the markets' expectation, and consecutive earnings growth. Hence, when even actual earnings surprises limitedly explained the stock market movements using multiple benchmarks, it became very difficult for the predicted earnings surprises to explain it.

6 Conclusion

The question: “*to what extent is it possible to more accurately forecast earnings than the analyst consensus using the contribution weighted model of Budescu and Chen (2014) and use the difference to predict stock market responses?”* is researched using two hypotheses.

The contribution weigthed model was about 10% statistically significantly more accurate than the analyst consensus. Therefore, evidence was found for hypothesis 1a. Accuracy was not further increased by employing a different weighting scheme (CAM) and imposing an extra condition (i.e. CWM(X)). Hence, not enough evidence was found for hypotheses 1b and 1c. The results seem to mainly stem from the CWM’s ability to identify superior performers. In other words, in terms of statistical significance this validates the use of the contribution weighted model in a new environment. However, in practice, 10% increase in accuracy is presumably not worth the effort for investors. Furthermore, the regression results to increase the practical relevance were not valid. As such, no conclusions could be drawn for hypotheses 1d to 1j.

As the practical significance of accuracy of the CWM was low, it was limitedly possible, after imposing extra conditions, to predict earnings surprises. As such, limited evidence was found for hypothesis 2a. Consequently, it was also not possible to predict market responses with the full sample. Also, no results were found using a subset of the sample. Therefore, no evidence was found for hypothesis 2b. That no evidence was found, was not entirely due to the proxy of the CWM, because also actual earnings surprises did not forecast market responses entirely. So even when a perfect proxy for actual earnings would have been found, it would only have been possible to predict stock market responses in a limited way.

6.1 Future recommendations

Future research could focus on refining the CWM by focussing on a more precise measure of the total contribution score. In this thesis, the total contribution score was based on the general relative forecast history of an analyst. Presumably, the level of analyst accuracy per company differs. For one company analyst (j) is more accurate than for another. Therefore, analysts who showed superior performance for a few companies and less for most, might have been given a negative total contribution score. This means that their forecast of the one company for which they showed superior performance was not included in the calculations of the forecast of the CWM. Hence, a method to refine the CWM, is to compute the total contribution score based the company specific or industry specific history of an analyst.

Another refinement could be to look at a different weighting scheme. No increases in accuracy were found for the ability of the CWM to weight forecasts of analysts. However, it might be possible to improve this by imposing certain conditions. Such as that an analyst should not be assigned a higher weight than a certain percentage (e.g. 50%). When one analyst is assigned a weight that is too high, accuracy of the CWM is expected to deteriorate. This stems from the principle of wisdom of the crowd, when the number of forecasters on which the forecast is based decreases, accuracy decreases.

Furthermore, a refinement could be made by looking at the percentage of positive contributors of 82%. The model could be improved by decreasing this percentage. When this percentage decreases, it will become more distinct from the analyst consensus. One possibility could be to compute contribution scores for an analyst by his ‘relative’ absolute relative difference from the analyst consensus rather than by the absolute relative difference from the analyst consensus. In other words, taking the computation of the contribution score one step further. To compute the contribution score of a forecast and divide it by the mistake of the analyst consensus. Alternatively, not normalise EPS by shares outstanding, and compute the contribution score of a forecast using values in EPS, and divide it by the mistake of the analyst consensus. The level of total contribution scores might then be less dependent on the size of a company's earnings. Even in general, computing contribution scores by 'relative' past relative difference could increase accuracy as it would control for inter-temporal changes and cross-sectional differences in task difficulty per company.

7 References

Alford, A. W., & Berger, P. G. (1999). A simultaneous equations analysis of forecast accuracy, analyst following, and trading volume. Journal of Accounting, Auditing & Finance, 14(3), 219-240.

Bartov, E., Givoly, D., & Hayn, C. (2002). The rewards to meeting or beating earnings expectations. Journal of accounting and economics, 33(2), 173-204.

Bhojraj, S., Hribar, P., Picconi, M., & McInnis, J. (2009). Making sense of cents: An examination of firms that marginally miss or beat analyst forecasts. The Journal of Finance, 64(5), 2361-2388.

Bradshaw, M. T. (2011). Analysts' forecasts: what do we know after decades of work?.

Brown, L. D. (2001). How important is past analyst forecast accuracy?. Financial Analysts Journal, 44-49.

Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2015). Inside the "black box" of sell-side financial analysts. Journal of Accounting Research, 53(1), 1-47.

Budescu, D. V., & Chen, E. (2014). Identifying expertise to extract the wisdom of crowds. Management Science, 61(2), 267-280.

Chen, E., Budescu, D. V., Lakshmikanth, S. K., Mellers, B. A., & Tetlock, P. E. (2016). Validating the contribution-weighted model: Robustness and cost-benefit analyses. Decision Analysis, 13(2), 128-152.

Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter?. Journal of Accounting and Economics, 27(3), 285-303.

Clement, M. B., & Tse, S. Y. (2005). Financial analyst characteristics and herding behavior in forecasting. The Journal of finance, 60(1), 307-341.

Das, S., Levine, C. B., & Sivaramakrishnan, K. (1998). Earnings predictability and bias in analysts' earnings forecasts. Accounting Review, 277-294.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... & Münkemüller, T. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36(1), 27-46.

Dunn, K., & Nathan, S. (2005). Analyst industry diversification and earnings forecast accuracy. The Journal of Investing, 14(2), 7-14.

Easterwood, J. C., & Nutt, S. R. (1999). Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism?. The Journal of Finance, 54(5), 1777-1797.

Galton, F. (1907). Vox populi (The wisdom of crowds). Nature, 75(7), 450-451.

Fried, D., & Givoly, D. (1982). Financial analysts' forecasts of earnings: A better surrogate for market expectations. Journal of Accounting and Economics, 4(2), 85-107.

Friesen, G., & Weller, P. A. (2006). Quantifying cognitive biases in analyst earnings forecasts. Journal of financial Markets, 9(4), 333-365.

Gleason, C. A., & Mills, L. F. (2008). Evidence of differing market responses to beating analysts' targets through tax expense decreases. Review of Accounting Studies, 13(2-3), 295-318.

Gordon, K. (1924). Group judgements in the field of lifted weights. Journal of Experimental Psychology 7, 389–400.

Henry, E., Robinson, T. R., & Stowe, J. D. (2010). Equity asset valuation (Vol. 27). John Wiley & Sons.

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. The Journal of Finance, 64(5), 2289-2325.

Hong, H., Kubik, J. D., & Solomon, A. (2000). Security analysts' career concerns and herding of earnings forecasts. The Rand journal of economics, 121-144.

Jacob, J., Lys, T. Z., & Neale, M. A. (1999). Expertise in forecasting performance of security analysts. Journal of Accounting and Economics, 28(1), 51-82. Keane, M. P., & Runkle, D. E. (1998). Are financial analysts' forecasts of corporate profits rational?. Journal of Political Economy, 106(4), 768-805.

Kasznik, R., & McNichols, M. F. (2002). Does meeting earnings expectations matter? Evidence from analyst forecast revisions and share prices. Journal of Accounting research, 40(3), 727-759.

Kinney, W., Burgstahler, D., & Martin, R. (2002). Earnings surprise “materiality” as measured by stock returns. Journal of Accounting Research, 40(5), 1297-1329.

Koller, T., Goedhart, M., & Wessels, D. (2010). Valuation: measuring and managing the value of companies (Vol. 499). john Wiley and sons.

Kothari, S. P. (2001). Capital markets research in accounting. Journal of accounting and economics, 31(1-3), 105-231.

Lim, T. (2001). Rationality and analysts' forecast bias. The Journal of Finance, 56(1), 369-385.

Mian, G. M., & Sankaraguruswamy, S. (2012). Investor sentiment and stock market response to earnings news. The Accounting Review, 87(4), 1357-1384.

Mikhail, M. B., Walther, B. R., & Willis, R. H. (1999). Does forecast accuracy matter to security analysts?. The Accounting Review, 74(2), 185-200.

Myers, J. N., Myers, L. A., & Skinner, D. J. (2007). Earnings momentum and earnings management. Journal of Accounting, Auditing & Finance, 22(2), 249-284.

O’brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality & quantity, 41(5), 673-690.

O'brien, P. C. (1988). Analysts' forecasts as earnings expectations. Journal of accounting and Economics, 10(1), 53-83.

Ramnath, S., Rock, S., & Shane, P. (2008). The financial analyst forecasting literature: A taxonomy with suggestions for further research. International Journal of Forecasting, 24(1), 34-75.

Rees, L. (2005). Abnormal returns from predicting earnings thresholds. Review of Accounting Studies, 10(4), 465-496.

Rees, L., & Sivaramakrishnan, K. (2007). The effect of meeting or beating revenue forecasts on the association between quarterly returns and earnings forecast errors. Contemporary Accounting Research, 24(1), 259-290.

Ronen, J., & Yaari, V. (2008). Earnings management (Vol. 372). Springer US.

Schipper, K. (1991). Analysts' forecasts. Accounting horizons, 5(4), 105.

Shon, J., & Zhou, P. (2011). Trading on Corporate Earnings News: Profiting from Targeted, Short-term Options Positions. FT Press.

Skinner, D. J., & Sloan, R. G. (2002). Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. Review of accounting studies, 7(2-3), 289-312.

Surowiecki, J. (2004). The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. Economies, Societies and Nations, 296.

Treynor, J. L. (1987). Market efficiency and the bean jar experiment. Financial Analysts Journal, 50-53.

Appendix

Appendix A

Legend

AR_{fit}	-	Abnormal return
CAR_{fit}	-	Cumulative abnormal return during earnings announcement. Proxies for response market
C_{jit}	-	Contribution C of an analyst j for earnings announcement i at time t
C	-	Contribution
$\overline{C_{jt}}$	-	Average of sum of C_{jit} for individual analyst j at time t
CWM	-	Based on a weighted average of positive contributors for an earnings announcement at time t. It stands for the Contribution Weighted Model
CAM	-	Based on normal average of positive contributors for an earnings announcement at time t. It stands for the Contribution Average Model.
CWM(X)	-	Based on a weighted average of positive contributors for an earnings announcement at time t, including an additional condition of level of absolute mistake.
ES_{it}	-	Predicted earnings surprise
ε	-	Minimum required level of $\overline{C_{jt}}$ for an analyst to be included in calculation of $P_{Gi(t+1)}$
F	-	Company
i	-	Earnings announcement
J	-	Analyst
M_{it}	-	The average of all analysts for an earnings announcement i at time t. In other words, the analyst consensus
NA	-	It stands for Normal Average. It takes the normal average of all analysts for an earnings announcement at time t
$P_{i(t+1)}$	-	Group forecast of analysts with positive $\overline{C_{jt}}$ for earnings announcement I
$P_{i(t+1)} - M_{it}$	-	Predicted earnings surprise
S_{it}	-	Absolute mistake of the NA, also called earnings surprise
X_{it}	-	Absolute mistake of the CWM
X_{jit}	-	Absolute difference between analyst j's forecast and actual earnings
$\overline{X_{jt}}$	-	Average of sum X_{jit} for individual analyst j at time t
ω_{jit}	-	Weight given to forecast of analyst (j) for earnings announcement (i) at earnings announcement (i) at time (t)

Appendix B

A typical row of the database is depicted in figure 4.

IBES Ticker Symbol	CUSIP/SEDOL	Company Name	Analyst Code	Measure (Data Type Indicator)	Estimate Value	Forecast Period End Date	Announce Date Analyst	Actual Value	Announce date of the Actual
MSFT	59491810	MICROSOFT	77578	EPS	0.58	30/06/2011	09/06/2011	0.69	21/07/2011
MSFT	59491810	MICROSOFT	49027	EPS	0.61	30/06/2011	29/04/2011	0.69	21/07/2011
MSFT	59491810	MICROSOFT	83598	EPS	0.58	30/06/2011	29/04/2011	0.69	21/07/2011
MSFT	59491810	MICROSOFT	48199	EPS	0.58	30/06/2011	29/04/2011	0.69	21/07/2011
MSFT	59491810	MICROSOFT	1043	EPS	0.57	30/06/2011	29/04/2011	0.69	21/07/2011
MSFT	59491810	MICROSOFT	40322	EPS	0.56	30/06/2011	29/04/2011	0.69	21/07/2011
MSFT	59491810	MICROSOFT	86602	EPS	0.59	30/06/2011	29/04/2011	0.69	21/07/2011

Figure 4. Database example

Here IBES, CUSIP/SEDOL, and Company Name are identifiers of the company and can be used to combine other data sources with the sample. Analyst code is the identifier of analyst (j). IBES aims to give a single analyst code to each analyst from start of the career to retirement, even after switching employer. Measure (Data Type Indicator) confirms that EPS is forecasted. Estimate Value is the analysts' EPS forecast ($p_{ji(t+1)}$). Forecast Period End Date is the (final) deadline by which the analyst can convey a forecast. This is not a hard deadline, it happens that analysts draw up forecasts after the deadline. These are also included in the database. The Announce date of the Actual denotes a hard deadline, as this is the announcement date of the Actual Value of EPS. Lastly, the Announce Date Analyst denotes the date by which the analyst drew up a forecast.

Not all forecasts and earnings announcements can be used. Some rows or cells did not hold information. Rows where 'Actual Value' was missing were removed as without actual value it is impossible to calculate for these earnings announcements it would not be possible to calculate C_{jit}. Likewise, rows without 'Analyst code', 'CUSIP/SEDOL', 'IBES Ticker', 'Company Name', and 'Announce date' were dropped. This totals to 10795 rows being dropped, which is named drop A.

Furthermore, some forecasts were announced before the previous earnings announcement date and were dropped, as it is unclear whether these are related to the previous or current earnings announcement. Also, forecasts with announce dates after next earnings announcement date were removed, as actual EPS has already been announced. Therefore, all forecasts are from between previous earnings announcement date + 1 day and next earnings announcement date – 1 day. This is in line with Bartov et al. (2002) and Friesen and Weller (2006). A total of 293707 were dropped, this is named drop B.

When analyst (j) has more than one forecasts for earnings announcement (i), the latest one is kept (closer to earnings announcement) and the earliest one removed (further away from earnings announcement). This is in line with other research, e.g. Mikhail et al. 1999, Clement and Tse (2005), and Hirshleifer (2009), who state that removing the earliest forecast mitigates the horizon effect. The horizon effect describes that forecasts further away from earnings announcements are generally less accurate. Therefore, by eliminating the oldest forecast, contribution of analyst (j) is not contaminated by older and generally less accurate forecasts. A total of 188704 were dropped and this is named drop C.

Additionally, data was taken from different databases. The forecasts for which the identifier did not match and for which it was impossible to link all relevant information, were removed. This involved 49417 forecasts and is named drop D.

Finally, Brown (1998) and Mikhail et al. (1999) argue that earnings announcements with less than 10 forecasts are less accurate. Also, Budescu and Chen (2014) state that the CWM is less accurate when the number of positive forecasters decreases. Therefore, it is adopted, all earnings announcements with less than 10 forecasts are dropped. This involved 250246 forecasts and it is named drop E.

Note, that dropping analysts does not affect the analyst consensus (NA), as for each earnings announcement, IBES provides summary statistics including the analyst consensus. In other words, it does not need to be calculated based on the forecasts that are left in the database.

Table 12. Reduction in sample size per drop

	Sample size
Full	1082683
Drop A	1071888
Drop B	778181
Drop C	589477
Drop D	540060
Drop E	289814

Appendix C

Minimum number of 10 forecasts made per analyst

In the beginning of the sample period $\overline{C_{Jt}}$ is based on few forecasts. Hence, it heavily fluctuates. When the number of forecasts increases, the $\overline{C_{Jt}}$ is becomes more calibrated. Therefore, when the number of forecasts increases CWM is expected to become more accurate. Figure 5 depicts the mean $\overline{C_{Jt}}$ of all analysts per forecast divided by the standard deviation. After 7 forecasts, the $\overline{C_{Jt}}$ fluctuates less. When analysts have made 10 forecasts, the standard deviation is 5.82 times bigger than the mean $\overline{C_{Jt}}$, and after 35 forecasts it is 5.24 times bigger.

Budescu and Chen's (2014) only use analysts in the calculation of CWM who have made a minimum of 10 forecasts. Table 13 depicts statistics of $\overline{C_{Jt}}$ before and after imposing this condition. The 1472 analysts that did not reach 10 forecasts are conditioned away. Furthermore, by imposing a minimum of 10 forecasts, outliers diminish. This can be derived from the decreased standard deviation and range.

Therefore, based on this information, Budescu and Chen's (2014) condition will be adopted in this thesis, as it aligns well with the sample. By imposing this condition, over the whole sample period, 607 earnings announcements will be dropped.

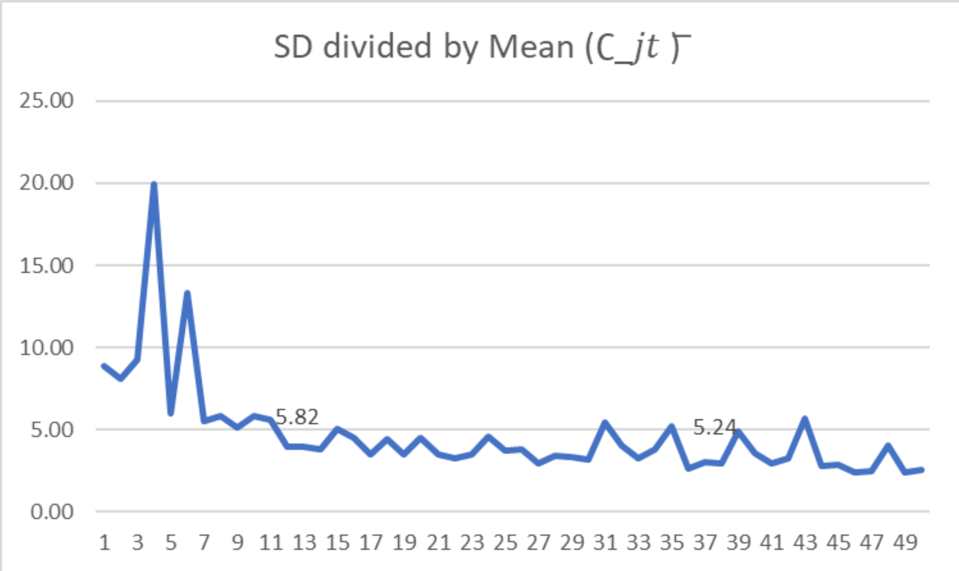

Figure 5. Fluctuation $\overline{C_{Jt}}$ per forecast

Table 13. $\overline{C_{Jt}}$ before and after imposing condition

	Obs	Mean	Std. Dev.	Min	Quartile 1	Median	Quartile 3	Max
$\overline{C_{jt}}$ per analyst (before)	4510	693	6777	-79573	0	162	672	414581
$\overline{C_{jt}}$ per analyst (after)	3038	783	2507	-45477	62	322	985	64380
Difference	-1472	90	-4270	34096	62	160	313	-350201

Minimum number of 5 positive contributors per earnings announcement

Budescu and Chen (2014) found in one of their studies that accuracy of the CWM compared to the AC deteriorates when the number of positive contributors per event drop below 3.

For this sample, Figure 6 depicts the effect on accuracy of the CWM compared to the AC when the number of positive contributors per earnings announcements changes. When this number drops below 5, CWM is less accurate than the AC.

Table 14 provides the datapoints of figure 6. N denotes number of positive contributors per earnings announcement. N EA denotes number of earnings announcements with N positive contributors. For instance, there are 442 earnings announcements with 5 positive contributors. From table 14 can be derived that, on average, the CAM and CWM perform better than the AC when there are 5 or more positive contributors per earnings announcement.

Therefore, based on this information, a minimum number of 5 positive contributors per earnings announcements is imposed. Over the whole sample period, 739 earnings announcements will be dropped.

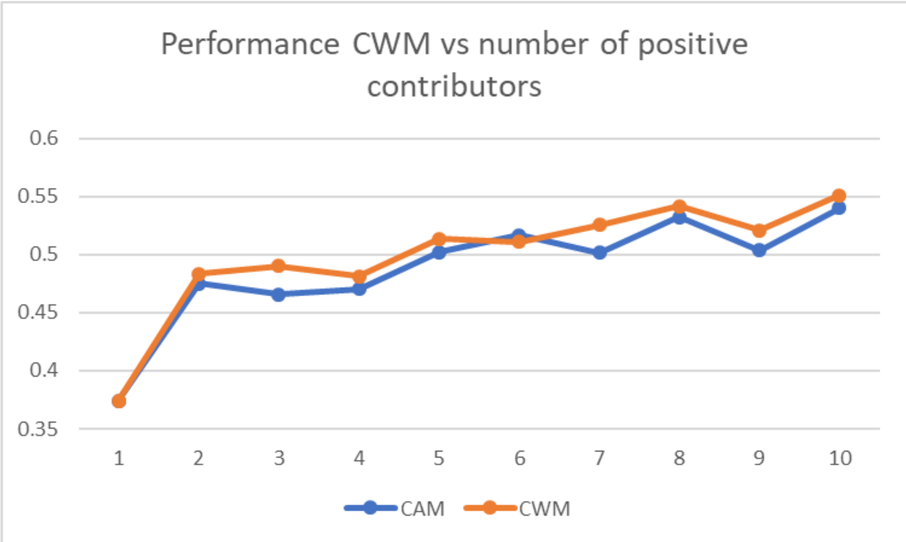

Figure 6. The performance of CAM and CWM conditional on a minimum number of positive contributors per earnings announcement

Table 14. Datapoints of figure 6

N	CAM	CWM	N EA
1	0.37415	0.37415	147
2	0.475	0.483333	120
3	0.465686	0.490196	204
4	0.470149	0.481343	268
5	0.502262	0.513575	442
6	0.516788	0.510949	685
7	0.501545	0.525232	971
8	0.532368	0.541738	1174
9	0.503687	0.520649	1356
10	0.540072	0.550903	1385

Appendix D

Table 15 depicts the various truncation levels and the effect on C_{jit} and $\overline{C_{jt}}$. Positive and negative values are separated. Total positive/negative C_{jit} denotes the sum of C_{jit}. Truncated value denotes the negative or positive value that is truncated away from total. Explains denotes, the percentage of negative or positive value that is truncated away. Truncated N denotes, the number of forecasts that are truncated away. Percentage (C_jt ̅) > 0 denotes the median percentage of positive contributors per earnings announcements.

Striking is that truncating away small percentages of forecasts of between 0.12% and 1.41%, decreases the sum of negative C_{jit} between 20.87% and 56.46%. Similar numbers are found for positive values of C_{jit}. Furthermore, the median number of positive contributors does not change much for different truncation levels. Nonetheless, the percentage of positive contributors only slightly decreases.

Table 15. Effect of truncating on C_{jit} and $\overline{C_{jt}}$

	0.5 SD	**1 SD**	**1.5 SD**	**2 SD**	**2.5 SD**	**3 SD**
Total negative C_jit	-156244076	-156244076	-156244076	-156244076	-156244076	-156244076
Truncated value	-88217417	-64679391	-51556352	-42373427	-36930171	-32608129
Explains	0.5646	0.4140	0.3300	0.2714	0.2364	0.2087
Truncated N	4099	1769	1013	639	468	358
Percentage of total N	0.0141	0.0061	0.0035	0.0022	0.0016	0.0012
Total positive C_jit	365751547	365751547	365751547	365751547	365751547	365751547
Truncated value	229177231	175817397	147019898	126301498	110864715	96741741
Explains	0.6266	0.4807	0.4020	0.3453	0.3031	0.2645
Truncated N	9569	4232	2565	1728	1246	884
Percentage of total N	0.0330	0.0146	0.0089	0.0060	0.0043	0.0031
Percentage (C_jt ̅) > 0	0.808	0.806	0.802	0.800	0.800	0.796

Table 16 depicts the effect of truncating 0.5 and 3 standard deviations of C_{jit} from 0. The level of absolute and relative mistakes for AC, CAM, and CWM are given. Notice that the mean relative absolute mistake does not change. The mean absolute mistake for the different models does change, but the analyst consensus changes accordingly. Lastly, the CAM and CWM do not become considerably more accurate than the analyst consensus for different truncation levels.

Table 16. Effect of truncating on absolute and relative absolute mistake of NA, CAM, and CWM

	No truncation			0.5 SD			3 SD		
	AC	**CAM**	**CWM**	**AC**	**CAM**	**CWM**	**AC**	**CAM**	**CWM**
Absolute mistake per EA									
Mean	**56413**	**50430**	**53591**	**49335**	**46748**	**46487**	**55382**	**50738**	**50666**
Standard deviation	524714	461317	470023	231764	231204	231076	519108	516255	515881
Skewness	90	88	84	25	25	25	94	95	95
Kurtosis	10027	9684	8956	993	1005	1007	10624	10776	10771
Relative absolute mistake									
Mean	**0.34**	**0.28**	**0.29**	**0.34**	**0.28**	**0.28**	**0.34**	**0.28**	**0.28**
Standard deviation	1.45	1.09	1.08	1.45	1.10	1.08	1.45	1.11	1.08
Skewness	26	22	21	26	21	21	26	21	21
Kurtosis	1039	785	762	1067	739	762	1046	725	764
More accurate than AC	N/A	**0.54**	**0.55**	N/A	**0.55**	**0.57**	N/A	**0.55**	**0.56**
N earnings announcements	17019	17019	17019	16651	16651	16651	16750	16750	16750

Appendix E

The boxplot and histogram in the figures 7 and 8 depict the distribution of the paired difference in mistakes between the CWM and CAM and CWM(X). Based on these figures, it is assumed that the normal distribution requirement of the Paired t-test is satisfied.

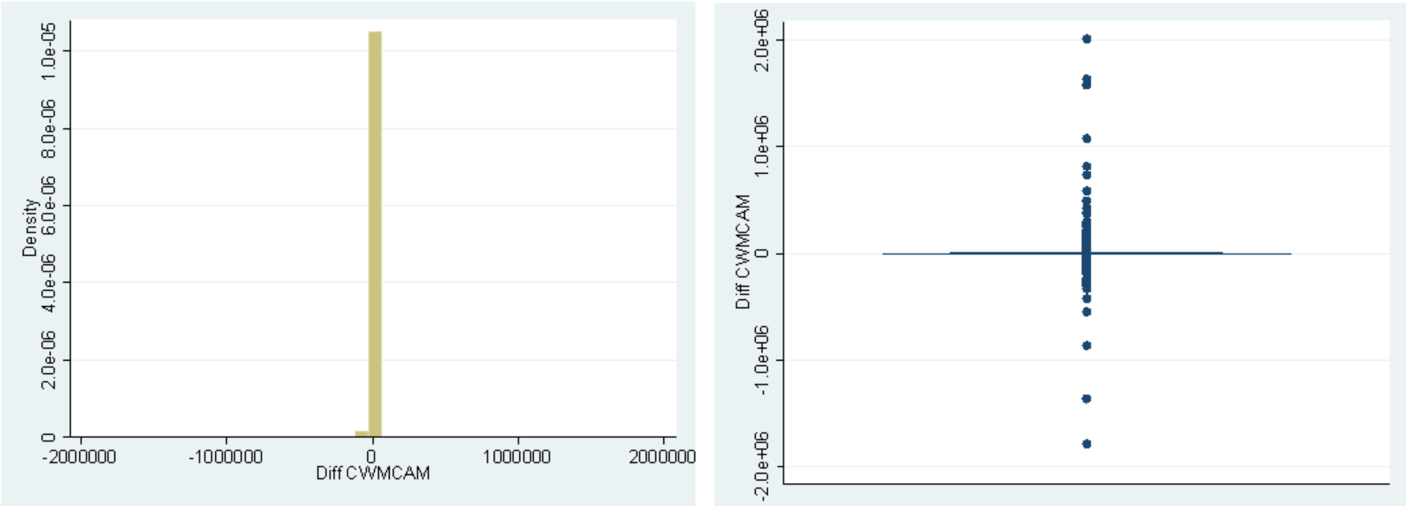

Figure 7. Histogram and boxplot of $X_{it(CWM)} - X_{it(CAM)}$

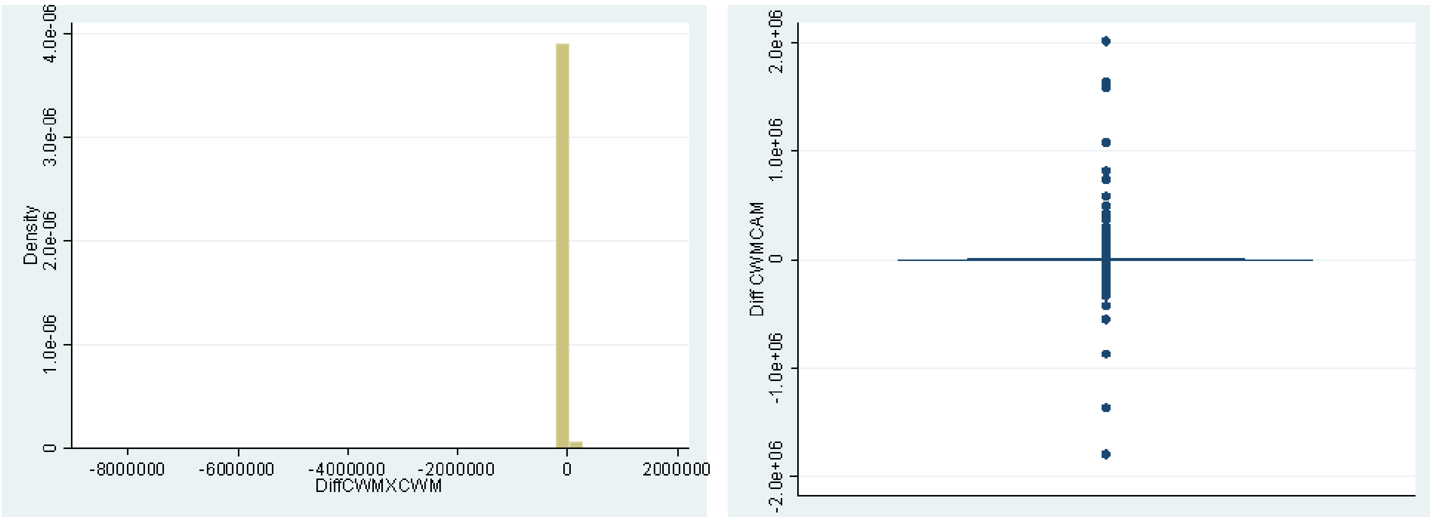

Figure 8. Histogram and boxplot of $X_{it(CWM)} - X_{it(CWM(X))}$

Appendix F

Below, the expected effects of the variables of the regression of hypothesis 1d to 1j are given.

Increases in 'MeanC' are expected to decrease 'CWM-AC', so a negative relation is expected. 'MeanC' increases when $\overline{C_{Jt}}$ of positive contributors increases compared to negative contributors, or when $\overline{C_{Jt}}$ of negative contributors decreases compared to positive contributors.

Decreases in 'PastAccuracy' are expected to decrease 'CWM-AC', thus a positive relation is expected. 'PastAccuracy' decreases when $\overline{X_{Jt}}$ of positive contributors decreases compared to negative contributors, or when $\overline{X_{Jt}}$ of negative contributors increases compared to positive contributors.

Increases in 'NForecasts' are expected to decrease 'CWM-AC', hence a negative relation is expected. The more calibrated the model will be, the higher accuracy is expected.

A negative relation is expected for 'PosConEA' because increases in 'PosConEA' are expected to decrease 'CWM-AC'. On the contrary, for 'NegConEA' the effect is uncertain, both increases and decreases can decrease 'CWM-AC'.

An increase in 'ContributionsJ' is expected to be associated with a decrease of 'CWM-AC'. Therefore, a negative relation is expected. 'ContributionsJ' increases when the ratio of positive contributions for positive contributors compared to negative contributors increases, or when the ratio of negative contributors compared to positive contributors decreases.

It is expected that when 'IndExpertise' and 'ComExpertise' decrease, 'CWM-AC' decreases. Therefore, a positive relation is expected. 'IndExpertise' and 'ComExpertise' decrease when either positive contributors' mean number of companies/industries followed (i.e. expertise increases) decreases relative to negative contributors, or when the mean number of companies or industries followed increases of negative contributors relative to positive contributors.

It is expected that increases in 'GenExperience' and 'ComExperience' decrease 'CWM-AC'. Hence, a negative relation is expected. Either an increase in experience of positive contributors compared to negative contributors, or a decrease in experience of negative contributors compared to positive contributors increases 'GenExperience' and 'ComExperience'.

A negative relation is expected for 'Size' and 'Following' because an increases is expected to decrease 'CWM-AC'. A positive relation is expected because a decrease in 'Age' is expected to decrease 'CWM-AC'.

Appendix G

Four assumptions are tested: misspecification error of the model, heteroskedasticity, normality of residuals, and autocorrelation.

A Ramsey-reset test was used to test for misspecification of the functional form of the model. The test reported a F-statistic of 149.29 with a p-value of 0.000. Therefore, there is enough evidence to reject the model that there are no missing non-linearities.

For heteroskedasticity, two tests were performed. The Cameron and Trivedi decomposition reported a chi-squared of 1728 and p-value of 0.0000. The Breusch-Pagan/ Cook-Weisberg test reported a chi-squared of 140147 and a p-value of 0.0000. This means that enough evidence is found to reject that the variance of the error term is constant.

The normality of the residuals is tested with the Skewness-Kurtosis (Jarque-Bera) test. First, residuals were predicted. Next, the Skewness-Kurtosis test was used with these fitted residuals. The observed p-value is 0.0000 for both skewness and kurtosis. Therefore, there is enough evidence that the non-normality of residuals is not satisfied.

Autocorrelation is tested using the Variance Inflation Factors (VIF), a very conservative threshold is 2 (Dormann et al., 2013), and a less strict is 4 (O'Brien, 2007). In this thesis, 1 variable is above 4 (i.e. 4.38), and 3 others above 2. Therefore, there is little evidence of autocorrelation.

Amending the data

It was tested whether changing the specification form (i.e. adding a squared form and log) for the explanatory variables, but this did not provide much improvement. Moreover, changing the way variables were computed. Specifically, rather than deducting the means of all analysts from the means of positive contributors (i.e. as described in the methodology section); the sum of positive contributors was tried, the sum of all analysts, the mean of positive contributors, and the mean of all analysts. None of which improved the model. Furthermore, adding variables was disregarded because these would not be based on theory relevant to this thesis and also that would not explain why the theorised variables have a low linear relation. Lastly, outliers of the dependent variable were removed but this also did not improve the model.

Appendix H

To create a proxy for actual earnings surprises, two new conditions are imposed to identify companies for which the CWM performs best:

1) A minimum of 70% accuracy of the CWM and CAM: in hypothesis 1, the CWM and CAM are compared to the NA. For some companies, the CWM and CAM perform better than the NA X% of earnings announcements. A percentage of 70% is arbitrarily imposed. Both the CWM and CAM should be more accurate than the NA, 70% of earnings announcements of company (f). The reason that both must be more accurate rather than one is that underlying factors could be at play when the CWM and CAM do not have a similar level of accuracy. For example, an underlying factor could be that there are only one or a few positive contributors with excessively large $\overline{C_{Jt}}$ for earnings announcement (i) at time (t). For these earnings announcements, too much weight is put on one or a few analysts, giving unreliable results for the CWM. In other words, requiring that the accuracy of the CWM and CAM both are above 70% accuracy over the NA, mitigates these underlying factors. This is called drop 70%.
2) A minimum of 8 earnings announcements for each company: A limited number of earnings announcements gives unlikely results. There are companies for which the percentage of accuracy of the CWM and CAM is based on one or a few earnings announcements. These percentages are likely to be unreliable and subject to change. Hence, a minimum of 8 earnings announcements is required which equals 2 years of earnings announcements. This should mitigate unreliable percentages. This is called drop minimum 8.

Drop 70%: from 1122 companies to 172 companies, and from 15414 earnings announcements to 2016 earnings announcements.

Drop minimum 8: from 172 companies to 88 companies, and from 2016 earnings announcements to 1692 earnings announcements.

Appendix I

In the tables below, correlation coefficients and their significance are provided for imposing two extra conditions. The first condition is that the company should have met the earnings target (i.e. the analyst consensus) of previous earnings announcement. This is in line with Kasznik and McNichols (2002), who found that investors reward companies with a return premium when they consecutively meat the earnings consensus, but also that missing it the next time is penalised. Therefore, it is assumed that when the earnings benchmark of previous earnings announcement is met, subsequent stock market responses to earnings announcements are more often in the hypothesised direction. The second condition is derived from Myers et al. (2007). They state that when companies have subsequent growth in earnings, a premium is observed, and penalised when they do not meet it after a string of growth. Hence, the same reasoning applies for this condition. It is expected that stock market responses are more often in the hypothesised direction.

Therefore, table 17, depicts meeting previous earnings benchmark, table 18 meeting previous growth benchmark, and table19 meeting both benchmarks. The results are similar to the correlations without the extra benchmarks. The correlation between actual earnings surprises and CAR is significant. The correlation between predicted earnings surprises and CAR are in most cases not significant. Only the Pearson correlation of the growth condition is significant. However, it is quite different from the Spearman correlation coefficient and is therefore probably biased. Likewise, the Pearson correlation between actual earnings surprises and CAR are higher than the Spearman correlation coefficient. Therefore, caution should be taken when interpreting these results.

Table 17. Pearson correlation (left) and Spearman correlation (right), meeting previous AC, 70% top performers

	CAR	Actual ES	Predicted ES
CAR	1		
Actual ES	0.2138	1	
	***0.0000		
Predicted ES	0.0517	0.6488	1
	0.1294	***0.0000	

	CAR sign	Actual ES sign	Predicted ES sign
CAR sign	1		
Actual ES sign	0.1592	1	
	***0.0000		
Predicted ES sign	0.0317	0.6417	1
	0.3527	***0.0000	

** denotes significant at 10% significance level. ** denotes significant at 5% significance level. *** denotes significant at 1% significance level.*

Table 18. Pearson correlation (left) and Spearman correlation (right), growth previous AC, 70% top performers

	CAR	Actual ES	Predicted ES
CAR	1		
Actual ES	0.2383	1	
	***0.0000		
Predicted ES	0.1142	0.6958	1
	***0.0007	***0.0000	

	CAR sign	Actual ES sign	Predicted ES sign
CAR sign	1		
Actual ES sign	0.1427	1	
	***0.0000		
Predicted ES sign	0.0378	0.7007	1
	0.2653	***0.0000	

*denotes significant at 10% significance level. ** denotes significant at 5% significance level. *** denotes significant at 1% significance level.*

Table 19. Pearson correlation (left) and Spearman correlation (right), both previous AC, 70% top performers

	CAR	Actual ES	Predicted ES
CAR	1		
Actual ES	0.2186	1	
	***0.0000		
Predicted ES	0.1020	0.7059	1
	***0.0154	***0.0000	

	CAR sign	Actual ES sign	Predicted ES sign
CAR sign	1		
Actual ES sign	0.1357	1	
	***0.0012		
Predicted ES sign	0.0221	0.6544	1
	0.5998	***0.0000	

*denotes significant at 10% significance level. ** denotes significant at 5% significance level. *** denotes significant at 1% significance level.*