Analyst Forecasting Errors and Their Implications for Security Analysis

David N. Dreman and Michael A. Berry

A comparison of 66,100 consensus estimates of Wall Street analysts with reported earnings for a large sample of NYSE, Amex, and OTC companies demonstrates that their forecasts differ significantly from actual reported earnings. A minority of estimates fall within a range around reported earnings considered acceptable to many professional investors. The error rates are not meaningfully affected by the business cycle or industry groupings. The average error also appears to be increasing over time. These findings question the use of finely calibrated earnings forecasts that are integral to the most common valuation models and indirectly question the valuation methods themselves.

A large part of the research budget of the brokerage industry is expended on hiring top analysts to provide accurate earnings estimates. Professional investors also rely on commercial earnings forecasting services such as Institutional Broker's Estimate System (IBES), Zacks, Value Line, and First Call, which maintain records of all estimates and rapidly relay brokerage house earnings forecast changes to the marketplace. First Call, for example, provides instant release of analysts' estimate changes together with detailed analysis for each company.

Financial academics and investment professionals agree that earnings are a major determinant of stock prices. The heart of modern security analysis centers on the attempt to predict stock price movements by fine-tuning near-term earnings estimates. This practice has continued in spite of the warnings by Graham and Dodd in the early 1930s and by other knowledgeable market observers over the decades about the difficulties of forecasting earnings precisely. A significant component of the research effort of the brokerage industry is directed at producing accurate short-term earnings estimates. The requirement for precise earnings estimates has been increasing in recent years. An examination of the reactions of stock prices to earnings surprises indicates that very small percentage misses may cause large changes in price.[1] Indeed, many market professionals consider a forecast error magnitude of plus or minus 10 percent of actual or forecast earnings enough to trigger a major stock reaction.[2]

Accurate earnings estimates are also essential for most contemporary stock valuation models. The intrinsic value theory of stock selection that is used extensively in earnings, dividend, and cash flow discount models is based on the ability of practitioners to forecast earnings accurately often a decade or more into the future. The growth and momentum schools of investing also require finely calibrated, precise earnings estimates years into the future to achieve the valuations they place on securities.

We examined the forecasting accuracy of analysts by comparing their consensus forecasts with reported earnings. We demonstrate that consensus forecasts, revised as recently as two weeks prior to the end of the quarter for which the earnings forecasts were made, deviate significantly and consistently from actual earnings. Using four different surprise measures, we found that only a relatively small percentage of earnings estimates fall into what many professional investors consider to be acceptable ranges around the reported earnings.[3] We believe that analysts' forecast errors are systematically too large for many analytical valuation methods to provide consistent results. This finding allowed us to hypothesize about some

David N. Dreman is chairman of Dreman Value Management L.P. in New Jersey. Michael A. Berry is the Wheat First Professor of Investments at James Madison University in Virginia and managing director of Dreman Value Management L.P.

behavioral aspects of the earnings forecasting process.

LITERATURE REVIEW

Neiderhoffer and Regan examined the 50 best-performing and the 50 worst-performing stocks listed on the NYSE in 1970 with respect to actual and forecast earnings.[4] The median earnings forecast was 7.7 percent for the "best" group, although the average increase in earnings was 21.4 percent. This forecast resulted in a price appreciation of the sample median by 44.4 percent. Earnings for the bottom 50 stocks declined by a median of 56.7 percent and price declined by 83 percent. The authors concluded that "it is clear that an accurate earnings estimate is of enormous value in stock selection (p. 71)."

Copeland and Mariani reviewed management estimates, because most institutional analysts interview management to fine-tune their earnings estimates.[5] They used deviation as a percent of actual 1968 earnings to compare the estimates of 50 executives published in the *Wall Street Journal* against year-end results. The absolute error was 20.1 percent. Green and Segall, McDonald, and Basi, Carey, and Twark also analyzed and compared management forecasts against actual results for the 1970–71 period.[6] Basi et al. used both absolute deviation and percentage of actual earnings to measure the size of analyst and management forecast errors. They showed that analysts and management, on average, tend to overestimate earnings. Both groups generated an average error greater than 10 percent. Company management in these four studies exhibited an average error of 14.5 percent, even after outliers resulting from nominal forecasts had been deleted.

The literature on analysts' forecast errors is similar. Basi et al. also studied the error rate of analysts for the 1970–71 period and found it to average 40 percent greater than that of the executives. Richards and Frazer found that analysts' mean consensus error for 1973 was 22.7 percent; in addition, analyst forecasts tended to cluster.[7] Richards, Benjamin, and Strawser used error as a percent of actual earnings to examine analysts' estimates between 1972 and 1976.[8] They found an average annual error rate of 24.1 percent. Analysts exhibited average error rates of 59.6 percent in 1974. This study also showed that the consensus earnings forecast deviated significantly from realized earnings and that individual estimates clustered.

Dreman noted in reviewing early studies that the composite forecast error from 1960 through 1976 was 16.6 percent.[9] He posited that practicing analysts try to fine-tune their earnings estimates to within a very narrow range, normally well under plus or minus 10 percent of reported earnings, so the average error rates found on consensus estimates are highly significant. Little and Rayner and Brealey also documented the randomness of earnings changes.[10] Cragg and Malkiel studied the earnings projections of large groups of security analysts. The researchers found that most analysts' estimates were simply linear extrapolations of recent trends.[11] Dreman postulated that if changes in earnings follow a random walk, projecting current trends into the future, as Malkiel suggests that analysts do, should lead to the significant forecasting errors that the literature demonstrates.[12]

Recently, researchers have reexamined the hypothesis that analysts are poor forecasters. Imhoff and Pare compared the forecasts of analysts and management using four surprise metrics: percent of forecast, percent of actual, absolute difference between forecast and actual, and percent of the standard deviation of the actual.[13] They also used four different types of naive econometric models for comparative purposes. They measured the relative errors between forecast and actual earnings and concluded that no significant differences are observed between the forecast agents. This result implies that analysts do not outperform naive econometric models in forecasting earnings.

Ou and Penman developed a single financial statement measure to forecast the change in direction in a company's earnings per share (EPS) during the next year.[14] Strober tested this measure and found that it has earnings forecasting value up to 36 months into the future.[15] He surmised that the measure impounds a risk factor not perceived by analysts in their expectations for future earnings and concluded that this forecasting model provides direct evidence of the inability of analysts to forecast earnings with a high degree of accuracy.

Ali, Klein, and Rosenfeld suggested that neither markets nor analysts recognize the time series properties of quarterly earnings surprises.[16] In their study, Ali et al. cited Bernard and Thomas.[17] Ali et al. also showed that analysts, on average, underestimate the permanence of the previous year's forecast error when forecasting earnings. Abarbanell and Bernard found that analysts do not use the time series properties of earnings correctly in forecasting earnings.[18] These results provide evidence supporting the hypothesis that analysts systematically misforecast future earnings.

Thus, the literature from 1967 forward clearly suggests that analysts consistently misforecast earnings but does not provide a rationale for the persistence, size, or increasing trend of the error.

METHODOLOGY

We analyzed consensus earnings estimates derived from the Abel Noser data base. This data base comprises approximately 1,200 companies followed by analysts from 1972 through March 1991. The study begins in 1974 to allow two years of previous earnings to form "standardized" surprise metrics. The Abel Noser data base was used because it contains 17 years of quarterly earnings estimates, the longest such data base that we are aware of.[19] In addition, we used the same measure the market appears to use to capitalize a firm's expected earnings, a single consensus point estimate of earnings. We analyzed 69 quarters of earnings surprise data. Because increasing error rates were noted after IBES and Zacks introduced quarterly estimate reporting in 1984, we concluded that the effect of not having formalized reporting in the early 1970s through the early 1980s by all the services is minimal.

The sample size increased over this timeframe. In 1974, the sample yielded 2,451 surprise observations with valid estimates; in 1990, it yielded 4,057 surprise observations. The data base contained 66,100 observations from the first quarter of 1974 through the first quarter of 1991, each representing a single firm's quarterly consensus earnings estimate.

The stocks in the Abel Noser sample were matched to the Compustat data base to determine the fiscal year and adjustment factors for stock splits for each company. Only companies with fiscal year-ends in March, June, September, or December were included in the study. A firm's share price was verified by matching Abel Noser data to the Compustat data base.

After 1981, companies included in the Abel Noser data base must have been followed by at least four analysts. In 1993, an Abel Noser company was followed by an average of 11 analysts. To eliminate the possibility of survivorship bias, we tracked all stocks deleted from the Abel Noser data base from 1980 forward. The returns derived from this sample of firms experiencing bankruptcies, mergers, and insufficient analyst coverage were similar to the results for the principal sample.

Two "standardized" surprise measures were calculated by dividing the difference between actual and forecast earnings per share by the standard deviation of actual earnings per share for the past eight quarters (SURP8) and the standard deviation of the change in actual earnings per share for the past seven quarters (SURPC7). This standardization permitted a test of a volatility-adjusted error on the sample as a whole and for each industry yearly and for the entire period. Standardized surprise metrics such as SUEs (standardized unexpected earnings) are often used in the academic literature to correlate with returns rather than to provide a measure of the size of the surprise. Note that the absolute and standardized measures cannot be directly compared with each other and that the value to investors of one versus the other is not at issue in this paper. We documented that the sizes of these surprises are large, on average, relative to contemporary investment practice.

In total, we defined the following four earnings surprise metrics:[20]

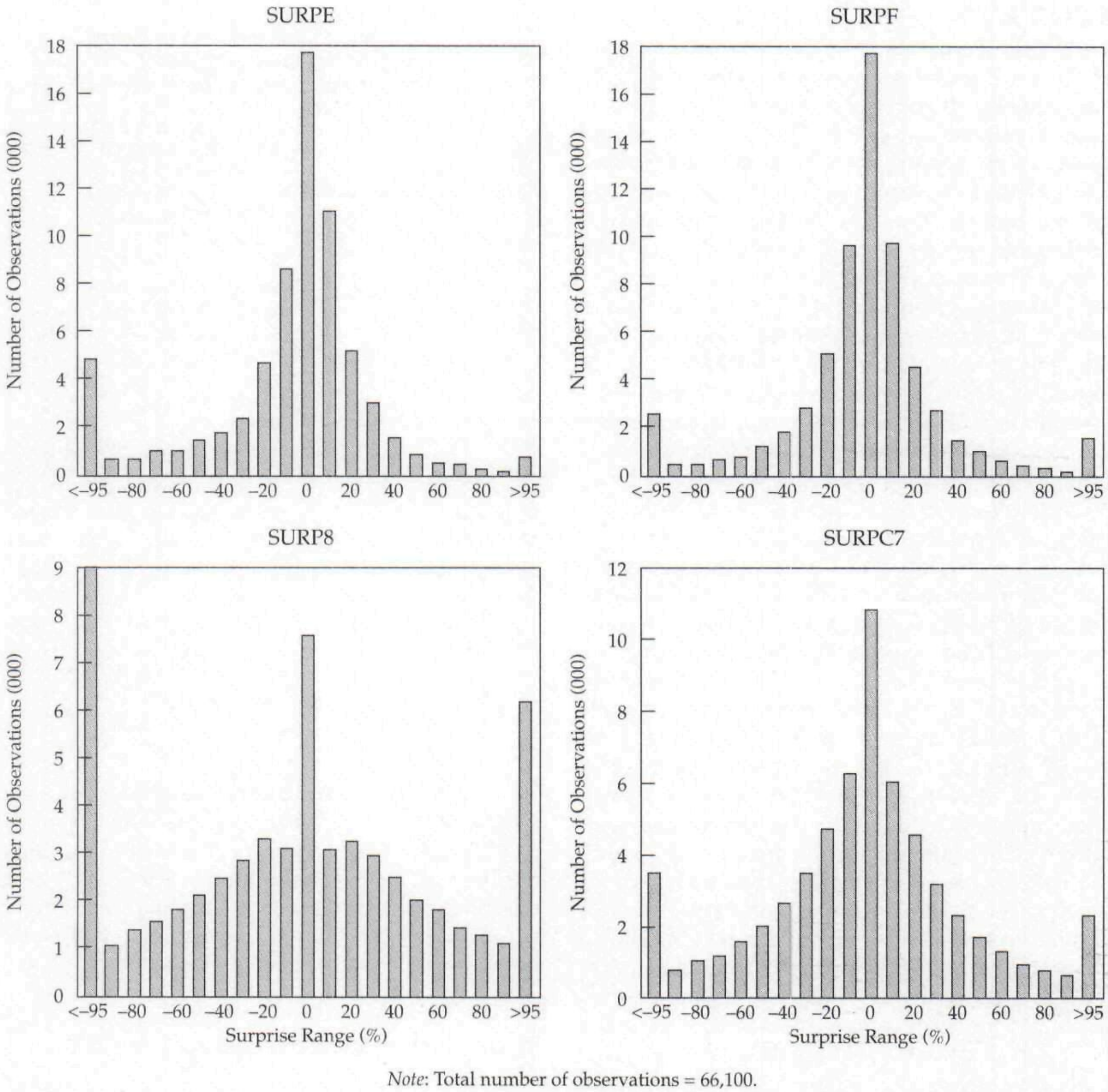
- SURPE: Consensus EPS surprise as a percent of absolute value of actual EPS—(Actual EPS – Forecast EPS)/|(Actual EPS)|
- SURPF: Consensus EPS surprise as a percent of absolute value of forecast EPS—(Actual EPS – Forecast EPS)/|(Forecast EPS)|
- SURP8: Consensus EPS surprise as a percent of the past eight-quarter volatility of actual EPS—(Actual EPS – Forecast EPS)/(Standard deviation of trailing eight-quarter actual EPS).
- SURPC7: Consensus EPS surprise as a percent of the past seven-quarter volatility of change in actual EPS—(Actual EPS – Forecast EPS)/(Standard deviation of trailing seven-quarter change in EPS).

The summary statistics and sampling distributions of these metrics were estimated and observations made regarding the absolute magnitude, central tendency, and distribution of observations of each of the metrics. Results of these tests are consistent with the previous forecasting literature.

For each year, the four quarterly consensus earnings surprises were estimated for each company in the sample. The sample was pooled across all companies and years, and t-statistics were estimated for each surprise metric to test the hypothesis that the mean surprise was different from zero. Descriptive statistics were estimated for positive and negative surprises for each surprise metric separately.

A second sample was created by deleting all surprises with reported or forecast EPS between

Figure 1. Histograms of Earnings Surprise Measures, Quarterly Observations, First Quarter 1974–Fourth Quarter 1991



Note: Total number of observations = 66,100.

plus and minus 10 cents for all four surprise metrics. This number was reduced to 55,650 stocks after the deletions. We had two motives for creating this sample. Neiderhoffer and Regan pointed out the difficulty of using an error metric with actual earnings as the denominator because this technique "becomes statistically cumbersome whenever the base (actual or forecast earnings) is small or negative."[21] By deleting all stocks with EPS between plus and minus 10 cents, we were able to control for a large part of this negative bias problem for the SURPE and SURPF results. Second, by using this technique, we controlled for the potential for outliers to dominate the results. We were able to determine the impact of large errors on stocks with small actual or forecast earnings on the distribution. Using this approach, we found that the impact of nominal earnings and forecasts is negligible and that large errors are valid misses, not outliers.[22]

RESULTS

The distributional results of each of the surprise metrics for the total sample are shown in Figure 1, the frequency distributions of earnings surprises. These histograms of quarterly earnings surprises appear to be approximately normally distributed with a central tendency around zero percent. The tails are "fatter" than expected, however, and the distribution slightly more peaked and negatively skewed in each case. This configuration is consistent with research showing that analysts tend to be overly optimistic in making earnings forecasts. The sampling distribution of the SURPE metric is skewed slightly to the positive side of the surprise distribution with the exception of a large number of large negative surprises.[23] This distribution is a result of the definition of this surprise metric, which tends to increase the size of negative surprises. The histogram of the SURPF metric (percent of forecast) appears to be more symmetric with fewer negative and more positive outliers than the SURPE metric. The two metrics representing standardized surprises exhibit a larger number of outliers, and the tails of these distributions appear to be "fatter" than normal, even though the mean and median more nearly coincide. One general conclusion from an inspection of the histograms is that a large number of outliers exist when surprise is measured by any of the four criteria. The four surprise metrics were reestimated for a reduced sample that excluded all actual and forecast earnings between plus and minus 10 cents. The results were not significantly different from those obtained for the full sample. The appendix addresses the issue of the identification and importance of outliers in this analysis.

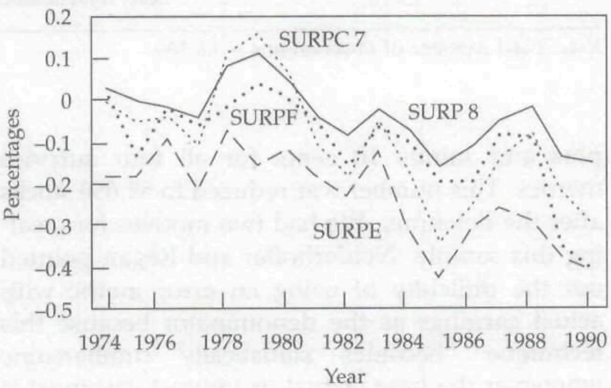
Table 1 makes evident that the mean surprise is negative irrespective of the choice of surprise metric. Specifically, the *t*-tests indicate that all the metrics generated average surprises that were less than zero at the 99.9 percent level of significance. A priori, we would expect analysts to achieve a mean-zero forecast error. These results verify that analysts tend to be optimistic over time in their forecasts. Negative surprises outnumbered positive surprises (SURPE) by 3,241 out of a sample of 66,100 observations, and the mean of negative surprises was always larger in absolute magnitude than that of the positive surprises. Table 1 also reveals that the average absolute value of the surprise over this period was large, averaging 43.8 percent of actual and 41.5 percent of forecast earnings.

Table 1. Descriptive Statistics for Earnings Surprise Measures, Quarterly Observations, First Quarter 1974–First Quarter 1991

Statistic	SURPE	SURPF	SURP8	SURPC7
All surprises (66,100 observations)				
Average absolute surprise	43.8%	41.5%	81.0%	42.2%
Mean	–0.250	–0.111	–0.136	–0.049
Standard deviation	2.208	1.961	1.409	0.620
Median	0.000	0.000	0.000	0.000
Maximum	49.000	48.000	30.425	30.500
Minimum	–216.000	–282.600	–78.160	–23.270
t-test for difference of mean from zero	–28.14	–14.07	–24.11	–19.64
Positive surprises (26,122 observations)				
Mean	0.234	0.316	0.706	0.392
Standard deviation	0.922	0.961	0.810	0.455
Median	0.117	0.132	0.477	0.254
Maximum	49.000	48.000	30.425	30.500
Minimum	0.002	0.002	0.002	0.001
Negative surprises (29,363 observations)				
Mean	–0.733	–0.514	–0.915	–0.452
Standard deviation	–0.734	2.630	1.530	0.537
Median	–0.184	–0.157	–0.554	–0.284
Maximum	–0.002	–0.002	–0.004	–0.002
Minimum	–216.000	–282.600	–78.160	–23.270

The SURPE and SURPF magnitudes are consistent with the findings of Basi et al. and Richards et al. for the 1972–76 period, thus confirming their findings but more importantly extending these results into a sample period in which one might expect surprises to be diminishing in both size and frequency of occurrence (see Figure 2).

Figure 2. Mean Value of Surprises over Time



PROPORTION OF SAMPLE OUTSIDE OF FORECAST BANDS

As shown in Table 2, the average quarterly earnings surprise for the entire sample in this time period was significantly greater than plus or minus 10 percent. Irrespective of the type of earnings surprise, on average, a minimum of 55.6 percent of SURPE and 55.5 percent of SURPF observations fell outside the 10 percent error bandwidth.[24] This target was exceeded, on average, in every year of the period and overall for both samples and both types of earnings surprises. These results indicate that analysts who try to fine-tune their forecasts to within plus or minus 5 percent or 10 percent of actual earnings are, on average, unsuccessful.

Table 2 also shows the proportion of analyst consensus forecasts that fell outside of practical error bands during this time period. For instance, in the full sample, 73.3 percent of all the SURPE estimates fell outside a plus or minus 5 percent interval around the actual earnings, 55.6 percent fell outside of plus or minus 10 percent, and 43.75 percent fell outside of plus or minus 15 percent. The proportions falling outside of these error bands for the other metrics was equally large. These results are significant in that the sample size is large and the time frame is 18 years. The proportions falling outside the respective error bands did not vary significantly in trend over time. For the reduced sample, the results are equally significant. Even with the largest surprises deleted, the proportions of analysts' estimates outside the three bands for SURPE were 71.1 percent, 50.7 percent, and 37.9 percent, respectively.[25] Both sets of these results imply that analysts miss their targets by at least plus or minus 10 percent half the time and plus or minus 5 percent almost three quarters of the time.[26]

The use of percentage bandwidths for the standardized surprises requires a slightly different interpretation than surprises measured as a percentage of actual or forecast earnings. Technically, the four metrics are not directly comparable. In the case of a standardized surprise metric, such as SURPC7, the measure is a percent of the volatility of the change in actual earnings. To judge the size of a 42.2 percent surprise in this case, we must consider the size of the standard deviation of the dollar change in actual earnings. If we assume a normal sampling distribution for the surprise metric, one standard deviation on either side of the mean encompasses 68.2 percent of the probability in the distribution. Because each year, on average, 62.75 percent of the surprises fell outside of 15 percent of one standard deviation and the absolute magnitude of the mean error was 42.2 percent of one standard deviation of earnings change (from Table 1), these errors appear to be quite large relative to volatility of earnings changes for the entire sample. For example, if the average stan-

Table 2. Proportion of Forecast Errors Outside of Percentage Bandwidths, First Quarter 1974–First Quarter 1991

	SURPE			SURPF			SURP8			SURPC7		
Year	+/–5%	+/–10%	+/–15%	+/–5%	+/–10%	+/–15%	+/–5%	+/–10%	+/–15%	+/–5%	+/–10%	+/–15%
1974	0.765	0.611	0.491	0.765	0.612	0.493	0.86	0.838	0.807	0.839	0.765	0.695
1975	0.778	0.625	0.499	0.779	0.623	0.503	0.865	0.836	0.796	0.83	0.747	0.661
1976	0.746	0.564	0.421	0.745	0.558	0.424	0.853	0.819	0.775	0.809	0.717	0.635
1977	0.706	0.507	0.383	0.707	0.51	0.391	0.85	0.82	0.772	0.818	0.722	0.632
1978	0.698	0.508	0.37	0.699	0.516	0.383	0.845	0.813	0.768	0.805	0.715	0.635
1979	0.72	0.526	0.398	0.72	0.535	0.41	0.864	0.827	0.782	0.827	0.734	0.652
1980	0.732	0.549	0.426	0.732	0.551	0.429	0.878	0.836	0.789	0.83	0.739	0.652
1981	0.728	0.538	0.417	0.73	0.538	0.421	0.875	0.831	0.783	0.825	0.73	0.64
1982	0.742	0.565	0.458	0.742	0.562	0.458	0.867	0.811	0.762	0.809	0.7	0.593
1983	0.741	0.566	0.449	0.743	0.566	0.451	0.876	0.821	0.764	0.817	0.705	0.609
1984	0.727	0.556	0.436	0.727	0.551	0.433	0.871	0.826	0.769	0.821	0.719	0.619
1985	0.752	0.58	0.474	0.752	0.571	0.462	0.893	0.843	0.792	0.834	0.727	0.634
1986	0.752	0.57	0.452	0.752	0.562	0.449	0.893	0.84	0.79	0.834	0.729	0.624
1987	0.718	0.542	0.424	0.72	0.543	0.425	0.888	0.829	0.769	0.831	0.705	0.598
1988	0.711	0.522	0.397	0.712	0.524	0.405	0.895	0.833	0.77	0.833	0.716	0.608
1989	0.72	0.532	0.422	0.719	0.528	0.416	0.9	0.836	0.774	0.834	0.712	0.607
1990	0.718	0.556	0.453	0.718	0.551	0.444	0.903	0.835	0.769	0.832	0.703	0.6
1991	0.744	0.599	0.505	0.743	0.596	0.49	0.906	0.837	0.765	0.83	0.707	0.602
Average	0.733	0.556	0.438	0.734	0.555	0.438	0.877	0.830	0.778	0.825	0.722	0.628

dard deviation of change in earnings is 25 cents, then analysts miss their target, on average, by 10.5 cents and 62.75 percent of the time they would miss actual earnings by a minimum of 4 cents.

We conclude that if analysts try to fine-tune their earnings estimates to within plus or minus 10 percent of actual earnings, they do not perform this task well.

FORECAST ERROR BY INDUSTRY

To determine whether a significant proportion of the overall mean and median earnings surprise was attributable to a small number of industries in a few time periods, we classified our sample by two-digit Standard Industrial Classification Code industries. Sixty-one industry portfolios were created, and all four surprise metrics were estimated for each industry for each year. In addition, we estimated median surprises for each year and industry and percentile ranking indicating the percentage of industries whose mean, median, and standard deviation of surprise exceeded decile limits.

A legitimate concern is whether the preponderance of surprises occur in highly volatile industries, thus skewing the findings for the sample as a whole. Table 3 shows the average results by industry. These statistics include the mean, median, and standard deviation of the absolute value of each of the four surprise metrics for each industry averaged over the entire time period. Irrespective of the metric, the surprises are large and emanate from many industries. Table 3 reveals that, over the entire time period, 90 percent of all industries exhibited mean surprises (SURPE) greater than 21.44 percent and median errors greater than 16 percent. Ten percent of the industries experienced average surprises greater than 84.33 percent of actual EPS. Using the standardized error measure SURPC7, 90 percent of all industries experienced mean errors greater than 27.67 percent of one standard deviation of actual EPS and median errors greater than 27 percent. Ten percent of the industries experienced average surprises over the entire time period greater than 48.06 percent of one standard deviation. Because these results are relative to changes in EPS, we consider this level of error to be large. Moreover, in examining decile boundaries, the distribution of these surprises was surprisingly uniform across industries.

A further conclusion may be drawn from this analysis. Standardized errors are large uniformly across industries, indicating that even on a volatility-adjusted basis, analysts err indiscriminately across industries. There is high earnings volatility in industries that are supposed to have high visibility and thus often are given high valuations. This volatility raises a question about whether many such valuations are excessive.

With respect to specific industry rankings, a number of results are evident. The tobacco products industry, for instance, exhibited the lowest rank for mean, median, and standard deviation of surprise for either of the two absolute measures. Although this industry ranked in the first decile for SURPE, it ranked in the seventh decile for SURPC7. Our expectation was that both mean and median surprise levels should be low in this indus-

Table 3. Deciles of Each Surprise Metric by Industry (Percent Industry Means within Decile)

Metric	10%	20%	30%	40%	50%	60%	70%	80%	90%
SURPE									
Mean	21.44%	27.44%	31.17%	35.39%	43.11%	47.61%	57.89%	69.56%	84.33%
Median	16.00	19.50	25.00	28.50	32.00	38.00	45.50	55.50	67.50
Standard deviation	12.52	17.34	21.76	24.78	32.41	34.98	41.28	53.06	90.00
SURPF									
Mean	20.72	22.22	27.06	29.00	36.83	44.44	54.39	61.61	85.22
Median	14.50	17.50	20.00	23.50	26.00	30.50	37.00	44.00	66.50
Standard deviation	10.29	12.51	16.72	18.79	25.41	37.53	49.41	63.41	101.92
SURP8									
Mean	53.33	64.00	69.33	72.67	76.61	78.78	81.06	86.50	90.61
Median	52.00	57.00	65.00	67.00	71.00	73.00	76.00	80.50	84.50
Standard deviation	13.46	15.31	18.34	20.88	22.61	29.25	35.48	41.97	48.94
SURPC7									
Mean	27.67	32.78	35.56	37.11	38.67	41.44	43.06	44.78	48.06
Median	27.00	30.50	33.50	35.50	37.00	38.50	40.50	41.50	46.00
Standard deviation	6.18	7.60	9.16	9.83	10.63	11.25	15.30	18.35	22.21

try because of the stability of demand, yet standardized surprise measures ranked the mean and medians quite high relative to other industries. In contrast, the food industry ranked in the third and second deciles, respectively, for absolute (SURPE) and standardized measures (SURPC7), indicating a similarity of rankings. Apparently, the choice of surprise metric is important.[27]

THE CYCLICAL BEHAVIOR OF EARNINGS SURPRISES

Another major question that might be raised is whether the large surprises are significantly influenced by periods of business expansion and recession, during which changing economic conditions make an analyst's task more difficult. We examined the surprise metrics in three periods of economic expansion and four periods of recession to determine whether their magnitude varied predictably across economic cycles. We hypothesized that during periods of recession, we would expect to see analysts' forecasts exceed reported earnings because they would not have fully adjusted their forecasting techniques to accommodate slow economic growth; during periods of economic expansion, we would expect to observe their forecasts fall below actual earnings and therefore exhibit more or larger negative errors.

Table 4 shows the proportion of analyst consensus for each economic expansion and contraction during the sample time period. No significant difference is apparent between the mean size of analyst errors in periods of expansion and recessions. Thus, economic conditions do not seem to affect analysts in making their earnings estimates. Clearly, however, this analysis shows that analysts tend to be overly optimistic in both expansions and recessions. Taking simple averages, the mean positive surprise (SURPE) in expansions was 23 percent and in recessions, 23 percent. For negative surprises, the corresponding statistics were −64 percent and −72 percent.[28] Uniformly across surprise metrics, the negative surprises in recessions appear to be slightly larger in absolute value than in expansions. Larger negative errors during recessions would imply that analysts' projections are optimistic. Clearly, however, the proportion of the

Table 4. Average of Earnings Surprise Measures Across All Expansions and Recessions, January 1974–March 1991
(number of observations in parentheses)

Surprise Measure	Expansion Dates						Recession Dates							
	April 1975–January 1980		August 1980–July 1981		December 1982–July 1990		November 1973–March 1975		January 1980–July 1980		July 1981–November 1982		July 1991–March 1991	
SURPE														
Positive average	0.20	(6,669)	0.24	(1,236)	0.25	(12,447)	0.22	(1,315)	0.20	(1,209)	0.30	(2,224)	0.23	(1,022)
Negative average	–0.60	(5,571)	–0.52	(1,110)	–0.80	(15,748)	–0.67	(1,373)	–0.57	(1,136)	–0.71	(2,673)	–0.93	(1,752)
All average	–0.14	(12,240)	–0.11	(2,346)	–0.31	(28,195)	–0.21	(2,688)	–0.15	(2,345)	–0.23	(4,897)	–0.47	(2,774)
(Zero observations)		(1,862)		(301)		(2,594)		(393)		(259)		(569)		(183)
SURPF														
Positive average	0.27	(6,669)	0.31	(1,236)	0.32	(12,447)	0.37	(1,315)	0.37	(1,209)	0.38	(2,224)	0.31	(1,022)
Negative average	–0.37	(5,571)	–0.38	(1,110)	–0.53	(15,748)	–0.58	(1,373)	–0.40	(1,136)	–0.60	(2,673)	–0.82	(1,752)
All average	–0.02	(12,240)	–0.01	(2,346)	–0.14	(28,195)	–0.10	(2,688)	0.00	(2,345)	–0.14	(4,897)	–0.38	(2,774)
(Zero observations)		(1,862)		(301)		(2,594)		(393)		(259)		(569)		(183)
SURP8														
Positive average	0.79	(6,669)	0.78	(1,236)	0.63	(12,447)	1.00	(1,315)	0.86	(1,209)	0.66	(2,224)	0.59	(1,022)
Negative average	–0.87	(5,571)	–0.86	(1,110)	–0.91	(15,748)	–1.10	(1,373)	–0.85	(1,136)	–0.91	(2,673)	–1.08	(1,752)
All average	0.3	(12,240)	0.00	(2,346)	–0.21	(28,195)	–0.06	(2,688)	0.03	(2,345)	–0.18	(4,897)	–0.44	(2,774)
(Zero observations)		(1,862)		(301)		(2,594)		(393)		(259)		(569)		(183)
SURPC7														
Positive average	0.45	(6,669)	0.41	(1,236)	0.35	(12,447)	0.56	(1,315)	0.46	(1,209)	0.35	(2,224)	0.30	(1,022)
Negative average	–0.46	(5,571)	–0.43	(1,110)	–0.45	(15,748)	–0.54	(1,373)	–0.42	(1,136)	–0.46	(2,673)	–0.46	(1,752)
All average	0.03	(12,240)	0.01	(2,346)	–0.09	(28,195)	0.00	(2,688)	0.03	(2,345)	–0.08	(4,897)	–0.17	(2,774)
(Zero observations)		(1,862)		(301)		(2,594)		(393)		(259)		(569)		(183)

Note: Cycle dates from the National Bureau of Economic Research.

overall surprise did not emanate from either economic expansions or recessions. Therefore, we concluded that large earnings surprises do not emanate from economic business cycles.

DO ANALYSTS' FORECAST ERRORS INCREASE WITH TIME?

To try to identify a trend in the magnitude of analysts' errors, we regressed the surprise for each stock in time period t against the surprise for the period $t - 1$. We estimated this statistical relationship to determine whether the surprises appear to be increasing over time. This process was replicated, with suitable adjustments for autocorrelation, for each surprise type for the entire time period for both samples, as well as for the positive, negative, and pooled surprise subsamples. The regression equation took the following form:

$$(\text{Avg. surprise})_t = \alpha + \delta\,(\text{Avg. surprise})_{t-1} + \epsilon_t,$$

where $\epsilon_t \sim N(0, \sigma^2)$; $E(\epsilon_i, \epsilon_j) = 0\ \forall i \neq j$.

The interpretation of the regression intercept, α, is the mean error at the beginning of the time period. The coefficient δ may be interpreted as the average percent change in the mean error between two quarters over the entire time period. Thus, if δ is significant and positive, positive errors are increasing over time. Because of the presence of autocorrelation in the residuals, Cochran–Orcutt transformations were applied to the data. Table 5 indicates that analysts errors are increasing over time. We obtained highly significant t-statistics on the slope coefficients, δ. With one exception, these results obtained for the entire sample and for positive as well as negative surprises. In addition, as Table 5 shows, the intercept of each of the regressions is highly significant for all four metrics, indicating that analysts tend to be optimistic in their forecasting. Intercepts for negative surprises were much larger than for positive surprises, reaffirming both an increasing error trend and the analysts' tendency toward overoptimism.

The rates of change for the entire sample, σ, were large and highly significant for each of the four metrics. These rates were approximately the same, and the intercept terms were also highly significant for all metrics. The rate of change for positive surprises for the four metrics appeared to be larger than for negative surprises and was highly significant.

This observation supports the previous one: The size and trend of consensus forecast errors make the dependence on most forecasting techniques that require single-point earnings estimates to be close to actual earnings unreliable as a primary investment technique.

Table 5. Regression Test Results, Trend in Analysts' Forecasting Errors (Full Sample)

Metric	α[a]	t-Statistic	δ[a]	t-Statistic
All surprises				
SURPE	–0.094	–3.46***	0.636	6.48***
SURPF	–0.031	–2.51**	0.745	8.61***
SURP8	–0.052	–2.53**	0.652	7.23***
SURPC7	–0.022	–2.21*	0.603	6.35***
Positive surprises				
SURPE	0.041	2.45**	0.825	11.80***
SURPF	0.120	3.84***	0.617	6.34***
SURP8	0.137	3.45***	0.796	14.47***
SURPC7	0.063	2.90**	0.828	15.29***
Negative surprises				
SURPE	–0.395	–4.89***	0.447	4.01***
SURPF	–0.242	–4.38***	0.519	4.82***
SURP8	–0.517	–5.05***	0.428	3.83***
SURPC7	–0.389	–6.95***	0.138	1.12

[a] Estimated value from the first equation.
* Statistically significant at the 95 percent level of confidence.
** Statistically significant at the 99 percent level of confidence.
*** Statistically significant at the 99.9 percent level of confidence.

If earnings surprises are increasing over time, these results suggest that despite increased availability of data bases and real-time reporting services, the analysts' processes for forecasting earnings are flawed. The large size of the forecasting error, even after controlling for the business cycle and industry groupings, casts doubt on the viability of valuation methodologies such as the growth or discounted cash flow approaches that require accurate near- and long-term, single-point estimates of earnings.

HOW WELL DO ANALYSTS FORECAST?

To quantify the size of forecast errors, we regressed the actual earnings for a company on the consensus forecast. We estimated the following model:

$$(\text{Actual Earnings})_t = \alpha + \beta(\text{Forecast})_t + \epsilon_t,$$

where $\epsilon_t \sim N(0, \sigma^2)$.

This regression framework was estimated for the entire sample pooled, as well as for positive and negative surprises. In addition, the regression parameters were estimated for each year for the entire sample and for positive and negative surprises respectively. The regression framework as applied to each of the error metrics provides an estimate of the size of surprise and a test of analyst overoptimism. This framework is independent of the construction of the error metric, relying instead on the regression of actual earnings on the consensus forecasts.

We tested the α and β coefficients for statistical significance using Student's t-tests and compared the coefficient of determination, R^2, across positive and negative surprise samples to ascertain differential forecasting ability. If analysts are excellent forecasters at the consensus level, we would expect the β coefficient to be equal to 1 and the α coefficient to be zero. We report results for the SURPE statistics only. In particular, we sought to determine the following:

- Do analysts miss their forecasts by a statistically significant amount?
- What is the estimate of the percent size of the miss?
- Are analysts optimistic or pessimistic on average?
- Is the error increasing over time?

Table 6 shows the results of this regression analysis. The alphas were statistically significant at the 99.9 percent level for the entire sample. This finding means that analysts overestimate earnings, on average, by a significant amount. Pooling all stocks (those with both positive and negative surprises) revealed that analysts tended to overestimate earnings by an average of 3.6 percent; the error was much larger (15 percent) on those stocks that received negative surprises. Those receiving positive surprises exhibited earnings that were, on average, 8 percent above the forecast level. This analysis reconfirms the negative bias to surprises and the tendency toward analyst optimism that we observed previously. Results from the reduced samples were similar and confirm these findings.

Table 6. Regression Results, Actual Quarterly EPS on Consensus Analyst's Forecasts
(t-statistics in parentheses)

Coefficient	Full Sample	Positive Surprises	Negative Surprises
Sample 1[a]			
Observations	66,100	26,122	29,363
β	0.99	1.03	1.00
	(305.68)***	(275.75)***	(196.52)***
α	–0.01	0.08	–0.15
	(–7.05)***	(34.37)***	(–42.33)***
R^2	0.58	0.67	0.56
Sample 2[b]			
Observations	52,582	23,735	23,766
β	1.02	1.06	1.00
	(286.80)***	(238.24)***	(193.69)***
α	–0.03	0.05	–0.14
	(–11.43)***	(17.92)***	(–36.95)***
R^2	0.60	0.66	0.60

[a] Includes all observations regardless of the value of quarterly earnings or consensus forecasts, first quarter 1974–first quarter 1991.

[b] Excludes observations with absolute values, quarterly earnings, or consensus forecasts less than 10 cents, first quarter 1974–first quarter 1991.

*** Statistically significant at the 99.9 percent level of confidence.

CONCLUSION

In this study, we examined a data base of consensus analysts' forecasts from 1974 through the first quarter of 1991 and found that errors are larger than one might expect; that they are increasing over time; and that analysts are optimistic on average, because the mean error is significantly negative irrespective of the surprise metric. Forecasting errors also appear to be large across industries and through various stages of the business cycle.

Two major conclusions are evident from the study. The first is that the average forecast error of more than 20 percent of actual EPS (SURPE)—more than 40 percent using nominal estimated and reported earnings—is too high for investors to rely on consensus forecasts as a major determinant of stock valuation. Second, regardless of surprise metric, only a small percentage of estimates fall into a range considered acceptable. On average, 56 percent of the estimates measured as a percent of actuals fall outside a plus or minus 10 percent range, a level that many Wall Street professionals consider minimally acceptable; approximately 45 percent fall outside a plus or minus 15 percent range. These results indicate that, on average, large earnings surprises are the rule rather than the exception.[29]

The observed frequency, size, and increasing trend of all of the error metrics for quarterly earnings estimates bring into question many important methods of stock valuation, which rely on precise earnings estimates sometimes years into the future. The growth, earnings momentum, discounted cash flow, and earnings yield techniques, for example, require fine-tuned estimates often a decade or more into the future. Thus, a significant portion of current security analysis requires a precision in earnings forecasts that is increasingly difficult for analysts to meet.[30]

A final conclusion of this study is that in spite of our own earlier findings, analysts, money managers, and investors appear to ignore the industry's poor forecasting record, although it questions the viability of many important stock valuation methods. Neither the consistency nor the size of forecasting errors, which are well documented, have been addressed. Although the frequency of large surprises in the overall sample is predictably high, market professionals react to forecast errors as though each change is unique and has a very small probability of occurring, thus warranting extensive analysis and earnings revision. We believe this phenomenon may have a behavioral explanation.

If analysts and investment professionals learn from past mistakes, as rational decision makers are expected to do, far less emphasis should be placed on forecasting within their valuation models. Analysts, given the findings, should also use broadband rather than single-point forecasts. At present, they do not. The prevailing belief is that earnings can be fine-tuned. Few recognize the persistent nature of large forecasting errors or have the ability to make adjustments for them.

We believe this lack of recognition of a major shortfall in contemporary investment methodology is likely to have its roots in a behavioral explanation. These findings may be explained by research in the discipline of psychology, which suggests that the accuracy of judgmental forecasts is influenced by cognitive biases that arise when the processing of complex information is simplified (Tversky and Kahneman[31]). Even when warned about the existence of such biases, forecasters appear not to be able to adjust for their effects (Fischhoff[32]).

Our findings raise another interesting question. Is it possible that the "best" analysts' judgmental forecasts may not be the "best" forecasts careerwise? Are analysts drawn to the consensus opinion either openly or unknowingly by the safety of the group? An estimate that is far off the consensus might pose career dangers, whereas an estimate near the group may provide the analyst with a much higher degree of safety, regardless of how inaccurate it may prove to be.

The above conclusions lead us to believe that behavioral factors may play an important role for analysts in forecasting earnings.[33]

FOOTNOTES

1. Analysts' consensus estimates as reported by a number of services, including IBES, Zacks, and First Call, are fine-tuned to the penny. Compaq Computers quarterly estimates would be stated as 75 cents a share for the second quarter of 1994 versus 40 cents one year earlier. The *Wall Street Journal*, *Barron's*, and the *New York Times* provide many examples of sharp stock declines on minimum variance between estimated and actual reported earnings. For example, in a one-week period between April 18 and April 25, 1994, Motorola reported a 46 percent increase in earnings. This increase was 1 cent short of consensus expectations (less than 1 percent), however, and Motorola's price fell 16⅛ dollars, a 15.2 percent decline. Another example was Zoll Medical, which fell 26.4 percent on a 1-cent earnings shortfall to expectations.
2. We use the error metrics, which are percentages of either actual or forecast earnings because these are most often presented in the *Wall Street Journal* and other financial publications, as well as by earnings services such as Zacks and IBES.
3. IBES reports that, since 1979, the mean absolute revision in earnings estimates for S&P 500 stocks is 12.9 percent from the beginning to the end of the year in which the forecast is made. Analysts revise their estimates by 6.34 percent in the first half of the year and 19.51 percent in the second half. This finding indicates that analysts target a high degree of accuracy in earnings forecasts. According to IBES, despite these estimate changes, analysts tend to be overoptimistic. Between 1980 and 1994, they missed their growth targets on the S&P 500 by an average of almost 65 percent (the mean earnings growth forecast was 18.41 percent versus the 6.28 percent average reported). Therefore, we hypothesized that 5 percent, 10 percent, and 15 percent error bands around reported earnings are significant targets sought by analysts and professional investors.
4. Victor Neiderhoffer and Patrick J. Regan, "Earnings Changes, Analysts' Forecasts and Stock Prices," *Financial Analysts Journal*, vol. 28, no. 3 (May/June 1972):65–71.
5. Ronald M. Copeland and Robert J. Mariani, "Executives' Forecasts of Earnings per Share Versus Forecasts of Naive Models," *The Journal of Business*, vol. 45, no. 4 (October 1972):497–512.
6. David Green and Joel Segall, "The Predictive Power of First Quarter Earnings Reports," *The Journal of Business*, vol. 40, no. 1 (January 1967):44–55; C.L. McDonald "An Empirical Examination of the Reliability of Published Predictions of Future Earnings," *The Accounting Review*, vol. 48, no. 1 (1973):502; and Bart A. Basi, Kenneth J. Carey, and Richard D. Twark, "A Comparison of the Accuracy of Corporate and Security Analysts' Forecasts of Earnings," *The Accounting Review*, vol. 51, no. 2 (April 1976):244–54.
7. R. Malcolm Richards and Donald R. Frazer, "Further Evidence of the Accuracy of Analysts' Earnings Forecasts: A Comparison among Analysts," *Journal of Economics and Business*, vol. 29, no. 3 (Spring/Summer 1977):193–97.
8. Malcolm R. Richards, James J. Benjamin, and Robert W.

Strawser, "An Examination of the Accuracy of Earnings Forecasts," *Financial Management*, vol. 6, no. 3 (Fall 1977): 78–86.

9. David Dreman, *Contrarian Investment Strategy* (New York: Random House, 1979):144.
10. I.M.D. Little and A.C. Rayner, *Higgledy Piggledy Growth Again* (Oxford, England: Basil Blackwell, 1966); and Richard A. Brealey, *An Introduction to Risk and Return from Common Stocks* (Cambridge, Mass.: MIT Press, 1968).
11. J.G. Cragg and Burton Malkiel, "The Consensus and Accuracy of Some Predictions of the Growth of Corporate Earnings," *The Journal of Finance*, vol. 23 (March 1968):67–84.
12. Dreman, *Contrarian Investment Strategy*, p. 147.
13. Eugene A. Imhoff and Paul V. Pare, "Analysis and Comparison of Earnings Forecast Agents," *Journal of Accounting Research*, vol. 20, no. 2 (Autumn 1982):429–39.
14. Jane A. Ou and Stephen H. Penman, "Accounting Measurement, Price–Earnings Ratio, and the Information Content of Security Prices," *Journal of Accounting Research*, vol. 27 (supp., 1989):111–52.
15. Thomas L. Stober, "Summary Financial Statement Measures and Analysts' Forecasts of Earnings," *Journal of Accounting and Economics*, vol. 15, no. 1 (1992):347–72.
16. Ashiq Ali, April Klein, and James Rosenfeld, "Analysts' Use of Information About Permanent and Transitory Earnings Components in Forecasting Annual EPS," *The Accounting Review*, vol. 67, no. 1 (January 1992):183–98.
17. Victor L. Bernard and Jacob K. Thomas, "Evidence that Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings," *The Journal of Accounting and Economics*, vol. 13 (1990):305–40.
18. Jeffrey S. Abarbanell and Victor L. Bernard, "Tests of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior," *The Journal of Finance*, vol. 47, no. 3 (July 1992):1181–208.
19. Value Line provided Abel Noser with quarterly earnings estimates prior to 1981. Zacks' quarterly estimates were incorporated in 1981, and IBES's in 1984. Today, Abel Noser uses Zacks, IBES, and First Call to develop consensus estimates.
20. Four different surprise metrics are used because the academic literature lacks consensus on the appropriate form of the metric. Reporting services, such as Zacks, tend to report surprise as a percent of forecast earnings or of actual earnings in their weekly documentation. Each measure has a unique set of statistical and interpretive problems. Our basic results appear invariant over the four metrics we used.
21. Neiderhoffer and Regan, "Earnings Changes, Analysts' Forecasts, and Stock Prices," p. 69.
22. Space considerations prevent these results from being presented here; they are available from the authors upon request.
23. More than 4,000 surprises (SURPE) are less than –95 percent while fewer than 1,000 are greater than +95 percent. In each of the four metrics, large negative errors (> |95 percent |) outnumber large positive surprises.
24. The proportion falling outside 10 percent for the standardized errors was even higher: 82.95 percent of SURP8 and 72.17 percent of SURPC7 errors fell outside the 10 percent bandwidth. Note, however, that these errors are with respect to the standard deviation of the changes in earnings (SURPC7).
25. The small-sample results are available from the authors upon request.
26. The results of the "reduced" sample indicate that SURPE and SURPF are identical as are SURP8 and SURPC7 with respect to proportions falling outside these error bands.
27. To save space, we have not included the ranking statistics or analysis for each of the 61 industries. Nevertheless, three findings stand out from this analysis: Standardized errors were uniformly large across industries; the absolute magnitude of these errors was high, averaging 40 percent; and the absolute measure exhibited higher volatility than the standardized measure in specific years. The findings also show very high earnings volatility in industries that should have high visibility. These results are available upon request.
28. The other surprise metrics yield the following average results:

Metric	*Surprise*	*Expansion*	*Recession*
SURPF	Positive	30%	35%
	Negative	−42	−60
SURPC7	Positive	−40	41
	Negative	−44	−47

29. These results applied to a greater extent with respect to SURPC7 and SURP8, the standardized errors. Although the interpretation is slightly different because of the relative nature of the errors, we found that 62.8 percent of SURPC7 and 77.8 percent of SURP8 estimates fell outside of plus or minus 15 percent of one standard deviation of earnings and earnings changes.
30. Many practitioners consider an earnings gain of 6–9 percent as average and 12–15 percent or more to be in the growth category. Thus, surprises of the magnitude shown make a high-growth company indistinguishable from an average-growth company.
31. A. Tversky and D. Kahneman, "Judgment under Uncertainty," *Science*, vol. 185 (1974):1124–31.
32. B. Fischhoff, "Debiasing," in *Judgement Under Uncertainty: Heuristics and Biases*, eds. D. Kahneman, P. Slovic, and A. Tversky (New York: Cambridge University Press, 1982): 422–44.

33. Support for this research was provided by the David Dreman Foundation. The authors thank W. Van Harlow, Richard Zeckhauser, Professor Domenic Cicchetti, Professor Mitchell Stern, Professor Nelson Woodard, and anonymous referees for their assistance in the preparation of this paper. The authors also thank the Abel Noser Corporation for the use of the Abel Noser data base.