Enhancing Earnings Predictability Using Individual Analyst Forecasts

MARTIN M. HERZBERG, JAMES GUO, AND LAWRENCE D. BROWN

MARTIN M. HERZBERG is director of research at Bridge/The DAIS Group in New York.

JAMES GUO completed a post-doctoral program with The DAIS Group during the initial development of its earnings model.

LAWRENCE D. BROWN is Controllers RoundTable research professor at Georgia State University in Atlanta.

Analyst earnings forecasts are broadly used in the investment decision-making process. Dividend discount, valuation, earnings momentum, and other quantitative models use analyst earnings forecasts as inputs to rank stocks on a relative basis. Performance that is based on quintile returns provides ample evidence of the value of analyst forecasts (Butman, Esser, and Herzberg [1998]; Herzberg [1998]).

The earnings consensus, or simple average of analyst forecasts of a firm's earnings, is the most prevalent piece of earnings data routinely considered. Firms such as First Call Corporation, I/B/E/S International Inc., and Zacks Investment Research engage in extensive data collection and distribution of both individual and consensus analyst earnings forecasts.

Given the simplistic approach used to compute an ordinary consensus, it seems that one should be able to develop means by which to generate superior earnings forecasts. Some, in fact, have provided evidence of this by using such earnings principles as forecast timeliness (O'Brien [1988]; Brown [1991]), forecast updating (Stickel [1990]), analyst track record (Sinha, Brown, and Das [1997]), forecast clustering (Mozes and Williams [1998]), and earnings preannouncements (Bhagat [1997]; Soffer, Thiagarajan, and Walther [1998]).

The earnings predictor model (EPM) introduced in this article uses all of these factors, among others, to generate superior earnings forecasts. We show that the earnings predictor model (EPM) is more accurate than the consensus in each one of more than 1,200 (non-independent) tests. We also show backtest results suggesting that abnormal returns can be achieved by buying (shorting) stocks when the EPM is above (below) the consensus estimate.

DATA

Our primary data source is the First Call Corporation. These data consist of individual analyst forecasts (also known as detail data) for over 7,000 stocks from 1989 to 1998. In addition to detailed forecasts, we obtain the First Call consensus and actual reported earnings per share data for the same period and preannouncement data for the period 1992–1998. Our secondary data source is The Wall Street Journal/Zacks Investment Research, which jointly compile and publish two categories of industry-determined "all stars" per year: estimate accuracy and recommendation accuracy (Dorfman [1997]). We include the estimate accuracy "all stars," and also derive several factors (discussed below): more recent forecasts, DAIS track record analysts, and updated individual analyst forecasts. All of these data are consolidated into an extensive customized data base.

Given the vast amount of data and its potential susceptibility to error, the raw source data are required to pass through numerous data integrity filters to insure quality control. Forecasts that did not clear the established criteria as defined by these data integrity filters are excluded from the process. The surviving forecasts are examined for timeliness, track record, clustering, and other earnings principles. The study uses the term clustering in two distinct ways:

1. An earnings interval measured in cents in which a predominant number of forecasts for a particular firm exists for a given forecast period (e.g., third fiscal quarter of 1997).
2. A time interval in which a large number of revisions occur subsequent to a period of few or no revisions.

Forecasts are updated, combined, classified, and assigned scores. Forecasts excluded from the process receive scores of zero. Ordinary forecasts are assigned scores of one. Superior forecasts as identified by the model (discussed below) are assigned a score of two, with higher scores given to the various combinations of superior forecasts.

METHODOLOGY

The earnings predictor model (EPM) employs a multifaceted approach (Herzberg [1998a]). It incorporates combinations of the following methodologies.

Forecast Recency

The rationale for this approach is obvious. A forecast that is only two days old is usually more accurate than one that is two months old. This is not surprising, because forecasts provided more recently are often based on information that was previously unavailable. We determine the number of more recent forecasts to be used, and impose a maximum time limit for forecast inclusions. Forecasts outside the maximum are assigned weights of zero. Aggregation is a key principle of earnings forecasting (Brown [1991]). It is frequently the case that, given two estimates, one may be too high and the other too low, but even a simple average of the two is closer to the actual than either estimate alone. Individual forecasts are, therefore, weighted and aggregated according to timeliness.

Updated Individual Analyst Earnings Forecasts

We update forecasts of analysts who have not revised on their own. Stickel [1990] shows that it is possible to predict how analysts will update their forecasts. The factors he considers are deviation from the consensus, change in the consensus, and stock price movement. He finds stock price movement to be less significant than the other two. In line with his findings, we use the actual revisions of analysts who have updated recently to update forecasts of those who have not updated recently.

Wall Street Journal/Zacks Estimate Accuracy All Stars

The Wall Street Journal, in conjunction with Zacks Investment Research, annually identifies and publishes estimate accuracy all-star analysts designated by industry. These "all-star" analysts are the ones who have provided the most accurate earnings estimates relative to the consensus for stocks in their respective industries, subject to satisfying certain requirements (Dorfman [1997]). To qualify, an analyst must cover:

- At least five stocks in the industry (exceptions are made for very small industries).
- At least two of the ten largest stocks in the industry.
- An industry for the full calendar year, and not have changed employers during that period.

DAIS Track Record Analysts

The EPM also uses a second form of track record, namely, analysts who have the best track record of predicting earnings for each firm. This firm-specific approach complements the WSJ/Zacks approach that evaluates track record on an industry-specific basis. A DAIS track record analyst must have provided forecasts for each of the four most recent quarters. We find, as Stickel [1992] does, that analysts who revise the most tend to be more accurate, and therefore we also consider frequency of forecast revisions.

Clustering

Computing a consensus restricted to estimates contained within an interval where a predominant number of the forecasts exist usually serves to quickly improve our earnings prediction. In turn, by observing the timing of these estimates, we garner information useful for updating and weighting individual forecasts.

Earnings Preannouncements

Earnings preannouncements by managers are most prevalent when there is negative news to be divulged, or when there exists a large dispersion in analyst estimates (Soffer, Thiagarajan, and Walther [1998]). Preannouncements are often precautionary moves by firms motivated by a desire to avoid potential problems with regulatory agencies. They take various forms: point estimates, ranges, or indications that the earnings results will come in above or below the consensus analyst forecast. The EPM quantifies the preannouncements, assigns values to them, and uses the preannouncement dates as conditioning events.

Summary

We first generate preliminary derived forecasts on an individual methodology basis. Each preliminary derived forecast can be viewed as a methodology in its own right, with the objective of generating superior earnings predictions. The model then generates EPM forecasts according to optimally weighted combinations of the preliminary derived forecasts. In essence, the EPM reviews all the analyst forecasts, employs a host of methodologies to identify superior ones, and combines them in a manner designed to optimally generate earnings estimates that are superior to an equally weighted consensus.

HISTORICAL PERFORMANCE OF EPM

In this section, we backtest the EPM to measure:

1. The accuracy of its earnings forecasts.
2. Its performance as a stock selection tool.

Accuracy of EPM Earnings Forecast

To backtest the accuracy of its predictions, we generate EPM quarterly earnings forecasts daily from the beginning of 1993 to the end of 1997, comprising a total of more than 1,200 periods. Differences in forecast accuracy between EPM and the First Call consensus estimate were compared using three alternative metrics: average absolute error, accuracy frequency, and weighted accuracy frequency.

Average absolute error is defined as the absolute value of the difference between the actual quarter earnings number and the model (EPM or consensus) forecast. We do not deflate because of interpretation problems associated with deflating (Degeorge, Patel, and Zeckhauser [1999]). Accuracy frequency measures the percentage of the time that the EPM forecasts are closer than the consensus forecasts to the actual reported earnings numbers. The weighted accuracy is similar to the accuracy frequency, but it weights each observation according to the magnitude of its accuracy superiority.

To understand the difference between the two frequency measures, consider two quarters, each with actual earnings of 50 cents. Assume that the EPM and the consensus forecasts are 48 cents and 55 cents for the first quarter, and 48 cents and 50 cents for the second quarter. The accuracy frequency for EPM is 50% because it is closer to the actual in one of the two quarters. The weighted accuracy frequency for EPM is 60%, because, when it is more accurate than the consensus, its advantage is 3 cents; when it is less accurate than the consensus its disadvantage is only 2 cents. In this example, the 60% equals 3 cents divided by the summation of 3 cents plus 2 cents. In the more general case, a model's weighted accuracy frequency is computed as the summation of the dollar advantages of the model divided by the total of the summations of the dollar advantages of both models.

Exhibit 1 displays the backtest results for the five-year period, 9301 to 9712, and the two-year period,

EXHIBIT 1

EPM Historical Performance from 9301 to 9712 and from 9601 to 9712

Summary Comparison of EPM to Consensus

	Average Absolute Error	Accuracy Frequency	Weighted Accuracy Frequency
EPM	6.79 (6.43)	63% (63%)	71% (71%)
Consensus	7.48 (7.07)	37% (37%)	29% (29%)

The five-year results are based on more than 3.48 million observations. Accuracy frequency compares EPM to consensus when errors differ. Two-year (9601 to 9712) results are shown in parentheses.

9601 to 9712. For the five-year period, the EPM average absolute error is 6.79, which is 9.2% smaller than the consensus average absolute error of 7.48. For the two-year period, the EPM average absolute error is 6.43, which is 9.1% smaller than the consensus average absolute error of 7.07. The average absolute errors are computed including all observations.

The numbers in the tables relating to average absolute errors include ties. Excluding ties, the EPM advantage for the five-year period is 13.8%, reflecting average absolute errors of 7.75 and 8.99 for the EPM and consensus forecasts, respectively. The two-year results show a 13.9% advantage for the EPM, but this time the average absolute errors are lower for both the EPM (7.30) and consensus (8.48). To be conservative, we show all subsequent average absolute error results including ties.

The EPM accuracy frequency and the EPM weighted accuracy frequency results are the same for both the five-year and two-year periods, namely, 63% and 71%, respectively. Ties are excluded when evaluating accuracy frequency and weighted accuracy frequency because including ties makes it difficult to interpret the results.

The remaining exhibits are based on data for the five-year period. Exhibit 2 shows that the mean absolute error of EPM is smaller than that of the consensus for each of the over 1,200 periods that EPM is compared with the consensus. Exhibit 3 reveals that EPM's advantage in every one of the more than 1,200 trials also pertains to the accuracy and weighted accuracy frequency metrics, with the weighted accuracy frequency almost always greater than the corresponding accuracy frequency.

Exhibits 4, 5, and 6 measure the performance of EPM and consensus forecasts according to the eleven sectors of the S&P 500. Exhibit 4 displays the average absolute errors of the EPM and consensus forecasts for each sector. The transportation sector has the largest errors for both EPM and the consensus, while communication services has the smallest errors for both EPM and the consensus. The EPM average absolute errors are smaller than those of the consensus for all eleven sectors.

Exhibit 5 shows the ratios of the average absolute error of the consensus to that of the EPM. The consumer staples sector has the highest ratio and utilities has the smallest, indicating that EPM does best relative to the consensus for consumer staples and worst for utilities. The very limited advantage of the EPM in the case

EXHIBIT 2

EPM and Consensus *Mean Absolute Error* from 9301 to 9712

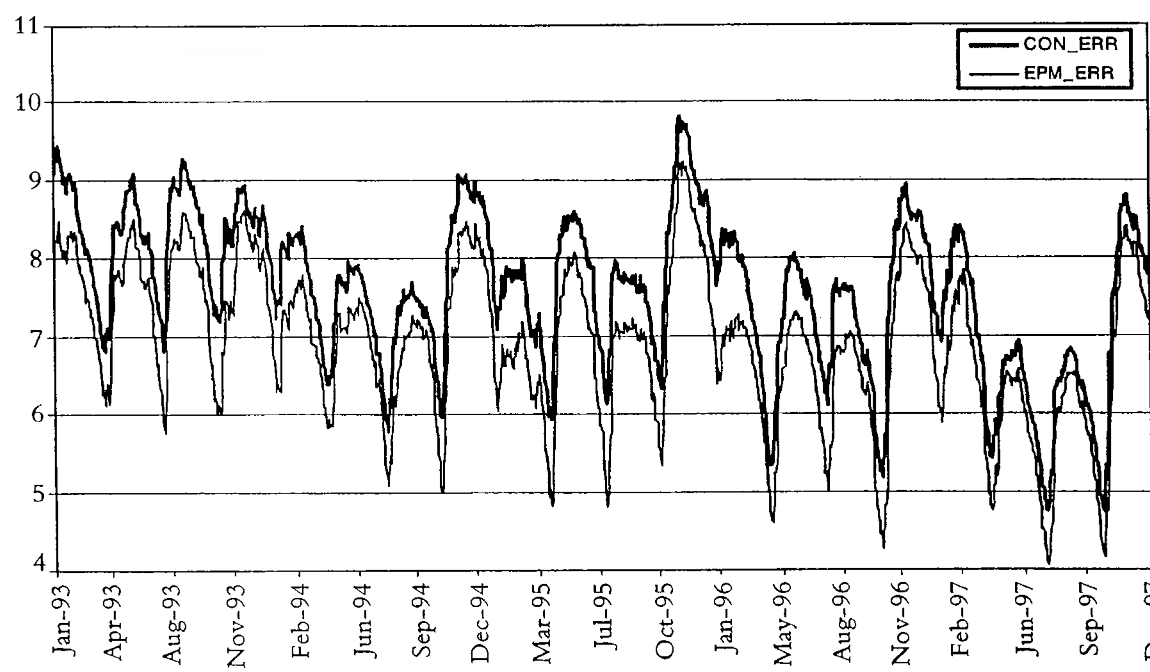


EXHIBIT 3
EPM *Accuracy Frequency* and *Weighted Accuracy Frequency* from 9301 to 9712

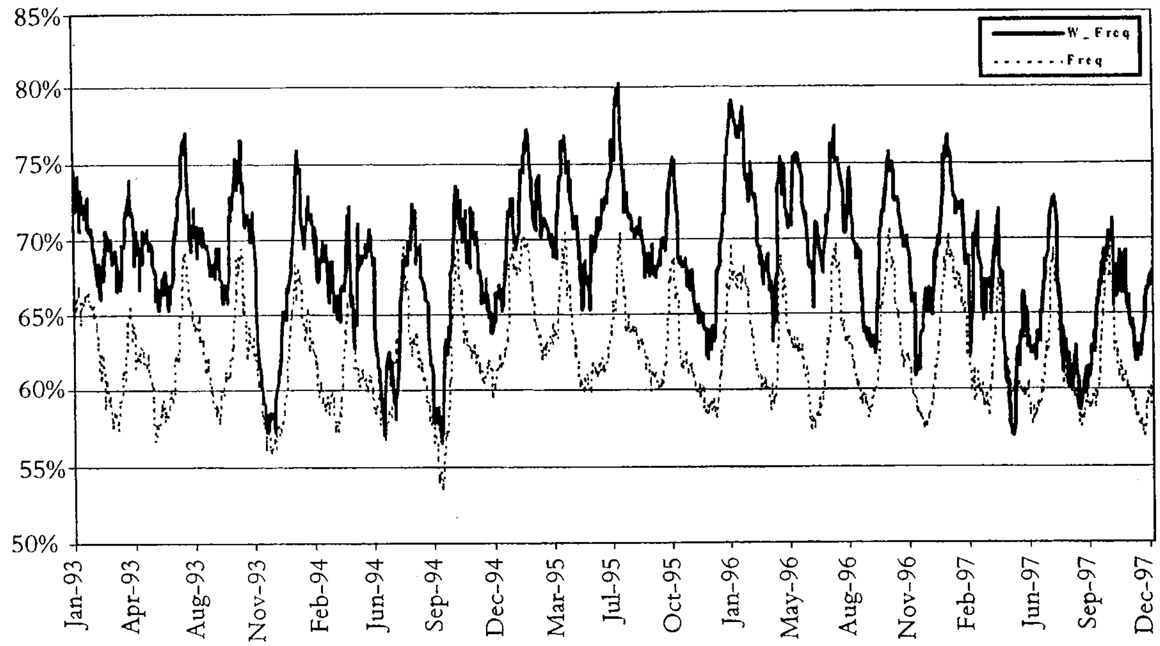
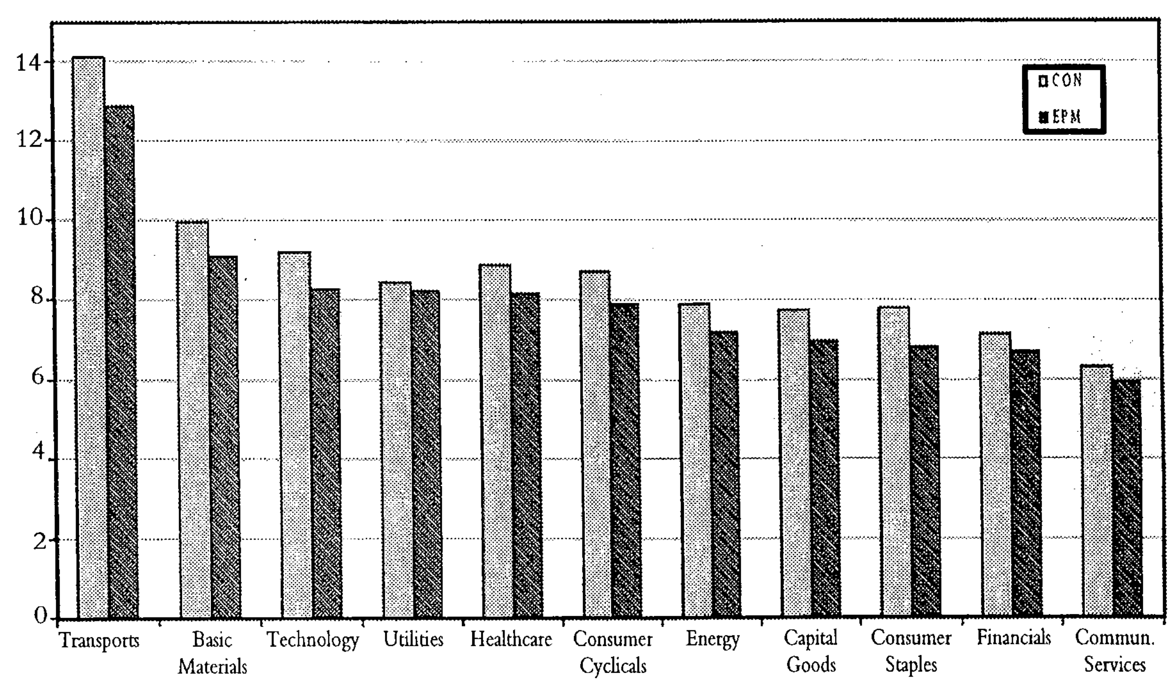


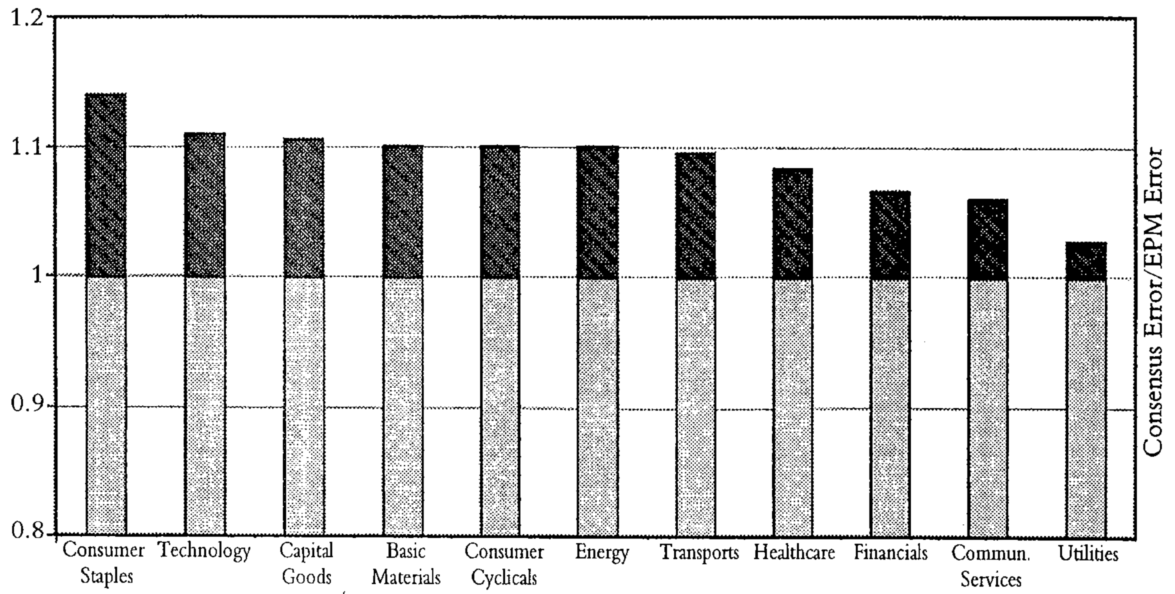
EXHIBIT 4
EPM and Consensus *Average Absolute Errors* by Sector from 9301 to 9712



Sector constituencies as of May 1998.

EXHIBIT 5

EPM *Average Absolute Error* Percentage Difference Relative to Consensus by Sector from 9301 to 9712



of utilities is due to the relatively low frequency of revisions for this sector.

Exhibit 6 reveals that the technology and consumer cyclical sectors have the highest EPM accuracy frequencies of the eleven sectors, respectively, while the consumer staples and technology sectors demonstrate

EXHIBIT 6

EPM *Accuracy Frequency* and *Weighted Accuracy Frequency* by Sector from 9301 to 9712

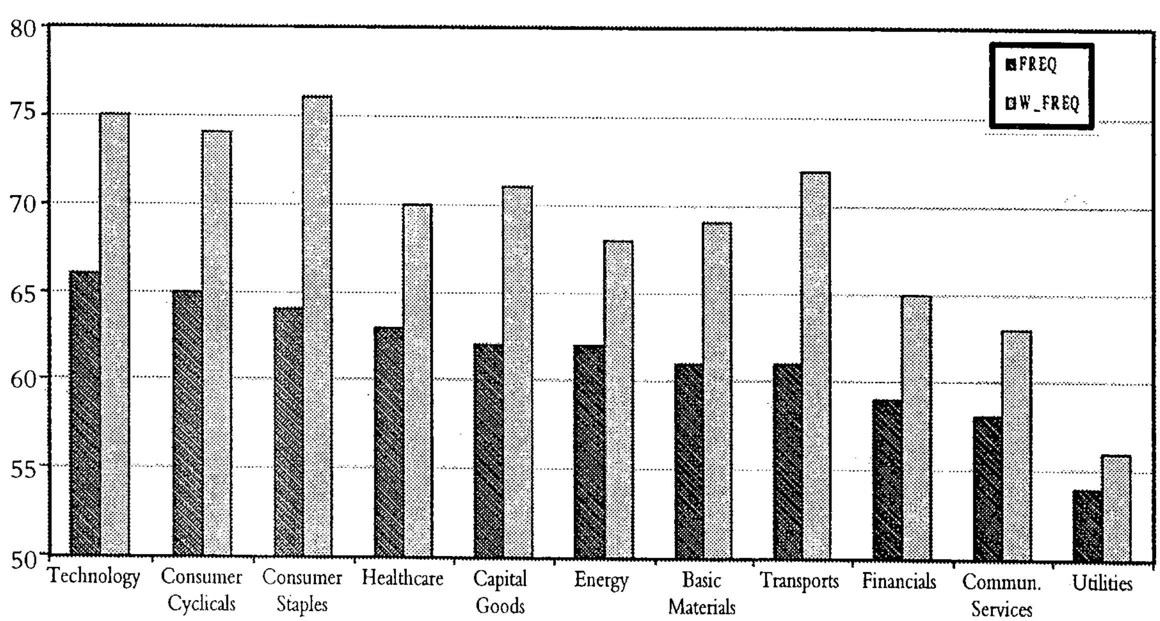


EXHIBIT 7

EPM Performance for Firms with Preannouncements *Excluding* and *Including* Preannouncement Data

Model	Average Absolute Error — Model	Average Absolute Error — Consensus	Accuracy Frequency	Weighted Accuracy Frequency
EPM *Excluding* Preannouncement Data	4.93	6.39	78%	89%
Preannouncement Estimate Only	5.50	6.39	60%	68%
EPM *Including* Preannouncement Data	4.82	6.39	77%	88%

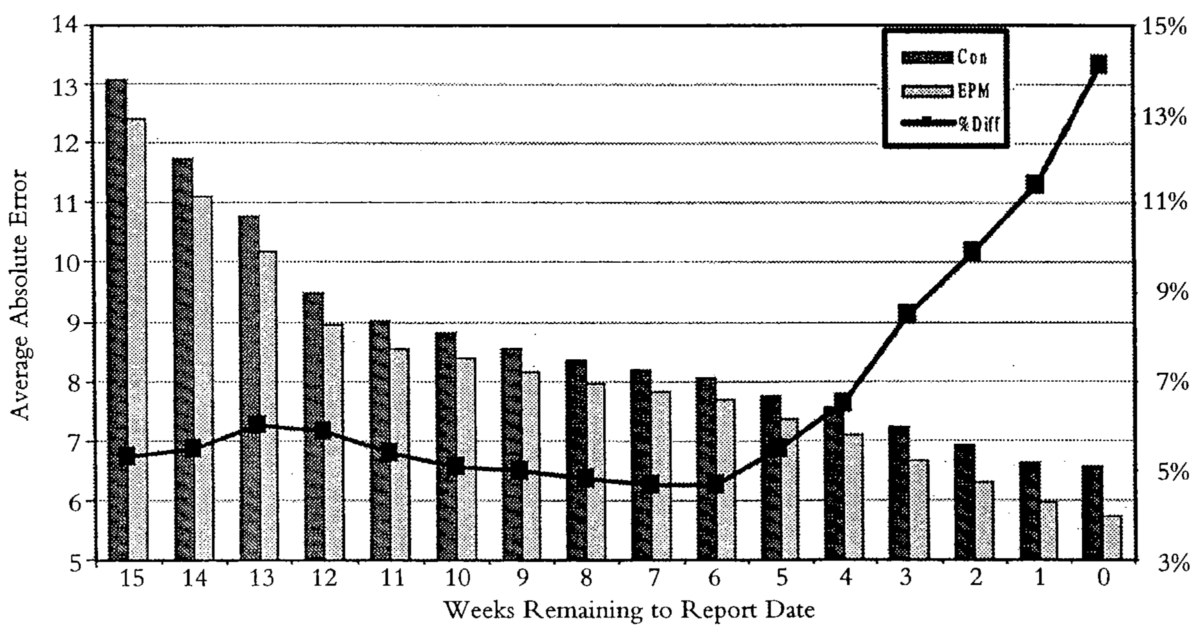
the highest weighted accuracy frequencies of the group, respectively. It is also evident that the EPM weighted accuracy frequency exceeds the EPM accuracy frequency for all eleven sectors.

Exhibit 7 examines a subset of the universe consisting of firm-quarters in which preannouncements occurred. The EPM was first tested without using preannouncement data. In other words, EPM ignored the preannouncements and generated earnings forecasts using its usual procedures of identifying superior analysts, combining them, and then classifying and scoring those forecasts. The average absolute error of EPM is 4.93 compared with the consensus average absolute error of 6.39, a 22.8% advantage for EPM. It is striking that, even without incorporating the preannouncement data into the model, EPM generated more accurate forecasts than the preannouncements themselves, which show an average absolute error of 5.50. This provides strong evidence of EPM's ability to evaluate the quality of the individual forecasts and assign weights accordingly. When EPM included the preannouncements, the average absolute error dropped to 4.82.

Exhibit 8 compares the EPM with the consensus based on time remaining to report date. The exhibit reflects the well-known result that the more time remaining to the report date, the larger the average absolute error (Crichfield, Dyckman, and Lakonishok [1978]). However, the EPM is more accurate than the consensus for every horizon. Its comparative advantage is fairly constant until about six weeks prior to the report date. Perhaps more concrete information relevant to the

EXHIBIT 8

EPM and Consensus Historical Performance Based on Time Remaining to Report Date



upcoming earnings announcements becomes available to firms. Those analysts who choose to revise their forecasts during this period, which is not a period of heightened activity, may be doing so based on more solid information. These more informed and, hence, more accurate forecasts are reflected in EPM's advantage relative to the consensus improving dramatically as the quarterly report approaches.

Stock Selection Performance

To backtest the EPM as a stock selection tool, we compute differences of the EPM and consensus forecasts for stocks in the DAIS Top 2000 universe monthly from 9301 to 9809. The stocks are sorted from most positive (EPM forecast higher than the consensus) to most negative (EPM forecast lower than the consensus) and assigned to five quintiles. Quintile 1 contains the most attractive stocks and quintile 5 has the least attractive ones. Equally weighted portfolios are constructed on the last day of each month. We assume trading occurs at closing prices and ignore transaction costs.

Exhibit 9 displays the monthly Q1 minus Q5 returns, which are seen to be positive in fifty-eight of the sixty-nine months (that is, 84% of the time) represented in this study. Exhibit 10 displays the total annualized returns for each of the five quintiles for each calendar year based on a monthly holding period. The exhibit also shows the annualized Q1 minus annualized Q5 returns. The returns for quintiles 1 to 5 are monotonic: 21.26%, 15.67%, 14.42%, 9.81%, and 5.69%. The Q1 minus Q5 returns are positive in all six years, ranging from a low of 10.34% in 1994 to a high of 19.61% in 1995. Thus, the total annualized Q1 minus annualized Q5 differential is 15.57%, providing strong evidence of the efficacy of the EPM model for stock selection.

CONCLUSIONS

The EPM generates earnings forecasts using a variety of proven earnings forecasting principles. At the core of this robust model is a series of distinct methodologies, each generating preliminary quarterly earnings forecasts. The EPM algorithm revises individual forecasts and makes use of management preannouncements. It identifies, aggregates, and scores these preliminary forecasts by taking into account factors such as timeliness, track record, clustering, and estimate dispersion. The model algorithm operates at the firm-specific level to generate final EPM forecasts according to optimally weighted combinations of the preliminary forecasts. The power of the model lies in its ability both to process vast

EXHIBIT 9

EPM Monthly Q1 Minus Q5 Returns — Top 2000 Universe from 9301 to 9809

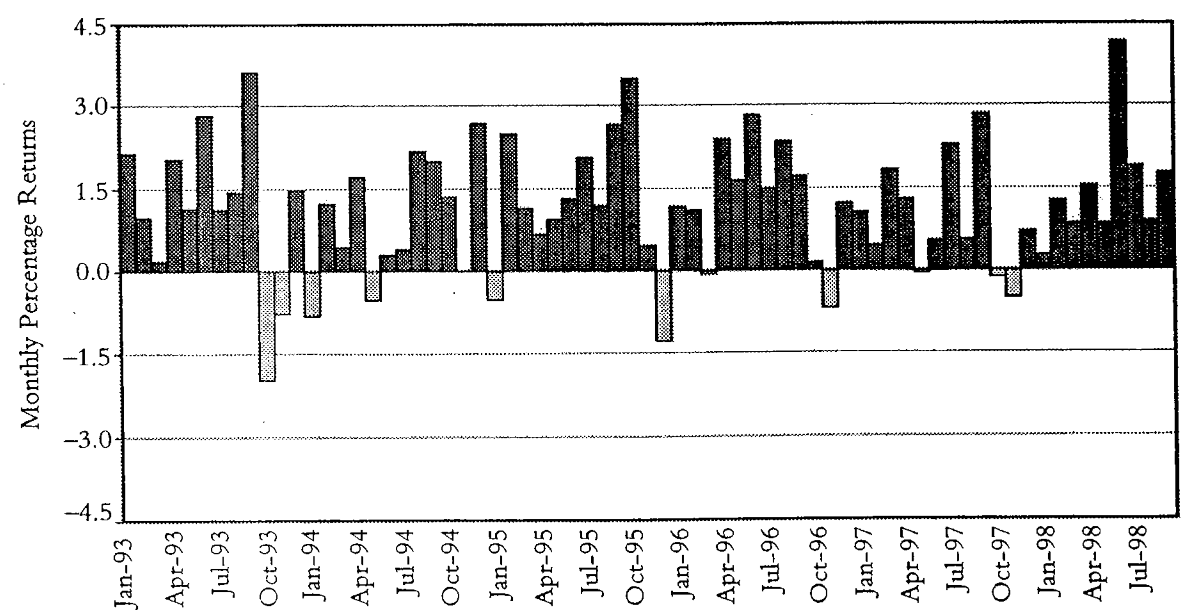
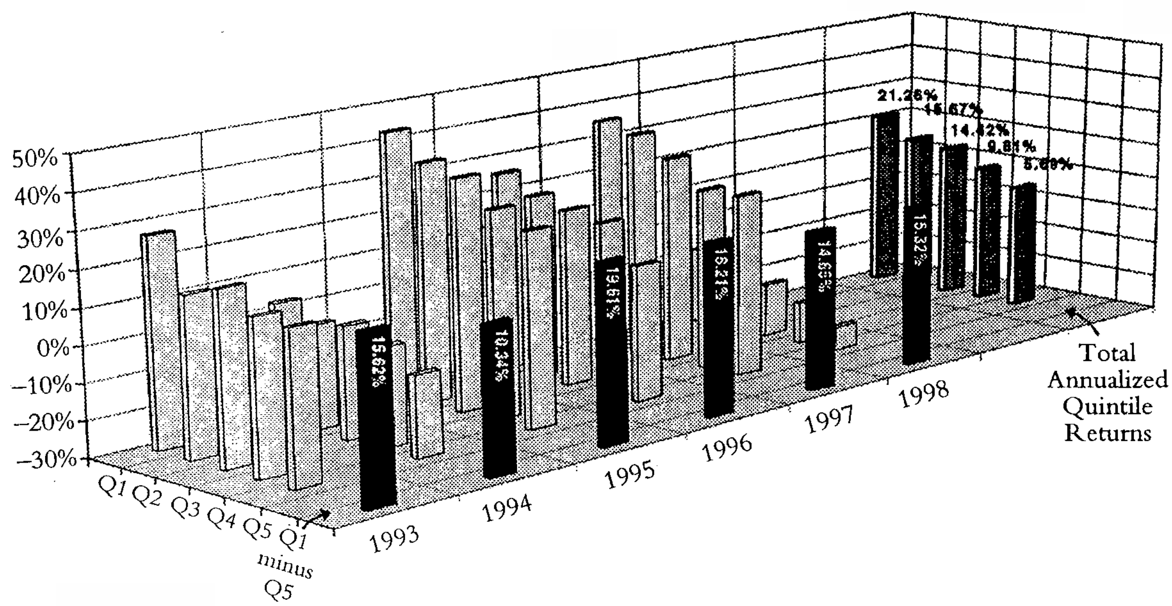


EXHIBIT 10

EPM Annualized Quintile Returns — Top 2000 Universe from 9301 to 9809



amounts of data and to ascertain those firm-specific components that best generate superior earnings predictions for each company at each point in time. The EPM attempts to model anticipated analyst behavior. It identifies and emphasizes the superior and timely analyst forecasts. Its approach is intuitive and logical.

The model forecasts are shown to be more accurate than the consensus in each of more than 1,200 (non-independent) backtests using three alternative metrics: mean absolute error, frequency accuracy, and weighted accuracy frequency. Backtests for the period 9301 to 9809 show that the EPM is very effective for stock selection purposes. It achieves purely monotonic quintile returns, demonstrating positive Q1 minus Q5 returns for each year of the study, with a total annualized Q1 minus annualized Q5 differential of 15.57%.

ENDNOTE

The authors would like to express their appreciation to Marc Miller for the construction and maintenance of the earnings model data base. They gratefully acknowledge Cristina Romero for very valuable assistance and input. They also thank Bob Butman and George Esser for helpful comments, and Stanley Levine for assistance with the First Call data.

REFERENCES

Bhagat, Sandip A. "Impact of Earnings Preannouncements." Presentation at The Third Annual Corporate Earnings Analysis Seminar, New York, 1997.

Brown, Lawrence D. "Forecast Selection When All Forecasts Are Not Equally Recent." *International Journal of Forecasting*, Vol. 7, No. 3 (November 1991), pp. 349-356.

Butman, Robert E., George J. Esser, and Martin M. Herzberg. "Equity Valuation Analysis." The DAIS Group, Monthly Research Publication, May 1998.

Crichfield, T., T. Dyckman, and J. Lakonishok. "An Evaluation of Security Analysts' Forecasts." *The Accounting Review*, Vol. 53, No. 3 (July 1978), pp. 651-668.

Degeorge, Francois, Jayendu Patel, and Richard Zeckhauser. "Earnings Manipulation to Exceed Thresholds." *Journal of Business*, Vol. 72, No. 1 (January 1999), pp. 1-33.

Dorfman, John R. "All Star Analysts 1997 Survey." *The Wall Street Journal*, June 19, 1997, pp. R1-R5.

Herzberg, Martin M. "Implementing EBO/EVA Analysis in

Stock Selection." *Journal of Investing*, Vol. 7, No. 1 (Spring 1998), pp. 45-53.

——. "Enhancing Earnings Predictability Using Individual Analyst Forecasts." Presentation at The DAIS Group's 16th Annual Equity Investment Technology Conference: A Forum for Practitioners, La Quinta, California, 1998a.

Mozes, Haim, and Patricia Williams. "Evidence of Mimicking in Analyst Forecasting Behavior." Working paper, Fordham University, 1998.

O'Brien, Patricia C. "Analysts' Forecasts as Earnings Expectations." *Journal of Accounting and Economics*, Vol. 10, No. 1 (January 1988), pp. 53-83.

Sinha, Praveen, Lawrence Brown, and Somnath Das. "A Re-Examination of Financial Analysts' Differential Earnings Forecast Accuracy." *Contemporary Accounting Research*, Vol. 14, No. 1 (Spring 1997), pp. 1-42.

Soffer, Leonard, Ramu Thiagarajan, and Beverly Walther. "Earnings Preannouncements." Working paper, Northwestern University, 1998.

Stickel, Scott. "Predicting Individual Analyst Earnings Forecasts." *Journal of Accounting Research*, Vol. 28, No. 2 (Autumn 1990), pp. 409-417.

——. "Reputation and Performance Among Security Analysts." *Journal of Finance*, Vol. 47, No. 5 (December 1992), pp. 1811-1836.