Aggregating Earnings per Share Forecasts

Greg Slavin
University of Pennsylvania

April 2007

ABSTRACT

Research analyst forecast accuracy has been the focus of many studies, but only Brown and Mohammad (2001) consider different ways of weighting individual forecasts to create consensus estimates of earnings per share. I add to the existing literature by suggesting a technique to reduce forecast errors by 29%, which is 13% more accurate than the method employed by Brown and Mohammad.

1. INTRODUCTION

Earnings per share, or net income divided by outstanding shares, is a widely used measure of firm profitability. Research analysts[1] employed by brokerages such as Goldman Sachs or Morgan Stanley forecast earnings per share for thousands of companies. Investment bankers use these forecasts as inputs into valuation models. Portfolio managers use these forecasts as criteria for security selection. Academics use these forecasts to write papers. Small changes in forecasts often result in significant changes in valuation. Data services such as Thomson Financial or Zack's aggregate individual analyst forecasts to create consensus forecasts. Typically, they publish the mean and median of analyst forecasts. Are these the best estimates of future earnings per share? The basic premise of my paper is that there are better ways of combining forecasts to estimate future earnings per share.

Studies of market association assert that the stock market puts little importance on published consensus forecasts[2]. However, little research has been done on creating more accurate consensus forecasts from analyst forecasts[3]. Two questions arise: How can we predict analyst accuracy? And if we can predict analyst accuracy, what weight do we place on each forecast? Brown and Mohammad (2001) predict individual analyst accuracy using characteristics such as forecast recency, forecast frequency, brokerage reputation, experience, and the number of

[1] Alternatively called "security analysts" or simply "analysts".

[2] Brown and Kim (1991) find that recent forecasts, such as the average of the last three forecasts, are more correlated with market returns than published consensus forecasts.

[3] Research has been done on improving forecast accuracy, though these studies typically rely on stochastic models. For example, Conroy and Harris (1987) find that the combination of a random walk model and the consensus of analyst forecasts improves forecast accuracy over using just a random walk model or consensus forecast.

companies and industries covered. They then combine forecasts using a weight based on each analyst's projected accuracy ranked relative to their peers. I improve on their methodology with two steps. First, I use past accuracy (in addition to their six factors) to predict individual analyst accuracy, which reduces forecast error by 1%. Second, my forecast weight accounts for the magnitude of analyst accuracy, whereas Brown and Mohammad use only ordinal rankings, which reduces forecast error by an additional 12%. Overall, my technique creates consensus forecasts 13% more accurate than Brown and Mohammad and 29% more accurate than current consensus forecasts.

Section 2 discusses prior research, identifying seven analyst characteristics that affect accuracy. Section 3 describes the data and sample selection. Section 4 explains the measurement of variables, showing descriptive statistics and correlations. Section 5 reports the results, showing regressions and consensus forecast errors. Section 6 summarizes and concludes the study.

2. PRIOR RESEARCH

Brown (1993) wrote 26 pages and referenced 174 academic papers (including 16 of his own) in his Magnus opus overview of prior research. I promise mine will be shorter. He concluded that research on forecasted earnings per share can be divided into association and accuracy. Association focuses on market reactions to changes in forecasts; for example, Chen, Chan and Steiner (2002) find that analysts at national brokerage firms are more influential than analysts at regional brokerage firms. Accuracy focuses on examining forecast errors and differences in analyst accuracy. O'Brien (1990) found that there are no systematic differences between analysts. However, Stickel (1992) discovered that All-American analysts (who are recognized by *Institutional Investor* as superior analysts) are more accurate than other forecasters. Sinha, Brown and Das (1997) conclude that O'Brien incorrectly accounted for differences in forecast recency. It is now widely accepted that analyst characteristics affect forecast accuracy. There are seven common characteristics recognized by the literature:

2.1. FORECAST RECENCY

Recent forecasts are more accurate than older forecasts. Recent forecasts have the advantage of incorporating additional information. For example, a forecast of annual earnings per share issued after a firm announces third quarter earnings would know with certainty three-quarters of actual annual earnings per share. Brown (1991) finds that using the most recent forecast, the average of the three most recent forecasts, or the average of all forecasts in the last month results in lower forecast errors than taking the average of all forecasts. Clement (1999) quantifies this relationship by claiming that each day a forecast ages reduces its relative accuracy by 0.35%, or 10.5% per month.

2.2. FORECAST FREQUENCY

Analysts who revise their forecasts frequently are more accurate than other forecasters. Jacob, Lys and Neale (1999) suggest that analysts only revise their forecasts when they have access to new information, making forecast frequency a proxy for information.

2.3. *Number of companies covered*

Analysts who follow many companies are less accurate than other forecasters. Clement (1999) suggests the number of companies followed is a proxy for portfolio complexity. Jacob, Lys and Neale (1999) support Clement, proposing that "the fewer the number of companies analysts follow, the more time per company they can spend refining their forecasts and the more accurate the forecasts should be."

2.4. *Brokerage reputation*

Analysts who work for prestigious brokerages are more accurate than other forecasters. Clement (1999) hypothesizes that premier brokerages provide superior resources to analysts in the form of better data sets and administrative support. Analysts at these brokerages may also have better access to private information from managers at the companies they follow. Finally, analysts at reputed brokerages are typically better compensated; consequently, working for a prestigious brokerage may indicate ability. While it is unclear which effect dominates, it is widely accepted that brokerage reputation effects forecast accuracy.

2.5. *Experience*

Analysts who have more experience are more accurate than other forecasters. Mikhail, Walther and Willis (1997) suggest that analysts "learn by doing", gaining basic skills with repetition and feedback, improving knowledge of idiosyncratic company operations and bettering their relationship with corporate management. Jacob, Lys and Neale (1999) and Clement (1999) disagree with "learning by doing", arguing instead that experience is a proxy for ability since experience is an indicator that the analyst has not been fired due to poor performance. In either case, more experience leads to more accurate forecasts.

2.6. *NUMBER OF INDUSTRIES COVERED*

Analysts who follow many industries are less accurate than other forecasters. Clement (1999) suggests that there are economies of scale to following firms in a particular industry. Jacob, Lys and Neale (1999) agree, suggesting that "industry specialization may allow the analyst to develop a depth of understanding that can provide considerable synergies in forecasting companies within that industry."

2.7. *PAST ACCURACY*

Analysts who were more accurate in the past are more accurate than other forecasters. Abarbanell and Bernard (1992) found forecast errors have positive serial correlation. Ali, Klein and Rosenfeld (1992) agree, contending that "analysts systematically underestimate the permanence of past forecast errors when forecasting future earnings." They find that adjusting for past accuracy improves forecast accuracy by 12%. Markov and Tamayo (2006) suggest that autocorrelation is consistent with rational analysts facing parameter uncertainty and gradually learning about the true values of parameters over time.

3. DATA AND SAMPLE SELECTION

3. 1. DATA

The data for this study were obtained from Thomson Financial's Institutional Broker Estimate System (I/B/E/S), available through Wharton Research Data Services. Each observation has eight variables: (1) firm identifier; (2) date earnings were reported; (3) date forecast was made; (4) forecasted earnings per share; (5) actual earnings per share; (6) brokerage identifier; (7) analyst identifier[4]; and (8) industry identifier for the firm being covered. All earnings per share figures represent "Fiscal Year 1", I/B/E/S's designation for the current fiscal year. The section on data and sample selection that follows is highly technical and detailed; casual readers are advised to skip to Section 4: Variables.

The years of note are 1989 – 2005, despite the data set beginning in 1983. There are three considerations in choosing the date range:

(1) The data set is left censored. Since the data starts in 1983, it is unknown how much experience analysts have in 1984. Clement (1999) and Hong and Kubik (2003) resolve this issue by including only analysts who began their careers after 1983. Alternatively, Clement and Tse (2005) restrict their data to 1989 – 1998. I have adopted Clement and Tse's starting date of 1989, with the acknowledgment that dispersion in experience will increase over time.

(2) Forecasts from the 1980s were less accurate than forecasts from the 1990s and 2000s. Philbrick and Ricks (1991) observe that Value Line forecasts outperformed I/B/E/S in the 1980s. However, Ramnath, Rock and Shane (2005) find that I/B/E/S estimates have improved over time due to competition (from Zack's and First Call). Likewise, Brown (1997) found that the accuracy of forecasts have increased over time. Consequently, I use data beginning from 1989.

---

[4] I was unable to obtain the I/B/E/S broker translation file used by Clement (1999), Lim (2001) and Herrmann and Thomas (2005) to exclude teams of analysts. Thus, analyst code may refer to a group of analysts.

(3) Including data from 2006 would lead to bias. For example, most financial institutions have fiscal year ends in November and have submitted their 2006 earnings by February 2007 (the month of my data collection); conversely, most retailers have fiscal year ends in January and have not submitted their 2006 earnings by February 2007. Brown (1997) finds industry specific differences in forecast errors. Consequently, I omit forecasts from 2006.

Table 1 displays the characteristics of the data over time, before sample selection. There were 81% more forecasts made by 70% more analysts working for 106% more brokerages in 2005 than in 1989.

Table 1
Characteristics of data over time

Year	Forecasts	Number of Firms	Number of Analysts	Number of Brokerages	Forecast / Firm	Forecast / Analyst	Analyst / Brokerage
1989	94,945	4,024	2,781	203	23.59	34.14	13.70
1990	99,486	4,006	2,767	205	24.83	35.95	13.50
1991	99,886	3,996	2,432	201	25.00	41.07	12.10
1992	100,851	4,219	2,325	220	23.90	43.38	10.57
1993	109,125	4,669	2,564	234	23.37	42.56	10.96
1994	110,259	5,198	2,972	239	21.21	37.10	12.44
1995	117,231	5,621	3,268	253	20.86	35.87	12.92
1996	127,509	6,334	3,633	280	20.13	35.10	12.98
1997	130,149	6,702	4,106	330	19.42	31.70	12.44
1998	142,282	6,535	4,467	356	21.77	31.85	12.55
1999	136,839	6,144	4,566	343	22.27	29.97	13.31
2000	130,056	5,701	4,696	332	22.81	27.70	14.14
2001	135,953	4,842	4,712	317	28.08	28.85	14.86
2002	135,553	4,772	4,987	287	28.41	27.18	17.38
2003	141,349	4,766	4,917	378	29.66	28.75	13.01
2004	158,900	5,280	4,689	423	30.09	33.89	11.09
2005	172,184	5,586	4,739	418	30.82	36.33	11.34
Total	2,142,557	17,479	15,006	923	24.48	34.20	12.90

3.2. *Sample selection*

Table 2 details the selection of the data. There are nine reasons for eliminating observations:

(1) Eliminate observations with missing values or duplicate observations.

(2) Eliminate observations more recent than 30 days or older than 365 days. Forecasts more recent than 30 days are typically extremely accurate and subsume all analyst-specific effects. Forecasts older than 365 days are considered stale and uninformative. However, there is little consensus on which cut-offs to use. For example, Hong, Kubik and Solomon (2000) use a range of 180 to 365 days, Sinha, Brown and Das (1997) use 5 to 180 days, and Herrmann and Thomas (2005) use 10 to 300 days. I have adopted the conventions of Clement (1999) and Clement and Tse (2005) and use 30 to 365 days.

(3) Eliminate observations for which the forecasted and actual earnings per share were greater than the 99^{th} percentile or less than the 1^{st} percentile of the distribution of earnings per share. This is consistent with Brown and Mohammad (2001), who suggest that these observations are the most likely to contain errors. However, this process also eliminates very large firms (such as Berkshire Hathaway) and firms with negative earnings (such as technology firms during the 1990s).

(4) Use the most recent observation for analyst *i*'s forecast of firm *j*'s earnings per share during time *t*. This is consistent with the extant literature[5].

(5) Use observations only if firm *j* in time *t* had five or more analysts providing forecasts. In order for aggregation to work, there must be several forecasts to aggregate. However, there is little consensus on which cut-off to use. For example, Bonner, Walther and Young (2003) and Clement and Tse (2005) require only two analysts. I use a cut-off of five analysts, consistent with Mikhail, Walther and Willis (1999), Park and Stice (2000), and Brown and Mohammad (2001).

[5] O'Brien (1990), Dugar and Nathan (1995), Mikhail, Walther and Willis (1997), Sinha, Brown and Das (1997), Jacob, Lys and Neale (1999), Clement (1999), Hong, Kubik and Solomon (2000), Brown and Mohammad (2001), Hong and Kubik (2003), Clement and Tse (2005), and Hermann and Thomas (2005).

Note that this condition reduces forecast bias in the sample. According to Conroy and Harris (1987), Elgers, Lo and Murray (1995), Brown (1997) and Lim (2001), bias is greater for small companies, which typically have less analyst coverage. Smaller firms have more volatile earnings which are difficult to predict – excluding them from the sample improves accuracy.

(6) Eliminate observations if the identity of the analyst could not be established.

(7) Eliminate observations if PMAFE is undefined due to the mean error for firm *j* in time *t* equaling zero, as explained in Section 4.

(8) Eliminate observations for which analyst *i* did not make a forecast for firm *j* in time *t-1*. This restriction is necessary to lag variables and is consistent with Brown and Mohammad (2001).

(9) Only forecasts between 1983 and 2005 were used, as explained in Section 3.1.

Table 2
Selection of data

Initial observations obtained from I/B/E/S	2,976,842
Missing values	-268,961
Duplicates	-837
Forecast more recent than 30 days	-334,180
Forecast older than 365 days	-32,329
Outside 1st and 99th percentiles of Forecast or Actual	-60,110
Not most recent forecast by analyst i for firm j in time t	-1,488,941
Covered by fewer than five analysts for firm j in time t	-108,842
Analyst code does not uniquely identify analyst	-13,802
PMAFE is undefined due to the mean error for firm j in time t equaling 0	-149
No observation for analyst i for firm j in time t-1	-282,116
Earnings before 1989 (1982 - 1988) or after 2005 (2006)	-78,544
Final observations used for analysis	308,031

4. VARIABLES

From the eight variables provided by I/B/E/S (see Section 3.1), we can create seven independent variables and one dependent variable. All names are consistent with Brown and Mohammad (2001). The theoretical justification for these variables can be found in Section 2.

AGE measures the number of calendar days between analyst *i*'s last forecast for firm *j* in time *t* and firm *j*'s fiscal year end. For example, if an analyst released his last forecast on November 15th for a firm with a fiscal year end of December 31st, AGE would equal 46 days. FREQ, short for frequency, measures the number of forecasts made by analyst *i* for firm *j* in time *t*. NCOS, short for number of companies, measures the number of firms covered by analyst *i* in time *t*. NTOP10 is a dummy variable equal to 1 if analyst *i* works for a top decile brokerage[6], where brokers are ranked by the number of analysts making forecasts in time *t*[7]. FEXP, short for firm experience, measures the number of years analyst *i* has made a forecast for firm *j*. NIND measures the number of industries – as defined by two-digit SIC codes – covered by analyst *i* in time *t*. All variables are mean-adjusted per Clement (1999) to control for firm-year effects, such as "voluntary management disclosures, mergers and strikes". The suffix D, as in DAGE and DFREQ, indicates mean-adjustment. For example, DAGE equals AGE for analyst *i* covering firm *j* in time *t* minus the average AGE for all analysts (including *i*) covering firm *j* in time *t*. Consistent with Brown and Mohammad (2001), DNCOS, DNTOP10, DFEXP and DNIND are lagged one period. Thus, if an analyst is covering General Motors in 2005, we measure how many companies the analyst covered in 2004.

While the literature is generally consistent in its treatment of the independent variables, there is little consensus on creating error terms. I have adopted the technique used by Clement (1999), Jacob, Lys and Neale (1999) and Brown and Mohammad (2001), who deflate absolute forecast

---

6 Clement (1999) finds that his results are robust to using a continuous measure of brokerage size as well.

7 Mikhail, Walther and Willis (1999) use three different proxies for the reputation of brokerage houses: (1) the size of the brokerage house, measured by the number of analysts making forecasts; (2) the profitability of buy recommendations issued by the brokerage house; and (3) the dollar magnitude of investment banking business. They find all three measures provide qualitatively similar results. Hong and Kubik (2003) also use three different proxies for the reputation of brokerage houses: (1) the size of brokerage houses; (2) the number of All-American forecasters; and (3) Carter and Manaster's (1990) ranking based on the prestige of firm name location in newspaper "tombstone" advertisements. They find all three measures provide qualitatively similar results.

error by average firm-year errors to allow comparison across companies and years independent of differences in forecast difficulty[8]. FE, short for forecast error, measures the absolute difference between forecasted and actual earnings per share. PMAFE, short for proportional mean absolute forecast error, equals the forecast error for analyst *i* covering firm *j* in time *t*, divided by the average forecast error of all analysts covering firm *j* in time *t*, minus 1. When the average forecast error is equal to zero, PMAFE is undefined; Section 3.2 discusses eliminating these observations. When an analyst has no forecast error, PMAFE is equal to -1. When the average forecast error is small, PMAFE is extremely large. To control against outliers, all PMAFE's greater than 1 are set equal to 1, consistent with Brown (1991) and Sinha, Brown and Das (1997). Thus, PMAFE ranges from -1 to 1, with negative values representing better than average performance and positive values representing worse than average performance. PMAFE has an attractive characteristic: a PMAFE of -0.50, for example, indicates a forecast 50% better than average; a PMAFE of 0.25 indicates a forecast 25% worse than average. LMAFE is PMAFE lagged one period.

[8] Alternative treatments include: using simply forecast error as in Stickel (1992); deflating forecast error by forecasted earnings per share as in Brown (1991), and Brown (1997); deflating forecast error by actual earnings per share as in Conroy and Harris (1987), Brown and Kim (1991) and Sinha, Brown and Das (1997); deflating forecast error by stock price as in Dugar and Nathan (1995), Park and Stice (2000), Lim (2001), Hong and Kubik (2003), Clement and Tse (2005) and Ramnath, Rock and Shane (2005); using mean square error as in Abarbanell (1991), and Lo and Elgers (1998); and deflating mean square error by actual earnings per share as in Lobo (1991) and Ali, Klein and Rosenfeld (1992). However, Abarbanell (1991) and Stickel (1992) conclude that the choice of deflator does not qualitatively affect results. In the literature, only the choice of forecast standard error as a deflator affects conclusions, as in Alexander (1995).

Table 3 provides descriptive statistics for unadjusted and adjusted variables, showing the median and the 10th, 25th, 75th and 90th percentiles of the pooled distribution.

Table 3
Descriptive statistics for unadjusted and adjusted variables

	Percentiles				
	10	25	50	75	90
Panel A: Unadjusted Variables					
AGE	47	60	76	162	258
FREQ	1	2	3	5	6
NCOS	8	11	16	22	33
NTOP10	0	0	0	0	1
FEXP	2	2	3	6	9
NIND	1	3	7	13	21
FE	0.01	0.02	0.05	0.15	0.41
Panel B: Adjusted Variables					
DAGE	-71.09	-47.74	-20.50	38.43	126.83
DFREQ	-1.70	-0.86	0.13	1.20	2.35
DNCOS	-10.60	-6.00	-1.27	4.00	11.82
DNTOP10	-0.20	-0.15	-0.10	0.00	0.76
DFEXP	-2.77	-1.43	-0.14	1.08	3.33
DNIND	-5.61	-2.25	-0.10	2.36	6.45
PMAFE	-0.87	-0.58	-0.18	0.24	1.00

AGE = Report Date - Estimate Date
FREQ = Forecasts for firm j in time t by analyst i
NCOS = Companies followed by analyst i
NTOP10 = 1 if analyst i works for top decile broker
FEXP = Years of forecasts for firm j by analyst i
NIND = Industries followed by analyst i
FE = | Actual EPS - Forecasted EPS |

DAGE = $AGE_{i,j,t} - AGE_{j,t}$
DFREQ = $FREQ_{i,j,t} - FREQ_{j,t}$
DNCOS = $NCOS_{i,j,t-1} - NCOS_{j,t-1}$
DNTOP10 = $NTOP10_{i,j,t-1} - NTOP10_{j,t-1}$
DFEXP = $FEXP_{i,j,t-1} - FEXP_{j,t-1}$
DNIND = $NIND_{i,j,t-1} - NIND_{j,t-1}$
PMAFE = $(FE_{i,j,t} / FE_{j,t}) - 1$

Table 4 shows correlation coefficients for unadjusted and adjusted variables. All correlations are statistically significant at 1%. As expected, DNCOS and DNIND are positively correlated (0.636). When an analyst covers many industries (high DNIND), they also cover many firms (DNCOS). Also as expected, DAGE and DFREQ are negatively correlated (-0.552). When an analyst makes many forecasts for firm *j* in time *t* (high DFREQ), their forecasts are more recent (low DAGE). While this high multi-collinearity may have statistical implications for the coefficients of regression (see Section 5.1), it is advantageous to include all four variables for predicting PMAFE.

Table 4
Correlation coefficients for unadjusted and adjusted variables

Panel A: Unadjusted Variables

	AGE	FREQ	NCOS	NTOP10	FEXP	NIND	FE
AGE	1.000						
FREQ	-0.545	1.000					
NCOS	-0.034	-0.051	1.000				
NTOP10	-0.019	0.011	-0.012	1.000			
FEXP	-0.039	0.036	0.097	0.038	1.000		
NIND	-0.035	-0.003	0.630	0.020	0.077	1.000	
FE	0.144	-0.060	-0.004	-0.001	-0.020	0.002	1.000

Panel B: Adjusted Variables

	DAGE	DFREQ	DNCOS	DNTOP10	DFEXP	DNIND	PMAFE	LMAFE
DAGE	1.000							
DFREQ	-0.552	1.000						
DNCOS	0.012	-0.031	1.000					
DNTOP10	-0.016	0.006	-0.023	1.000				
DFEXP	-0.032	0.010	0.116	0.018	1.000			
DNIND	-0.017	0.024	0.636	0.024	0.116	1.000		
PMAFE	0.418	-0.297	0.031	-0.021	-0.026	-0.002	1.000	
LMAFE	0.034	-0.057	0.033	-0.017	-0.008	0.001	0.064	1.000

AGE = Report Date - Estimate Date
FREQ = Forecasts for firm j in time t by analyst i
NCOS = Companies followed by analyst i
NTOP10 = 1 if analyst i works for top decile broker
FEXP = Years of forecasts for firm j by analyst i
NIND = Industries followed by analyst i
FE = | Actual EPS - Forecasted EPS |

DAGE = $AGE_{i,j,t} - AGE_{j,t}$
DFREQ = $FREQ_{i,j,t} - FREQ_{j,t}$
DNCOS = $NCOS_{i,j,t-1} - NCOS_{j,t-1}$
DNTOP10 = $NTOP10_{i,j,t-1} - NTOP10_{j,t-1}$
DFEXP = $FEXP_{i,j,t-1} - FEXP_{j,t-1}$
DNIND = $NIND_{i,j,t-1} - NIND_{j,t-1}$
PMAFE = $(FE_{i,j,t} / FE_{j,t}) - 1$

5. RESULTS

5.1. REGRESSIONS

Table 5 shows regression coefficients (multiplied by 100) and t-statistics. Model A, displayed in panel A, regresses PMAFE on DAGE, DFREQ, DNCOS, DNTOP10, DFEXP and DNIND per Brown and Mohammad (2001). Model B, displayed in panel B, regresses PMAFE on LMAFE in addition to the aforementioned six variables. Ordinary Least-Squares regressions are run for each year in the sample, as well as cross-sectionally across the sample.

$$\text{(A)} \quad PMAFE = \beta_1 DAGE + \beta_2 DFREQ + \beta_3 DNCOS + \beta_4 DNTOP10 + \beta_5 DFEXP + \beta_6 DNIND + \mu$$

DAGE is expected to have a positive sign: more recent forecasts (low DAGE) are more accurate (low PMAFE). DAGE has a statistically significant positive sign in all 16 years. Its average coefficient of 0.0025 is also economically significant: moving from the 90th percentile (126.8 days older than average) to the 10th percentile (71.1 more recent than average) increases forecast accuracy by 50.1%.

DFREQ is expected to have a negative sign: issuing more forecasts (high DREQ) leads to more accurate forecasts (low PMAFE). DFREQ has a statistically significant negative sign in all 16 years. Its average coefficient in regression B of -0.0477 is also economically significant: moving from the 10th percentile (1.7 less forecasts than average) to the 90th percentile (2.3 more forecasts than average) increases forecast accuracy by 19.3%.

DNCOS is expected to have a positive sign: covering fewer companies (low DNCOS) leads to more accurate forecasts (low PMAFE). DNCOS has a statistically significant positive sign in 11 of the 16 years. It has a negative sign for two years, though neither is significant. Its average coefficient of 0.0011 is not economically significant: moving from the 90th percentile (covering 11.8 more companies than average) to the 10th percentile (covering 10.6 less companies than average) increases forecast accuracy by 2.4%. The relationship between DNCOS and PMAFE weakens after 2000.

DNTOP10 is expected to have a negative sign: working for a large brokerage (high DNTOP10) leads to more accurate forecasts (low PMAFE). DNTOP10 has a statistically significant negative sign in 9 of the 16 years. It has a positive sign for four years, though none are significant. Its average coefficient of -0.0317 is not economically significant: moving from the 10th percentile to the 90th percentile increases forecast accuracy by 3.0%.

DFEXP is expected to have a negative sign: having many years of experience (high DEXP) leads to more accurate forecasts (low PMAFE). DEXP has a statistically significant negative sign in 12 of the 16 years. It has a positive sign for one year, though it is not significant. Its average coefficient of -0.0058 is not economically significant: moving from the 10th (having 2.7 fewer years of experience than average) to the 90th percentile (having 3.3 more years of experience than average) increases forecast accuracy by 3.5%. The relationship between DFEXP and PMAFE weakens after 2001.

DNIND is expected to have a positive sign: covering fewer industries (low DNIND) leads to more accurate forecasts (low PMAFE). Unexpectedly, DNIND has a statistically significant *negative* sign in 9 of the 16 years. It has a positive sign for only one year, and it is not significant. Its average coefficient of -0.0019 is not economically significant: moving from the 10th percentile (covering 5.6 industries less than average) to the 90th percentile (covering 6.4 more industries than average) increases forecast accuracy by 2.3%.

Brown and Mohammad (2001) find similar results. Indeed, their DAGE coefficient is 0.002 and their DFEXP coefficient is -0.005. The biggest difference occurs with DNIND, where their coefficient is 0.003 (and significant at 5%). Most likely, this arises because of the broker translation file mentioned in Section 3.1. Brown and Mohammad eliminate teams of analysts from their sample, which I am unable to do. Teams of analysts can cover more industries than any individual analyst. Brown and Mohammad's inter-quartile range for DNIND was -1.0 to 0.9, and Clement (1999) has a range of -1.5 to 1.0, compared to -2.2 to 2.3 for my sample. The unadjusted numbers in Table 3 clearly reflect the influence of teams: the 90th percentile of NIND is 21, which means 10% of analysts covered more than 21 industries. This is quite unrealistic for an individual analyst. I attribute all inconsistencies between Brown and Mohammad, Clement and my results to this sample selection attribute.

$$\text{(B)} \quad \text{PMAFE} = \beta_1\text{DAGE} + \beta_2\text{DFREQ} + \beta_3\text{DNCOS} + \beta_4\text{DNTOP10} + \beta_5\text{DEXP} + \beta_6\text{DNIND} + \beta_7\text{LMAFE} + \mu$$

LMAFE is expected to have a positive sign: being more accurate in the past (low LMAFE) leads to more accurate forecasts (low PMAFE). LMAFE has a statistically significant positive sign in all 16 years. Its average coefficient of 0.1099 is economically significant: moving from the 90th percentile (being 100% less accurate in the past than average) to the 10th percentile (being 87% more accurate in the past than average) increases forecast accuracy by 20.5%.

Adding LMAFE to the regression does not appreciably increase the adjusted R^2 of the regression, as displayed in Panel C. (Note that the R^2 refers to a mean-adjusted regression with firm-year effects removed. In other words, it is equivalent to running a regression against the residuals of the regression of error on firm-year effects. The relatively low R^2s are not indicative of a weak relationship between error and analyst-specific characteristics). Furthermore, adding LMAFE to the regression has little effect on the sign, magnitude or significance of the six variables. LMAFE changes the sign of NCOS in one year, though not significantly. Adding LMAFE reduces the significance of DNTOP10 for three years, DFEXP for one year and DNIND for one year. Overall, the effects of adding LMAFE to the regression are minor.

However, LMAFE itself is highly significant. This is consistent with Brown (2001), who finds that a two-factor model with DAGE and LMAFE is as accurate as a six-factor model with DAGE, DGEXP[9], DFEXP, DNCOS, DNIND and DNTOP10. Brown suggests combining the models, though he concludes, "Whether its slight predictive advantage justifies its additional complexity depends on the decision context."

[9] DGEXP, general experience, was not used in order to be consistent with Brown and Mohammad (2001). Mikhail, Walther and Willis (1997) and Clement (1999) use both general experience and firm experience.

Table 5
Regression model estimates (with t-statistics in parenthesis)

Panel A: $PMAFE = \beta_1 DAGE + \beta_2 DFREQ + \beta_3 DNCOS + \beta_4 DNTOP10 + \beta_5 DFEXP + \beta_6 DNIND + \mu$

Year	DAGE	DFREQ	DNCOS	DNTOP10	DFEXP	DNIND
1989	0.18	-4.24	0.07	-11.60	-1.22	0.00
	(28.73)	(-13.14)	(1.88)	(-7.92)	(-4.37)	(-0.06)
1990	0.22	-3.95	0.07	-6.65	-1.51	-0.11
	(32.74)	(-11.97)	(2.97)	(-4.56)	(-5.79)	(-2.05)
1991	0.24	-3.52	0.11	-4.30	-0.69	-0.17
	(35.65)	(-12.12)	(5.16)	(-3.33)	(-3.32)	(-3.43)
1992	0.20	-2.85	0.19	-4.40	-0.76	-0.20
	(29.79)	(-9.93)	(7.01)	(-3.45)	(-3.85)	(-3.52)
1993	0.21	-4.71	0.16	-1.68	-0.63	-0.14
	(32.81)	(-15.27)	(4.30)	(-1.33)	(-3.28)	(-2.52)
1994	0.20	-4.74	0.12	-1.89	-0.75	-0.11
	(31.47)	(-14.62)	(3.50)	(-1.48)	(-4.17)	(-1.88)
1995	0.22	-5.38	0.17	-2.75	-0.68	-0.03
	(32.34)	(-16.61)	(4.56)	(-2.15)	(-4.10)	(-0.50)
1996	0.24	-5.95	0.18	1.76	-0.70	-0.11
	(37.19)	(-19.99)	(4.60)	(1.30)	(-4.60)	(-1.69)
1997	0.21	-5.44	0.16	0.21	-0.43	-0.13
	(33.82)	(-18.20)	(4.19)	(0.16)	(-3.05)	(-1.94)
1998	0.27	-6.01	0.17	1.02	-0.53	-0.28
	(43.26)	(-22.41)	(4.45)	(0.87)	(-3.91)	(-3.91)
1999	0.26	-4.56	0.11	-5.38	-0.60	0.02
	(42.40)	(-16.53)	(2.80)	(-4.26)	(-4.51)	(0.25)
2000	0.29	-5.68	0.07	-8.31	-0.26	-0.16
	(48.27)	(-20.27)	(1.44)	(-6.75)	(-2.00)	(-1.84)
2001	0.36	-6.95	0.19	-3.46	0.03	-0.41
	(57.10)	(-25.96)	(3.24)	(-2.78)	(0.18)	(-3.72)
2002	0.34	-4.75	0.10	-3.83	-0.08	-0.59
	(54.81)	(-17.91)	(1.17)	(-3.13)	(-0.54)	(-4.45)
2003	0.32	-3.80	-0.04	-0.75	-0.23	-0.32
	(52.29)	(-15.40)	(-0.41)	(-0.60)	(-1.55)	(-2.49)
2004	0.29	-3.82	-0.12	1.32	-0.19	-0.25
	(47.78)	(-15.96)	(-1.55)	(1.04)	(-1.39)	(-2.08)
Average	0.25	-4.77	0.11	-3.17	-0.58	-0.19
	(19.00)	(-17.71)	(4.99)	(-3.46)	(-5.80)	(-4.83)
Pooled	0.26	-4.85	0.12	-3.06	-0.45	-0.15
	(167.60)	(-70.96)	(14.00)	(-9.83)	(-11.72)	(-8.94)
Sign	+	-	+	-	-	+

Table 5, continued

Panel B: $PMAFE = \beta_1 DAGE + \beta_2 DFREQ + \beta_3 DNCOS + \beta_4 DNTOP10 + \beta_5 DFEXP + \beta_6 DNIND + \beta_7 LMAFE + \mu$

Year	DAGE	DFREQ	DNCOS	DNTOP10	DFEXP	DNIND	LMAFE
1989	0.18	-3.99	0.06	-11.08	-1.12	-0.01	8.09
	(29.00)	(-12.38)	(1.63)	(-7.59)	(-4.01)	(-0.09)	(10.51)
1990	0.22	-3.77	0.07	-6.10	-1.41	-0.10	8.05
	(33.00)	(-11.44)	(2.80)	(-4.19)	(-5.42)	(-1.86)	(9.42)
1991	0.24	-3.28	0.10	-3.62	-0.63	-0.15	10.76
	(35.59)	(-11.32)	(4.72)	(-2.82)	(-3.07)	(-3.00)	(14.32)
1992	0.20	-2.57	0.16	-4.17	-0.66	-0.15	10.14
	(29.92)	(-8.99)	(6.06)	(-3.29)	(-3.37)	(-2.71)	(13.77)
1993	0.21	-4.48	0.13	-1.60	-0.56	-0.12	9.21
	(32.76)	(-14.55)	(3.71)	(-1.26)	(-2.90)	(-2.18)	(12.28)
1994	0.20	-4.35	0.11	-1.55	-0.67	-0.08	10.92
	(31.58)	(-13.46)	(3.01)	(-1.22)	(-3.76)	(-1.46)	(14.69)
1995	0.22	-4.92	0.14	-2.40	-0.59	0.00	11.51
	(32.86)	(-15.22)	(3.72)	(-1.88)	(-3.55)	(-0.02)	(15.50)
1996	0.24	-5.45	0.15	1.79	-0.61	-0.08	12.13
	(38.21)	(-18.37)	(3.85)	(1.33)	(-4.06)	(-1.22)	(17.40)
1997	0.21	-4.86	0.12	-0.01	-0.38	-0.08	11.59
	(34.76)	(-16.26)	(3.31)	(-0.01)	(-2.68)	(-1.22)	(16.99)
1998	0.27	-5.59	0.15	1.00	-0.48	-0.25	10.34
	(44.04)	(-20.86)	(3.95)	(0.86)	(-3.56)	(-3.59)	(15.08)
1999	0.27	-4.08	0.08	-4.79	-0.59	0.06	9.41
	(43.42)	(-14.76)	(2.19)	(-3.81)	(-4.42)	(0.88)	(13.78)
2000	0.29	-5.28	0.04	-7.62	-0.23	-0.12	9.33
	(49.22)	(-18.85)	(0.90)	(-6.21)	(-1.79)	(-1.36)	(13.39)
2001	0.37	-6.19	0.15	-2.34	0.03	-0.31	13.31
	(59.16)	(-23.05)	(2.59)	(-1.89)	(0.20)	(-2.87)	(18.71)
2002	0.35	-3.77	0.05	-3.45	-0.14	-0.40	12.78
	(57.62)	(-14.13)	(0.57)	(-2.85)	(-0.95)	(-3.09)	(19.70)
2003	0.33	-2.93	0.04	-0.74	-0.27	-0.29	14.41
	(54.99)	(-11.85)	(0.45)	(-0.60)	(-1.84)	(-2.30)	(21.59)
2004	0.30	-3.06	-0.11	1.03	-0.18	-0.17	13.81
	(49.87)	(-12.78)	(-1.37)	(0.83)	(-1.35)	(-1.40)	(21.53)
Average	0.26	-4.29	0.09	-2.85	-0.53	-0.14	10.99
	(17.71)	(-16.40)	(5.38)	(-3.34)	(-5.91)	(-4.61)	(22.58)
Pooled	0.26	-4.36	0.11	-2.68	-0.41	-0.11	11.23
	(171.66)	(-63.77)	(11.90)	(-8.67)	(-10.77)	(-6.95)	(65.41)
Sign	+	-	+	-	-	+	+

Table 5, continued

Panel C: Adjusted R^2

Year	Obs.	Model A	Model B
1989	15,355	0.12	0.12
1990	13,331	0.15	0.16
1991	16,678	0.13	0.14
1992	17,393	0.09	0.10
1993	17,550	0.12	0.13
1994	17,510	0.12	0.13
1995	17,501	0.13	0.15
1996	18,376	0.17	0.18
1997	18,823	0.14	0.15
1998	19,290	0.20	0.21
1999	19,691	0.17	0.18
2000	18,849	0.23	0.24
2001	17,861	0.34	0.35
2002	17,615	0.29	0.30
2003	18,634	0.23	0.25
2004	20,712	0.19	0.21
Pooled	308,031	0.18	0.19

Note: Panels A and B display ordinary least squared regression estimates for two models. Model A, displayed in Panel A, regresses PMAFE against DAGE, DFREQ, DNCOS, DNTOP10, DEXP and DNIND. Model B, displayed in Panel B, adds a lagged term, LMAFE. All coefficients are multipled by 100. Panel C compares adjusted R^2s for the two models.

5.2. *Aggregation*

Aggregating individual forecasts allows us to create consensus forecasts. Currently, I/B/E/S publishes mean and median consensus forecasts for firm *j* in time *t*. A mean, or average, consensus is equivalent to assigning a weight of 1 / N to each forecast, where N is the number of analysts making forecasts for firm *j* in time *t*. Thus, if there are five analysts covering a firm in a particular year, each forecast is given a weight of 20%. See Example 1 for an illustration of this method. The forecast error (in this example, 0.12) is the absolute difference between the consensus forecast (0.65) and the actual earnings per share (0.77).

Brown and Mohammad (2001) suggest a new weighting method, which I call Method 1. In the "Best" case, named "Perfect" by Brown and Mohammad, they rank *ex post* realized PMAFEs for firm *j* in time *t*, with the most accurate forecast (lowest PMAFE) receiving a rank of N and the least accurate forecast (highest PMAFE) receiving a rank of 1. They assign a weight to each forecast equal to RANK (PMAFE) / [N x (N+1) / 2]. In Example 1, the most accurate forecast is weighted 33% (= 5 / 15) while the least accurate forecast is weighted 7% (= 1 / 15). This leads to a consensus forecast of 0.69 and a forecast error of 0.08, which is an improvement on average weighting. In the “Model” case, they create *ex ante* estimated PMAFEs (alternatively, PMAFE hat) using regressions from Section 5.1. Estimated PMAFEs for time *t* are created using regression results from time *t-1*. For example, the estimated PMAFE for an analyst covering General Motors in 2005 uses analyst characteristics from 2005 (except for lagged variables DNCOS, DNTOP10, DFEXP, DNIND and LMAFE) multiplied by regression coefficients from 2004. “Best”, as the name suggests, represents the best possible consensus forecast for the weighting method. “Model” approaches “Best” as the regression explains more variation in forecast errors. In Example 1, “Model” produces a forecast error of 0.10, which is 0.02 (or 20%) worse than “Best”.

I introduce two alternative weighting methods. Method 2 assigns a weight of $PMAFE^2$ / $\Sigma(PMAFE^2)$ for all PMAFEs less than zero. The weights are designed to sum to 100%. This method should be superior to Method 1 as it considers only forecasts better (or expected to be better) than average. Furthermore, it accounts for the magnitude of forecast accuracy. In Example 1, Method 1 only gives a weight of 33% to the most accurate forecast, while Method 2 gives a weight of 83%. Likewise, Method 1 gives a weight of 7% to the least accurate forecast, while Method 2 gives no weight. In Example 1, Method 2 produces a forecast error of 0.02 in the “Best” case and 0.10 in the “Model” case.

Method 3 assigns a weight of [MAX(PMAFE) - PMAFE] / Σ[MAX(PMAFE - MAX(PMAFE)]. This method gives a weight of zero to the least accurate (highest PMAFE) forecast and linearly gives larger weights to more accurate (lower PMAFE) forecasts. The weights are designed to sum to 100%. Like Method 2, Method 3 is superior to Method 1 because it accounts for the magnitude of forecast accuracy. Unlike Method 2, Method 3 puts less weight on the most accurate forecasts by also considering forecasts less accurate than average. It is expected that Method 3 has errors greater than Method 2. In Example 1, Method 3 produces a forecast error of 0.06 in the “Best” case and 0.09 in the “Model” case.

Example 1
Aggregating forecast errors for one firm in one year

Analyst	Forecast	FE	PMAFE	$\widehat{\text{PMAFE}}$
1	0.54	0.23	0.98	0.83
2	0.60	0.17	0.47	2.10
3	0.67	0.10	-0.14	-0.90
4	0.70	0.07	-0.40	-0.35
5	0.76	0.01	-0.91	0.05
Actual	0.77			

Average : $w_i = 1 / N$

Analyst	Forecast	Weight
1	0.54	20.0%
2	0.60	20.0%
3	0.67	20.0%
4	0.70	20.0%
5	0.76	20.0%
	Consensus	0.65
	Error	$\boxed{0.12}$

Method 1 : w_i = RANK(PMAFE) / [N x (N+1) / 2]

		Best		Model	
Analyst	Forecast	Numerator	Weight	Numerator	Weight
1	0.54	1	7%	2	13%
2	0.60	2	13%	1	7%
3	0.67	3	20%	5	33%
4	0.70	4	27%	4	27%
5	0.76	5	33%	3	20%
	Consensus		0.69		0.67
	Error		$\boxed{0.08}$		$\boxed{0.10}$

Method 2 : $w_i = \text{PMAFE}^2 / \Sigma(\text{PMAFE}^2)$ for all PMAFE < 0

		Best		Model	
Analyst	Forecast	Numerator	Weight	Numerator	Weight
1	0.54	0.00	0%	0.00	0%
2	0.60	0.00	0%	0.00	0%
3	0.67	0.02	2%	0.81	87%
4	0.70	0.16	16%	0.12	13%
5	0.76	0.84	83%	0.00	0%
	Consensus		0.75		0.67
	Error		$\boxed{0.02}$		$\boxed{0.10}$

Method 3 : w_i = [MAX(PMAFE) - PMAFE] / Σ[MAX(PMAFE) - PMAFE]

		Best		Model	
Analyst	Forecast	Numerator	Weight	Numerator	Weight
1	0.54	0.00	0%	1.27	14%
2	0.60	0.52	11%	0.00	0%
3	0.67	1.12	23%	3.00	34%
4	0.70	1.38	28%	2.45	28%
5	0.76	1.90	39%	2.05	23%
	Consensus		0.71		0.68
	Error		$\boxed{0.06}$		$\boxed{0.09}$

Example 1 explained how to calculate consensus forecasts and consensus forecast errors for firm *j* in time *t*. Table 6 shows median forecast errors across all firms in time *t* expressed in cents[10].

Table 6
Aggregate forecast errors using median absolute error expressed in cents

Year	Average	Method 1			Method 2			Method 3		
		Best	Model A	Model B	Best	Model A	Model B	Best	Model A	Model B
1990	8.40	5.04	6.93	6.76	1.90	5.33	5.29	4.00	6.29	6.31
1991	5.48	3.29	4.80	4.75	1.43	4.65	4.65	2.83	4.54	4.57
1992	4.80	2.73	4.26	4.24	1.27	4.02	4.01	2.50	4.17	4.21
1993	4.30	2.62	3.83	3.83	1.18	3.52	3.52	2.18	3.71	3.70
1994	4.50	2.57	3.71	3.76	1.09	3.33	3.25	2.16	3.50	3.52
1995	4.27	2.56	3.71	3.76	1.10	3.49	3.23	2.00	3.50	3.50
1996	4.71	2.70	3.96	3.92	1.21	3.41	3.34	2.25	3.83	3.75
1997	4.39	2.67	3.72	3.69	1.29	3.33	3.26	2.33	3.53	3.52
1998	5.33	3.23	4.46	4.48	1.57	3.88	3.85	2.78	4.21	4.26
1999	5.34	3.20	4.15	4.10	1.62	3.68	3.77	2.67	4.00	4.00
2000	6.40	3.76	4.92	4.90	1.82	4.10	4.19	3.04	4.63	4.58
2001	6.10	2.91	4.13	4.07	1.38	3.05	3.05	2.23	3.65	3.59
2002	4.62	2.54	3.43	3.48	1.24	2.93	2.93	2.14	3.09	3.14
2003	4.00	2.36	3.14	3.12	1.16	2.84	2.83	2.06	3.00	3.00
2004	4.33	2.56	3.73	3.70	1.26	3.44	3.46	2.28	3.59	3.55
2005	5.00	3.00	4.32	4.26	1.55	4.04	3.98	2.65	4.24	4.23
Average	5.12	2.98	4.20	4.18	1.38	3.69	3.66	2.51	3.97	3.96
Pooled	4.97	2.87	4.04	4.02	1.35	3.59	3.54	2.43	3.84	3.84

Note: This table displays median absolute forecast errors expressed in cents for different aggregation methods. For example, the left column aggregates forecasts using simple averages. Each forecast by analyst i covering firm j in time t is given a ($^1/_N$) weight, where N is the number of analysts covering firm j in time t. This leads to a consensus forecast of $C_{i,j,t} = \Sigma w_{i,j,t} F_{i,j,t}$, which is then compared to actual earnings, $A_{i,j,t}$. The error is expressed as the absolute difference between $C_{i,j,t}$ and $A_{i,j,t}$. The table shows the median absolute forecast error across firms j for time t. For example, in 1990 there was a median absolute forecast error of 8.40 cents, or \$0.084, for the simple average method.

Average $w_i = \dfrac{1}{N}$

Method 1 $w_i = \dfrac{\text{RANK}(\widehat{\text{PMAFE}})}{N \times (N+1) / 2}$

Method 2 $w_i = \dfrac{\widehat{\text{PMAFE}}^2}{\Sigma\, \widehat{\text{PMAFE}}^2}$ for all $\widehat{\text{PMAFE}} < 0$

Method 3 $w_i = \dfrac{\text{MAX}(\widehat{\text{PMAFE}}) - \widehat{\text{PMAFE}}}{\Sigma\, [\text{MAX}(\widehat{\text{PMAFE}}) - \widehat{\text{PMAFE}}]}$

Best $\widehat{\text{PMAFE}} = \text{PMAFE}$

Model A $\widehat{\text{PMAFE}} = \beta_0 + \beta_1\text{DAGE} + \beta_2\text{DFREQ} + \beta_3\text{DNCOS} + \beta_4\text{DNTOP10} + \beta_5\text{DEXP} + \beta_6\text{DNIND} + \mu$

Model B $\widehat{\text{PMAFE}} = \beta_0 + \beta_1\text{DAGE} + \beta_2\text{DFREQ} + \beta_3\text{DNCOS} + \beta_4\text{DNTOP10} + \beta_5\text{DEXP} + \beta_6\text{DNIND} + \beta_7\text{LMAFE} + \mu$

[10] The analysis is robust to defining errors as mean, as opposed to median, forecast errors. However, the distribution of errors is positively skewed, so mean errors are significantly larger than median errors.

There are three ways to interpret Table 6: (1) compare errors across years; (2) compare errors across models; and (3) compare errors across weighting methods.

5.2.1. COMPARE ERRORS ACROSS YEARS

Errors were greatest in 1990, 2000 and 2001. These were turbulent years for the equity markets, with firms frequently reporting earnings dramatically below forecasts. Otherwise, forecast errors have been remarkably consistent over the past 16 years.

5.2.2. COMPARE ERRORS ACROSS MODELS

Looking at Method 1, the "Best" case has an average forecast error of 2.98 cents. Using Model A to estimate PMAFE generates an average forecast error of 4.20 cents, or 41% more error than the "Best" case. Adding LMAFE to the regression in Model B leads to an average forecast error of 4.18 cents, or 40% more error than the "Best" case. Consequently, Model B has 1% less error than Model A[11], which is not economically significant. Thus, adding LMAFE to the regression, as discussed in Section 5.1, does not significantly improve our ability to aggregate forecasts.

5.2.3. COMPARE ERRORS ACROSS WEIGHTING METHODS

The average weighting method, which is commonly used throughout the financial industry, has a median forecast error of 5.12 cents ($0.00512) over the past 16 years[12]. Looking across Model B cases, Method 1 has a median forecast error of 4.18 cents, which is an improvement of 0.95 cents, or 18%. Method 2 has a median forecast error of 3.66 cents, an improvement of 1.46 cents, or 29%. Method 3 has a median forecast error of 3.96 cents, an improvement of 1.16 cents, or 23%. All three methods are superior to average weighting. Furthermore, as expected, Method 2 reduces error by 12% relative to Brown and Mohammad's Method 1. Method 3 reduces error by 5% relative to Method 1. These results are very economically significant.

---

[11] The results are robust to using Method 2 or Method 3 to compare Models A and B. In either case, Model B reduces forecast error by 1% relative to Model A.

[12] The results are robust to defining forecast error as $(\text{Actual} - \text{Forecast})^2$, as opposed to Abs(Actual – Forecast).

6. CONCLUSIONS

Earnings per share forecasts made by research analysts can be combined in many ways. One of the most common methods in practice is to equally weight each analyst's forecast to create a consensus forecast. Equal weighting produces a median forecast error of 5.12 cents. Brown and Mohammad (2001), in the only study to date that examines forecast aggregation, are able to improve forecast accuracy by weighting forecasts with analyst specific characteristics. There have been a lot of numbers in this paper, but before I present more I would like you to step back and think about that for a second: Brown and Mohammad know *nothing* about the firms they are forecasting – they don't know if General Motors will sell 5 or 5 million cars – yet by looking at analyst characteristics (How any years of experience does the analyst have? How many companies does the analyst cover?), they are more accurate than highly skilled experts. Back to the numbers: Brown and Mohammad's weighting method produces a median forecast error of 4.20 cents, 18% more accurate than equal weighting.

I employ Brown and Mohammad's general methodology to further improve forecast accuracy. The first step is to regress analyst accuracy against analyst characteristics, such as forecast recency, forecast frequency, brokerage reputation, experience, and the number of companies and industries covered. In addition to their six analyst characteristics, I consider past accuracy[13]. Accuracy is predicted using the previous year's regression coefficients. Brown and Mohammad weight analysts using an ordinal system where the magnitude of predicted error is inconsequential. I improve on their methodology by weighing analyst forecasts using squared predicted accuracy. My weighting method produces a median forecast error of 3.66 cents, 29% more accurate than equal weighting and 13% more accurate than Brown and Mohammad.

Brown and Mohammad found that a trading strategy of purchasing securities whose model-based forecasts exceeded equal weighted forecasts by 5% resulted in abnormal returns of 1.87%. It remains to be seen whether a 13% improvement in forecast accuracy translates into a 13% improvement in abnormal returns.

[13] A practitioner may be better off using forecast recency and past accuracy, as opposed to the full seven factor model. These two characteristics are the easiest to obtain and explain the most variation in forecast error.

6.1. *Future research*

Future research is needed on two parallel tracks. First, alternative weighting schemes should be considered. Methods 2 and 3 were improvised and, though they have appealing economic qualities, have little statistical justification. Second, additional analyst characteristics should be considered in improving predictions of analyst accuracy. There are four characteristics in the literature that deserve further study:

(1) Francis and Philbrick (1993) find that analysts issuing "Sell" recommendations are more likely to have optimistic – and consequently, inaccurate – forecasts. They contend that analysts attempt to compensate companies for negative recommendations in order to maintain manager access.

(2) Dugar and Nathan (1995) find that analysts covering companies with whom their brokerage has an investment banking relationship are likely to issue optimistic – and consequently, inaccurate – forecasts. It is widely accepted that brokerages encourage optimistic forecasts to generate remunerative underwriting business.

(3) Herrmann and Thomas (2005) find that forecasts rounded to the nearest nickel are less accurate than other forecasts. Analysts who "heap" their forecasts tend to use ad hoc methods of forecasting which are less accurate than detailed models.

(4) Clement and Tse (2005) find that analysts making bold forecasts revisions are more accurate than herding analysts. A bold forecast revision, which moves away from the consensus forecast, indicates that the analyst is using new information.

REFERENCES

Abarbanell, J., 1991. "Do analysts' earnings forecasts incorporate information in prior stock price changes?" *Journal of Accounting and Economics*, Vol. 14, 147-165

Abarbanell, J. and V. Bernard, 1992. "Test of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior". *Journal of Finance*, Vol. 47, No. 3, 1181-1207

Ackert, L and W. Hunter, 1995. "Rational Expectations and Security Analysts' Earnings Forecasts". *The Financial Review*, Vol. 30, No. 3, 427-443

Albrecht, W., O. Johnson, L. Lookabill, and D. Watson, 1977. "A Comparison of the Accuracy of Corporate and Security Analysts' Forecasts of Earnings: A Comment". *The Accounting Review*, Vol. 52, No. 3, 736-740

Alexander, J., 1995. "Refining the Degree of Earnings Surprise: A Comparison of Statistical and Analysts' Forecasts". *The Financial Review*, Vol. 30, No. 3, 469-506

Ali, A., A. Klein and J. Rosenfeld, 1992. "Analysts' Use of Information about Permanent and Transitory Earnings Components in Forecasting Annual EPS". *The Accounting Review*, Vol. 67, 183-198

Ashton, A. and R. Ashton, 1985. "Aggregating Subjective Forecasts: Some Empirical Results". *Management Science*, Vol. 31, No. 12, 1499-1508

Ball, R. and E. Bartov, 1996. "How Naïve Is the Stock Market's Use of Earnings Information?" *Journal of Accounting and Economics*, Vol. 21, 319-337

Bates, J. and C. Granger, 1969. "The Combination of Forecasts". *Operational Research Quarterly*, Vol. 20, No. 4, 451-468

Bonner, S., B. Walther and S. Young, 2003. "Sophistication-Related Differences in Investors' Models of the Relative Accuracy of Analysts' Forecast Revisions". *The Accounting Review*, Vol. 78, No. 3, 679-706

Brandon, C. and J. Jarrett, 1979. "Revising Earnings per Share Forecasts: An Empirical Test". *Management Science*, Vol. 25, No. 3, 211-220

Brandon, C., J. Jarrett and S. Khumawala, 1983. "Revising Forecasts of Accounting Earnings: A Comparison with the Box-Jenkins Method". *Management Science*, Vol. 29, No. 2, 256-263

Brown, L., 1991. "Forecast Selection When All Forecasts Are Not Equally Recent". *International Journal of Forecasting*, Vol. 7, 349-356

Brown, L., 1993. "Earnings Forecasting Research: Its Implications for Capital Markets Research". *International Journal of Forecasting*, Vol. 9, 295-320

Brown, L., 1997. "Analyst Forecasting Errors: Additional Evidence". *Financial Analysts Journal*, 81-88

Brown, L, 2001. "How important is past analyst forecast accuracy?" *Financial Analysts Journal*, 44-49

Brown, L. and K. Kim, 1991. "Timely Aggregate Analyst Forecasts As Better Proxies for Market Earnings Expectations". *Journal of Accounting Research*, Vol. 29, No. 2, 382-385

Brown, L. and E. Mohammad, 2001. "Profiting from Predicting Individual Analyst Earnings Forecast Accuracy". Working paper, Georgia State University

Budescu, D., A. Rantilla, H. Yu and T. Karelitz, 2003. "The Effects of Asymmetry among Advisors on the Aggregation of Their Opinions". *Organizational Behavior and Human Decision Processes*, Vol. 90, No. 1, 178-194

Bunn, D., 1975. “A Bayesian Approach to the Linear Combination of Forecasts”. *Operational Research Quarterly*, Vol. 26, No. 2, Part 1, 325-329

Butler, K. and L. Lang, 1991. “The Forecast Accuracy of Individual Analysts: Evidence of Systematic Optimism and Pessimism”. *Journal of Accounting Research*, Vol. 29, No. 1, 150-156

Carleton, W., C. Chen and T. Steiner, 1998. “Optimism Biases Among Brokerage and Nonbrokerage Firm’s Equity Recommendations: Agency Costs in the Investment Industry”. *Financial Management*, Vol. 27, 17-30

Carter, R and S. Manaster, 1990. “Initial Public Offerings and Underwriter Reputation”. *Journal of Finance*, Vol. 45, 1045-1067

Chen, C., K. Chan, and T. Steiner, 2002. “Are All Security Analysts Equal?” *The Journal of Financial Research*, Vol. 25, No. 3, 415-430

Clement, M., 1999. “Analyst Forecast Accuracy: Do ability, resources, and portfolio complexity matter?” *Journal of Accounting and Economics* 27, 285-303

Clement, M. and S. Tse, 2005. “Financial Analyst Characteristics and Herding Behavior in Forecasting”. *The Journal of Finance*, Vol. 55, No. 1, 307-341

Conroy, R. and R. Harris, 1987. “Consensus Forecasts of Corporate Earnings: Analysts’ Forecasts and Time Series Methods”. *Management Science*, Vol. 33, No. 6, 725-738

Das, S., C. Levine and K. Sivaramakrishnan, 1998. “Earnings Predictability and Bias in Analysts’ Earnings Forecasts”. *The Accounting Review*, Vol. 73, No. 2, 277-294

Diether, K., C. Malloy, and A. Scherbina, 2002. “Differences of Opinion and the Cross Section of Stock Returns”. *The Journal of Finance*, Vol. 57, No. 5, 2113-2141

Dugar, A. and S. Nathan, 1995. “The Effect of Investment Banking Relationships on Financial Analysts’ Earnings Forecasts and Investment Recommendations”. *Contemporary Accounting Research*, Vol. 12, No. 1, 131-160

Easterwood, K. and S. Nutt, 1999. “Inefficiency in Analysts’ Earnings Forecasts: Systematic Misreaction or Systematic Optimism?” *The Journal of Finance*, Vol. 54, No. 5, 1777-1797

Elgers, P. and M. Lo, 1994. “Reductions in Analysts’ Annual Earnings Forecast Errors Using Information in Prior Earnings and Security Returns”. *Journal of Accounting Research*, Vol. 32, No. 2, 290-303

Elgers, P., M. Lo and D. Murray, 1995. “Note on Adjustments to Analysts’ Earnings Forecasts Based Upon Systematic Cross-Sectional Components of Prior-Period Errors”. *Management Science*, 1392-1396

Francis, J., 1997. “Discussion of Self-Selection and Analyst Coverage”. *Journal of Accounting Research*, Vol. 35, 201-208

Francis, J. and D. Philbrick, 1993. “Analysts’ Decisions as Products of a Multi-Task Environment”. *Journal of Accounting Research*, Vol. 31, 216-230

Herrmann, D. and W. Thomas, 2005. “Rounding of Analyst Forecasts”. *The Accounting Review*, Vol. 80, No. 3, 805-823

Hogarth, R., 1979. “A Note on Aggregating Opinions”. *Organizational Behavior and Human Performance*, Vol. 21, No. 1, 40-46

Holden, K. and D. Peel, 1988. “Combining Economic Forecasts”. *The Journal of the Operational Research Society*, Vol. 39, No. 11, 1005-1010

Hong, H. and J. Kubik, 2003. “Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts”. *The Journal of Finance*, Vol. 58, No. 1, 313-350

Hong, H., J. Kubik and A. Solomon, 2000. "Security Analysts' Career Concerns and Herding of Earning Forecasts". *The Rand Journal of Economics*, Vol. 31, No. 1, 121-144

Jacob, J., T. Lys and M. Neale, 1999. "Expertise in Forecasting Performance of Security Analysts". *Journal of Accounting and Economics*, Vol. 28, No. 1, 51-82

Lawrence, M., R. Edmundson and M. O'Connor, 1986. "The Accuracy of Combining Judgemental and Statistical Forecasts". *Management Science*, Vol. 32, No. 12, 1521-1532

Lim, T., 2001. "Rationality and Analysts' Forecast Bias". *The Journal of Finance*, Vol. 56, No. 1, 369-385

Lo, M. and P. Elgers, 1998. "Alternative adjustments to analysts' earnings forecasts: relative and complementary performance". *The Financial Review*, Vol. 33, 99-114

Lobo, G., 1991. "Alternative Methods of Combining Security Analysts' and Statistical Forecasts of Annual Corporate Earnings". *International Journal of Forecasting*, Vol. 7, No. 1, 57-63

Lorek, K., C. McDonald and D. Patz, 1976. "A Comparative Examination of Management Forecasts and Box-Jenkins Forecasts of Earnings". *The Accounting Review*, Vol. 51, No. 2, 321-330

Makridakis, S. and R. Winkler, 1983. "Averages of Forecasts: Some Empirical Results". *Management Science*, Vol. 29, No. 9, 987-996

Markov, S. and A. Tamayo, 2006. "Predictability in Financial Analyst Forecast Errors: Learning or Irrationality?" *Journal of Accounting Research*, Vol. 44, No. 4, 725-761

Matsumoto, D., 2002. "Management's Incentives to Avoid Negative Earnings Surprises". *The Accounting Review*, Vol. 77, 483-514

Mikhail, M., B. Walther, and R. Willis, 1997. "Do Security Analysts Improve Their Performance with Experience?" *Journal of Accounting Research*, Vol. 35, 131-157

Mikhail, M., B. Walther and R. Willis, 1999. "Does Forecast Accuracy Matter to Security Analysts?" *The Accounting Review*, Vol. 74, No. 2, 185-200

Mikhail, M., B. Walther and R. Willis, 2003. "The Effect of Experience on Security Analyst Underreaction". *Journal of Accounting and Economics*, Vol. 35, 101-116

O'Brien, P., 1990. "Forecast Accuracy of Individual Analysts in Nine Industries". *Journal of Accounting Research*, Vol. 28, No. 2, 286-304

Park, C. and E. Stice, 2000. "Analyst Forecasting Ability and the Stock Price Reaction to Forecast Revisions". *Review of Accounting Studies*, Vol. 5, 259-272

Philbrick, D. and W. Ricks, 1991. "Using Value Line and IBES Analyst Forecasts in Accounting Research". *Journal of Accounting Research*, Vol. 29, No. 2, 397-417

Ramnath, S., S. Rock and P. Shane, 2005. "Value Line and I/B/E/S earnings forecasts". *International Journal of Forecasting*, Vol. 21, 185-198

Sinha, P., L. Brown and S. Das, 1997. "A Re-Examination of Financial Analysts' Differential Earnings Forecast Accuracy". *Contemporary Accounting Research*, Vol. 14, 1-42

Stickel, S., 1990. "Predicting Individual Analyst Earnings Forecasts". *Journal of Accounting Research*, Vol. 28, No. 2, 409-417

Stickel, S., 1992. "Reputation and Performance among Security Analysts". *The Journal of Finance*, Vol. 47, No. 5, 1811-1836

Winkler, R. and R. Clement, 1992. "Sensitivity of Weights in Combining Forecasts". *Operations Research*, Vol. 40, No. 3, 609-614

Winkler, R. and S. Makridakis, 1983. "The Combination of Forecasts". *Journal of the Royal Statistical Society*, Vol. 146, No. 2, 150-157