

Are Analysts Still Overconfident ?

Student ID 

Department of Economics
The University of Warwick
March 2014

Abstract: This paper analyzes the effect of overconfidence on the performance of analysts using their forecasts on the earnings for following quarter. The results show that analysts forecast performance only depends on publicly available information rather than private information. This result is not due to analysts' completely ignoring their private information but there is an overconfidence story on behind. Analysts having accurate forecasts for some consecutive periods gain too much overconfidence that they ignore public information and underperform in their next forecast. After realizing his/her poor performance the analyst lost his/her overconfidence. In the following periods they ignore this time their private information. This cyclical to opposing process follow each other and overall cancel each other.

Content

1. Introduction.....	3
2. Theoretical Foundations and Related Literature.....	3
3. Empirical Design.....	5
4. Empirical Results.....	8
5. Conclusion.....	9
6. Tables.....	11
7. References.....	15

“There is considerable evidence that people are more likely to arrive at the conclusions that they want to arrive at, but their ability to construct seemingly reasonable justifications for these conclusions.” (Kunda, 1990)

1. Introduction

One of the most important agents of financial markets are analysts. How financial analysts do their forecasts is an important issue since these forecasts are a huge and actual natural data set for the real decision making by individuals.

The two important cognitive biases play a key role during the decision making process of analysts for their forecasts, which are overconfidence and self attribution. These are short term/temporary dynamics. In this paper, predictions of individual analysts, specifically the accuracy *and* deviation of analysts’ forecast from the median are studied.

The results show only public information is significant in explaining the forecast accuracy of the analysts namely only *CMD* and *CME* is significant in the regressions. One possible explanation is that accuracy of predictions in short term, periodically, lead to increase the possibility of inaccurate predictions for the following period. After having accurate predictions for some consecutive periods analysts become overconfident. So that in the next periods they underperform by deviating too much since they use their private information overconfidently. And when they underperform due to the increased overconfidence an opposite force comes into the picture and analysts this time do not pay much attention to their private information but mainly use public information. Finally these two opposing forces follow each other successively and in the end they cancel each other in the overall.

The results are not consistent with the hypothesis that there is a positive relationship between analysts’ predictions and overconfidence.

The rest of the paper is constructed as follow. Section two discusses theoretical foundations and related literature. Section three presents the empirical design, and in section four empirical results are discussed. Finally, section five concludes.

2. Theoretical Foundations and Related Literature

Self-Serving Attribution

People tend to exaggerate their own abilities, self-attribution bias refers to situations where good outcomes are attributed to skill, whilst bad outcomes are attributed to bad luck (Montier,

2002). And it is one the reasons of being optimistic, which leads to do inferior predictions. Therefore, Barber and Odean (2001) stated that analysts overconfidence results from self-attribution bias and they infer their own abilities from their successes and failures.

Overconfidence

Huck and Oechssler (2000) stated that subjects primarily prefer the heuristic “follow your (private) signal” rather than choosing according to Bayes’ rule. Therefore, overconfident investors, who believe that the accuracy of their knowledge about the value of an asset is higher than its actual value, trade more than others and that leading to decrease their expected utilities (Odean 1999).

Hilary and Menzly (2006) and Kraemer (2006) claims that from the existing literature, “this lead them to put excessive weight on their private information and discount public signals, such as market reactions and other analysts’ forecast”. After having a set of good estimations by analysts, overconfidence situation happens in the analysts’ ability to estimate the future earnings. Hence, the possibility of inaccurate forecasts of next periods is likely to increase *and* deviation from the consensus is likely to be more than the lack of this cognitive biasedness.

HM states that overconfidence resulting from previous accurate forecasts reduces the likelihood for their next forecast to be superior to the other analysts’ and, in turn, the next resulting less accurate prediction may trigger a negative feedback mechanism that reduces overconfidence. Hence, this overconfidence phenomenon may repeatedly emerge and disappear. Therefore, the success in previous predictions induces the emerging of overconfidence, which, then results in some increases in the possibility of bad predictions for the future and so causes to some reductions for overconfidence. In addition, analysts with overconfidence underperform as a result of overestimation when compared to these analysts’ own expected performance.

Previous literature mainly focused on long term changes by using yearly data such as Hong et al (2000). However, in this research, it is focused on short term periodic changes by employing quarterly data as following (Hilary and Menzly 2006) rather than long term.¹

¹ The four quarter period is preferred rather than monthly and yearly period, because of the reasons as mentioned in Hilary and Menzly (2006). Firstly, reducing the period too much such as monthly, may not allow us capture the sequence of the analysts’ forecast. Secondly, extending the period too much such as year by year, it might demonstrate dynamics in long run subsumed by control variables.

As a final word, HM (2006) studied overconfidence issue before the global financial crises. To my knowledge, previously no paper has examined overconfidence issue in this context including global financial process.² Indeed, the crises has not finished yet, therefore, only six years, twenty four quarters are taken.

3. Empirical Design

The data, definition of independent variables and methodology are explained to empirically examine the two hypothesis. One is that a higher tendency of deviation from the consensus, due to short term/temporary success *and* precision of forecasts by individual analysts.

3.1 Data

In order to retrieve the raw analyst forecast data, benefited from I/B/E/S dataset (Institutional Brokers Estimate System) which is accessed by WRDS, and randomly 14 firm were selected in S&P 100 Index, USA. And the period of the data is from third quarter of 2007 to third quarter of 2013, since this study focuses just after the global financial crises. To capture the effect also during the crisis the data used starts at 2007 as the early signal of the GFC.

The selection process of analysts' forecast is similar to Hilary and Menzly (2006) in terms of following four requirements. Firstly, this study concentrates on forecasts made by analysts for the next quarter. The firms are excluded which do not have a December year-end report. Secondly, if there are more than one prediction for each analyst, the earliest forecast is taken into account for each firm in each quarter, since I want to isolate the effects over the estimations such as different distribution, arising due to updated information and news. Thirdly, as also employed in HM, originally by Easterwood and Nutt (1999), in order to have a meaningful measure of the consensus, only the firms covered at least with four analysts for a particular quarter. Finally, analysts having at least four forecasts are chosen in this sample, for a given company in the preceding quarters.

3.2 Definition of Dependent Variables

In the light of the stated framework above, two different regressions are used to understand the effect of short-term/temporary past success of the analyst captured by both the deviation

² There is a related study which (Abbes, 2013) examined this kind of issue about same period, however, the context and methodology is different from this study.

from the consensus and the error of the forecast. In order to test these regressions, two distinct variables are determined. Firstly:

$$DEV_t^i = \alpha + \beta FREQ_t^i + \gamma_1 CME_t^i + \gamma_2 CMD_t^i + \gamma_3 T_t^i + \gamma_4 LOSS_t^i + \gamma_5 ACCURACY_t^i + \gamma_6 MCAP_t^i + \varepsilon_t^i$$

DEV is defined as the deviation from the consensus, which can be described as the difference between the consensus forecast and the individual analyst forecast *i* in absolute terms. The consensus is computed as the median forecast for each company and for each quarter *t*.³ *DEV* captures the propensity of an analyst for discounting public signals/information as a proxy. Secondly:

$$ERROR_t^i = \alpha + \beta FREQ_t^i + \gamma_1 CME_t^i + \gamma_2 CMD_t^i + \gamma_3 T_t^i + \gamma_4 LOSS_t^i + \gamma_5 ACCURACY_t^i + \gamma_6 MCAP_t^i + \varepsilon_t^i$$

ERROR is defined as the forecast error, which can be described as the difference between the value of the distinction between forecasted and actual of earnings in absolute terms. Earnings are taken as EPS (earnings per share) from the database.

3.3 Methodology

Assuming that a short period of successive good estimations make analysts more overconfident. There are three success factors. The first of them *FREQ* is produced as a discrete variable (0 to 4) which is described as the number of superior estimations in the previous four quarters. The superior estimation is a prediction that his or her individual error is smaller than the absolute value of the median error. Overconfidence of subjects can be increasing or decreasing as a result of superior and inferior predictions in the last four quarters.

After making one inferior estimation, overconfidence of the financial analysts may disappear, however, overconfidence can build up again uniformly with each superior estimation.

³Because of the biases from extreme observations, it is preferred to use the median rather than mean.

Alternatively, *STREAK 1* as a discrete variable (0 to 4), is calculated as the number of successive good estimations before the present forecast is made.⁴

The next one is *STREAK 2* shows how many successive quarters the financial analyst had a bad forecast. The combination of a specific firm and analyst is considered, as panel entity, precisely assuming each critique of different analyses of an analyst.

As mentioned above, three alternative versions are used in terms of past short success, including, *FREQ*, *STREAK 1*, *STREAK 2*. *FREQ* is used to capture the frequency of superior estimations whereas *STREAK 1* and *STREAK 2* captures the length of good and bad predictions, respectively. The other variables are defined below.

The median of all financial analysts' error for each company and current quarter is defined as *CME*. Earnings shocks and other companies' shocks are expressed by this Current Median Error variable. Likewise, the median of deviation of the financial analysts, for each company and current quarter, is described as *CMD*. The heterogeneity in financial analyst judgement about future earnings is measured by this variable, Current Median Deviation. The possible sign of the *CME* in the *ERROR*, regression and *CMD* in the *DEV* regression are expected to be positive.⁵

Additionally, in order to proxy for the set of information in terms of size, time is produced as a variable, *T*. It captures the number of days are from forecast date to the end of quarter. This *T*'s sign is expected to be negative in regressions, *DEV* and *ERROR*.

HM mention that, in the literature (e.g. Abarbanell and Lehavy, 2003) it is found that firms with negative profits may have an upwardly biased consensus forecasts. Thus losses may explain some anomalies that are pointed out in the literature. *LOSS* in the present quarter is produced as a dummy variable so as to manage this possible impact. A positive sign is expected for that in the second regression.

Because of the analyst's skill and tendency about deviation from the consensus are possible significant features. Hence, *ACCURACY* is also included in these cross-sectional regressions, which is the median distinction throughout the whole period of range for each analyst. It is

⁴Individuals tend to be more overconfident than being underconfident. So they do not underweight their failures as much as they overweight their successes (Fiske and Taylor, 1991). As stated in HM, it is supposed it is higher than the median performance and not less than the median performance.

⁵The predicted signs are taken as HM stated.

calculated as the difference between the absolute value of analyst error and median error of the analysts for each firm and quarter. Negative sign is expected for *ACCURACY* in *ERROR* regression.

As it has been showed in previous studies, there is a correlation between the size of companies and many factors. For these potential omitted variables, *MCAP* is introduced as the market capitalization for each company in order to proxy the size effect. *MCAP* may also proxy for the public information for each specific company. In both regressions, *MCAP* is expected to be negative.

The fixed effect method, for this panel data, is used to estimate coefficients in the regressions. Thus, the mean of every variable, which is computed for each combined company and analyst, is subtracted before starting to employ the regressions. Furthermore, the standard errors are groupwise heteroscedasticity consistent after clustering by company. Meaning that we have independency across groups (companies).

In addition, this model is especially appropriate for the hypotheses that I want to test. Namely whether overconfidence results with deviation of forecasts from the median and increased forecast error, focusing on the short term dynamics of financial forecasting. One of the drawbacks of the fixed effect model is that the variation in cross-section cannot be utilized in the estimation, since this variation is lost during the process.

4. Empirical Results

Table 1 presents descriptive statistics, the mean and standard deviation for both dependent and independent variables. As a result of using absolute value for both *DEV* and *ERROR* as dependent variables, and *CMD* and *CME* as independent variables, their means are positive. In terms of *T* (time), the average number of making the earliest forecast is 65 days prior to the last day of the quarter, in order to be forecasted. *ACCURACY*, the difference between the absolute value of analyst error and median error of the analysts for each firm and quarter, in the mean is expected, zero.

From unconditional means in Table 1 there is no significant. as the number of superior estimations in the previous four quarters.

In Table 2, there is no specific pattern (such as monotonically increasing or decreasing) for *DEV* and *ERROR* for conditional means. Indeed, I was expecting that there will be a

monotonic increase in the conditional means of *DEV* and *ERROR* regressions, as the value of *FREQ* or *STREAK 1* or *STREAK 2* increases (0 to 4). Although some increases, in general, there is no result as expect. The size of the data is could be one of the reasons.

Table 3a, fixed effect regression results of *DEV*, shows that only *CMD* is significant among control variables for all three specifications; *FREQ*, *STREAK 1* and *STREAK 2*. As expected *CMD*, the median deviation of all analysts affects *DEV* positively. The effect is approximately 1 to 1 (the coefficient is 1.05) which shows that the change in the consensus of the analysts is directly transferred to the individual analyst's deviation. This is a quite expected result which may be interpreted as all analysts use the available public information (which is proxied by *CMD* variable) in the same manner.

Having only *CMD* as significant may be due to not having big enough number of analysts in our data set. Since data is constructed from the raw data, due to time limitation I could include 14 firms in this study, indeed after cleaning the data and dropping some of the companies as explained in the data section, I left with 9 companies and totally 90 analysts. Even I have very few companies since I have 90 analysts it seems reasonable enough of individuals since we have totally 1710 number of observations.

Table 3b, fixed effect regression results of *ERROR*, shows that only *CME* is significant in all three specifications. Again as expected its sign is positive. Also *CME* effects the *ERROR* of an individual analyst approximately 1 to 1 (the coefficient is 0.95) indicating that any change in consensus affects the forecast error of an individual analyst directly. The interpretation of this result is again similar for the other regression. *CME* also proxies the public information available to all analysts and all analysts use the public information in similar way.

Furthermore, *ACCURACY* is omitted in the models since it suffers from multicollinearity problem in the regressions.

5. Conclusion

These results namely only having *CMD* in *DEV* regression and *CME* in *ERROR* regression suggest that on average only public information matters for the forecast precisions of the analysts. The variables proxying private information all turned out to be insignificant in all specifications of both regression models. Even private information is used during the forecast

process, since overconfidence fluctuates depending on whether the analyst perform an superior or inferior contrast in total the effect of private information cancels out.

The reason is that analysts start to ignore public information and put too much weight in their private information after performing good forecasts resulting with overconfidence and this decreases their forecast precision meaning that they will have a bigger deviation and error. This process is reverted when the analysts realize that they have poor performance so in that case they put more weight on public information and start to ignore their private information. This process like a feedback mechanism continuous in a cyclic process. In the end private information seems to have no effect on forecasting since the opposing effects explained above cancel out each other.

6. Tables

Table 1 **Descriptive Statistics**

VARIABLE	MEAN	STATNDARD DEVIATION
DEV	0.0013264	0.0028235
ERROR	0.0057355	0.0107422
CME	0.0058281	0.0108493
FREQ	1.826484	1.160372
STREAK 1	0.9196347	1.260759
STREAK 2	0.9351598	1.297726
CMD	0.0010311	0.0013677
T	65.0784	8.930656
MCAP	89.23478	117.2392
LOSS	0.0382862	0.1919739
ACCURACY	0	0

Note: This table presents descriptive statistics. *The two dependent variables are DEV and ERROR.* The former is the deviation from the consensus, which can be described as the difference between the consensus forecast and the individual analyst forecast in absolute terms, whereas the latter is the forecast error, which can be described as the difference between the value of the distinction between forecasted and actual of earnings in absolute terms. The three succes are that *FREQ* is used to capture the frequency of superior estimations whereas *STREAK 1* and *STREAK 2* captures the length of good and bad predictions, respectively. The other independent variables are *CME*, current median error of all analysts, *CMD*, current median deviation of all analysts, *T* (time) captures the number of days are from forecast date to the end of quarter, *MCAP*, market capitalization. *LOSS*, loss in the current quarter is a dummy variable. Finally, *ACCURACY* is the median distinction throughout the whole period of range for each analyst, however, is omitted in the models since it suffers from multicollinearity problem in the regressions.

Table 2 Conditional Means of *DEV* and *ERROR* on Success Factors (*FREQ*, *STREAK 1* and *STREAK 2*)

SUCCESS FACTOR/ DEPENDENT VARIABLE	MEAN OF SUCCESS FACTOR	DEV	ERROR
FREQ	0	0.0011102	0.0053503
	1	0.0012039	0.0061476
	2	0.0012967	0.0054154
	3	0.0016483	0.0062245
	4	0.0013684	0.0050815
STREAK 1	0	0.0012868	0.0058404
	1	0.0011421	0.0050533
	2	0.0013968	0.0058473
	3	0.0021736	0.0079493
	4	0.0013684	0.0050815
STREAK 2	0	0.0012571	0.0055614
	1	0.0013878	0.006044
	2	0.001521	0.0054737
	3	0.0016764	0.0073452
	4	0.0012147	0.0056709

Note: In this table, Column 2 shows the mean of success factors, *FREQ*, *STREAK 1* and *STREAK 2* as discrete variables, from 0 to 4. Column 3 presents the conditional means of *DEV* as a dependent variable on success factors, whereas, Column 4 presents the conditional means of *ERROR* as a dependent variable on success factors. *FREQ* is used to capture the frequency of superior estimations whereas *STREAK 1* and *STREAK 2* captures the length of good and bad predictions, respectively.

Table 3a Fixed Effect Regressions of *DEV*

VARIABLE/ DEPENDENT VARIABLE	DEV		
FREQ	0.0000503 (0.0001022)		
STREAK 1		0.0000344 (0.0000769)	
STREAK 2			0.0000478 (0.0000371)
CME	0.0302597 (0.0384016)	0.0300871 (0.0385389)	0.0299313 (0.0385759)
CMD	1.0521*** (0.0886186)	1.052667*** (0.0875732)	1.050864*** (0.0883427)
T	-0.0000324 (0.0000228)	-0.0000323 (0.0000227)	-0.0000319 (0.0000229)
MCAP	-2.22e-07 (2.69e-07)	-1.93e-07 (3.18e-07)	-2.26e-07 (2.63e-07)
LOSS	0.0002415 (0.00049)	0.0002404 (0.0004942)	0.0002431 (0.0004925)
ACCURACY	0 (omitted)	0 (omitted)	0 (omitted)
CONSTANT	0.0020941 (0.0013028)	0.0021449 (0.0013383)	0.0021137 (0.0014466)
R-Squared	0.3284	0.3281	0.3280

*** Significant at the one percent level.

Note: This table presents the fixed effect regression results of *DEV*, this dependent variable is defined as the deviation from the consensus, which can be described as the difference between the consensus forecast and the individual analyst forecast in absolute terms. The consensus is computed as the median forecast for each company and for each quarter. Apart from *CMD* as independent variables, are not significant. It is the median of deviation of the financial analysts, for each company and current quarter. The heterogeneity in financial analyst judgement about future earnings is measured by this variable, Current Median Deviation. In addition, *ACCURACY* is the median distinction throughout the whole period of range for each analyst, however, is omitted in the models since it suffers from multicollinearity problem in the regressions. Column 2, 3 and 4 show the result after controlling for *FREQ*, *STREAK 1* and *STREAK 2* which is used to capture the frequency of superior estimations, whereas captures the length of good predictions, and captures the length of bad predictions, respectively. The data is drawn from I/B/E/S dataset (Institutional Brokers Estimate System) which is accessed by WRDS, the period is between quarter three (2007) and quarter three (2013). The number of observation is 1710. R squared is roughly 33 percent. Standard errors are reported in brackets.

Table 3b Fixed Effect Regressions of *ERROR*

VARIABLE/ DEPENDENT VARIABLE	ERROR		
FREQ	0.0000899 (0.0001006)		
STREAK 1		-0.0000676 (0.0000632)	
STREAK 2			0.0000437 (0.0000208)
CME	0.9508969*** (0.0337514)	0.9501886*** (0.0335823)	0.950353*** (0.0332697)
CMD	1.1142619 (0.0655697)	1.1109009 (0.0650903)	1.1126086 (0.066484)
T	0.0000326 (0.0000265)	0.000033 (0.0000262)	0.0000332 (0.0000261)
MCAP	6.64e-07 (3.73e-07)	5.27e-07 (3.73e-07)	6.41e-07 (3.52e-07)
LOSS	0.0000729 (0.0003847)	0.0000877 (0.0003888)	0.0000775 (0.0003847)
ACCURACY	0 (omitted)	0 (omitted)	0 (omitted)
CONSTANT	-0.0020941 (0.0013028)	-0.0020543 (0.0015591)	-0.0021784 (0.0016148)
R-Squared	0.9410	0.9420	0.9417

*** Significant at the one percent level.

Note: This table presents the fixed effect regression of *ERROR*, this dependent variable is the forecast error, which can be described as the difference between the value of the distinction between forecasted and actual of earnings in absolute terms. Apart from *CME* as independent variables, are not significant. It is The median of all financial analysts' error for each company and current quarter. Earnings shocks and other companies' shocks are expressed by this Current Median Error variable. In addition, *ACCURACY* is the median distinction throughout the whole period of range for each analyst, however, is omitted in the models since it suffers from multicollinearity problem in the regressions. Column 2, 3 and 4 show the result after controlling for *FREQ*, *STREAK 1* and *STREAK 2* which is used to capture the frequency of superior estimations, whereas captures the length of good predictions, and captures the length of bad predictions, respectively. The data is drawn from I/B/E/S dataset (Institutional Brokers Estimate System) which is accessed by WRDS, the period is between quarter three (2007) and quarter three (2013). The number of observation is 1710. R squared is roughly 94 percent. Standard errors in parenthesis.

7. References

- Abbes, M. B., 2013, Does Overconfidence Bias Explain Volatility During the Global Financial Crisis?, *Transtion Studies Review*, 19(3): 291-312.
- Barber, B., Odean, T., 2001, Boys will be Boys: Gender, Overconfidence and Common Stock Investment, *Quarterly Journal of Economics*, 1: 261-292.
- Easterwood, J. C., Nutt, S.R., 1999, Inefficiency in Analysts' Earnings Forecasts: Systematic Misreaction oor Systematic Optimism, *Journal of Finance*, 54(5): 1777-1797.
- Fiske, S., Taylor S, E., 1991, Social Cognition, *McGraw Hill*, New York.
- Hilary, G., Menzly, L., 2006, Does Past Success Lead Analysts to Become Overconfident?, *Management Science*, 52(4): 489-500.
- Hong, H., Kubik, J. D., Solomon A., 2000, Security Analysts' Career Concerns and Herding of Earning Forecasts, *RAND Journal of Economics*, 31(1): 121-144.
- Huck, S., Oechssler, J., 2000, Informational Cascade in the Laboratory: Do They Occur for the Right Reasons, *Journal of Economic Psychology*, 21: 661-671.
- Kraemer, C., Noeth M., Weber M., 2006, Information Aggregation with Costly Information and Random Ordering: Experimental Evidence, *Journal of Economic Behavior & Organization*, 59(3): 423-432.
- Kunda, Z. 1990, The Case for Motivated Reasoning, *Psychological Bulletin*, 108(3): 480-498.
- Montier, J. 2002, Behavioural Finance: Insights into Irrational Minds and Markets, *Wiley*.
- Odean, T., 1999, Do Investors Trade Too Much?, *American Economic Review*, 89(5): 1279-1298.