The Decision Tree Approach to Stock Selection

An evolving tree model performs the best.

Eric H. Sorensen, Keith L. Miller, and Chee K. Ooi

ERIC H. SORENSEN is a managing director and head of Global Quantitative Research at Putnam Investments in Boston (MA 02109).

KEITH L. MILLER is a managing director and co-head of Global Quantitative Research at Salomon Smith Barney in New York (NY 10048).

CHEE K. OOI is a vice president in Global Quantitative Research at Salomon Smith Barney in New York (NY 10048).

One of the most common approaches to quantitative investment management is to reduce an investible universe to a manageable set of stocks that conform to a group of desired characteristics. Investment managers often use multiple screening techniques to better focus their research. Although most managers stop short of a pure quantitative process invoking optimization and higher mathematics, many use some form of quantitative screening. These screens constrain on criteria related to valuation, fundamental earnings performance, liquidity, momentum, and investment style. Many academic studies focus on the performance of quantitative criteria in predicting stock returns.[1]

Screening is helpful. It is, however, by no means a complete or strictly scientific process. For example, some stocks may be excluded from consideration on one criterion while meeting many other criteria. Alternatively, multivariate approaches may weight factors and arrive at a summary ranking for each stock. In this instance, again, some stocks may be arbitrarily included (or excluded) because of an extreme reading on one criterion, with other factors receiving too little (or too much) weight. In truth, multivariate ranking systems tend to be *ad hoc.*

In this report, we apply the CART methodology to an old problem—stock-picking. We develop a model for screening stocks. The model departs from tradition by determining the proper hierarchy and interaction of screening factors. Does a value criterion take priority over a momentum factor, or vice versa? Does value ever matter for some types of stocks, and if so how do the value factors interact with other criteria? We illustrate the importance of these higher-order considerations with our CART model for selecting stocks in the technology sector.

TREES AND RECURSIVE PARTITIONING

CART stands for classification and regression tree. The statistical approach is a specific technique from a general class of recursive partitioning algorithms (RPA). As the name implies, partitioning techniques separate observations in a binary and sequential fashion. The purpose is to improve prediction.

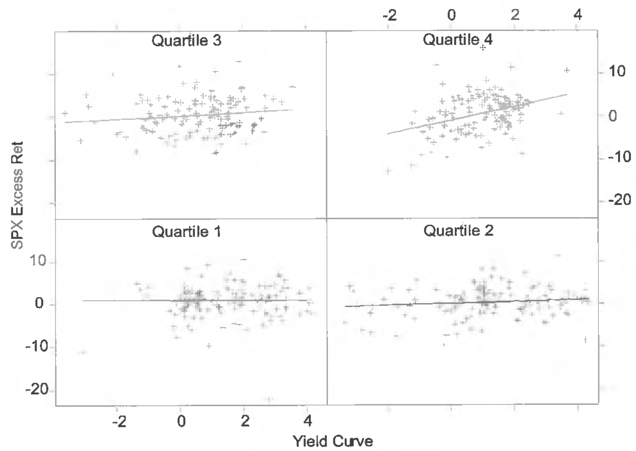
For example, assume we want to predict how fast a car can go by visual inspection of its exterior. We would collect data on a large number of cars that we already know how to rank on speed, from very fast to very slow. The data might include such variables as color, size, and tire width. On first pass, we might try partitioning the cars according to size—thinking that small cars are fast. A better model, however, may be to first partition on tire width, and subsequently partition on size. Small cars with wide tires are more likely to be classified as very fast than small cars with narrow tires. (See the appendix for details on RPA.)

Breiman et al. introduced CART in 1984, presenting a computational-statistical algorithm that predicts in the form of a decision tree. Initial applications were taken from medical diagnostics and prediction. Some have used the approach to model financial markets (see Kao and Shumaker [1999]; and Sorensen et al. [1994, 1996]). Most of this work using CART helps solve time series problems. For example, Kao and Shumaker [1999] estimate a time series to discriminate between growth stock and value stock returns.

The advantage of CART lies in determining the non-linear hierarchy of determinants that maximizes the order of a structure. The hierarchy is estimated in the form of a binary tree that produces combinations of conditions that reduce the dimensionality of the data and therefore improve the ability to predict. Effectively, the tree is a set of "if-then" rules that guide us to better decisions because the variables are allowed to take on a hierarchical prioritization and are also allowed to interact and have different influences under different conditions.

Properly deployed, CART is not a black box. The input variables and the dependent variable are comparable to the data set we would use in multiple linear

EXHIBIT 1
S&P 500 RETURNS–YIELD CURVE RELATIONSHIPS CONDITIONED ON CREDIT SPREADS



regression or discriminant analysis. The variable choices and specifications should be causal and rational, just as with traditional statistical methods. In traditional linear models, however, the final equation must treat all inputs as independent and additive, and as having the same coefficient (influence) at all times.

Linearity assumptions are clearly restrictive. For example, our S&P 500 market timing tree demonstrates that the relative value of the stock market (i.e., S&P 500 earnings yield versus Treasury bond yield) has a much more significant influence on future S&P performance when first conditioned on the state of the economy as measured by the yield curve (long Treasury bond yields minus T-bill rates) (see Sorensen et al. [1998]). A model that first asks what is the macro environment, and then considers the influence of relative value is far superior to one that assigns these two variables independent and competing influences.

Exhibit 1 illustrates a simple case of conditioning. We show the linear slope of the interaction of S&P 500 excess returns and the slope of the yield curve conditioned on different regimes of credit spreads. From the graphs, it is obvious that when credit spreads are the widest (quartile 4, or the upper right-hand panel), steeper yield curves tend to lead to better S&P 500 performance. The performance versus yield curve relationships are at best ambiguous in all the other regimes. By examining the data on a conditional basis, we have discovered a hidden relationship that would be easily overlooked by linear models.

Dividing the Data into Buckets

Using discrete categories to represent both the independent and dependent variables is a useful initial step in estimation. The output of the classification tree technique is a binary tree that assigns probabilities to the dependent variable outcome categories. The dependent variable might take on ranges such as quartiles or deciles. For example, we might want to model periods when large-cap stocks underperformed small-cap stocks by a significant degree, and vice versa. In this case, the data might be organized into periods of returns differences falling into three bins: 1) similar performance, 2) outperformance of large-cap, and 3) outperformance of small-cap.

The same discretization method is useful for specification of the independent variables. One categorical exogenous variable representing market conditions might be periods of 1) high volatility, 2) normal volatility range, and 3) low volatility.

In determining the structure of the tree, CART uses a mathematical algorithm to identify a variable and a corresponding threshold for the variable that splits the input observations into two subgroups: one comprising observations with variable values lower than the threshold and the other, observations with variable values higher than the threshold. The variable-threshold pair is selected so that it splits the observations into the two most homogeneous groups. This determines the top level of the tree, and gives us results on the critical level to split the tree, as well as the improvement in homogenizing the outcomes. (There are more details on variable selection and splitting threshold in the appendix.)

For example, if the market volatility were the most important input variable, we would find out how high (or low) a threshold volatility would have to pass to best explain historical return differentials between large-cap and small-cap stocks. Once the first cut is made at the top of the tree, there are recursive cuts that maintain the integrity of higher levels of the tree, and refine the classification outcomes.

For example, we might find that volatility partitioning is not the single most important variable but depends on some other variable. If so, the tree would split on the dominant variable, and then determine the relevant hurdle for volatility to enhance the prediction. Volatility may be subordinate to a hierarchical structure. It may be, for example, that high volatility favors small-cap stocks after a period of general market weakness, but favors large-cap stocks after a period of market strength. Moreover, the estimated classification tree may include different volatility thresholds at different points (nodes) in the tree, allowing a more complex set of interactions.

CART for Cross-Sectional Applications

The power of an RPA such as CART comes from 1) intuition on the tree structure hierarchy, 2) accounting for non-linearities in the data, 3) accounting for dependencies among variables, and 4) production of conditional probabilities for the outcomes. These advantages should surface for cross-sectional investment strategies as well as time series.

CART lends itself to cross-sectional problems—like stock selection. We pool the data over multiple periods. The model we estimate here is cross-sectional. That is, it seeks to predict categories of outcomes while pooling many observations as if all of them occurred in the same interval of time. Predicting the top speed of a car is a cross-sectional analysis. It may well be that we have sufficient observations of cars in a given time period to perform a thorough analysis. With stocks, however, we may have too *few* observations during one period to perform an adequate analysis.

We describe an example in an estimated tree structure that seeks to distinguish winners from losers in a universe of technology stocks. Using the Russell 1000 index constituents back to 1992, we combine the monthly return observations for the subset of technology stocks. At any time over these years, there may be 70-110 names within the Russell 1000 classified as "technology." By pooling the returns for each stock over monthly observations, we can create a dependent variable representing the distribution of relative performance that provides thousands of observations. The goal is to estimate a stable model that helps distinguish outperformers from underperformers month by month, based on reasonable stock and/or company attributes.

The Search for Homogeneity Within Sectors

What is the point of analyzing technology stocks as a set? It is simply the case that grouping stocks that show some homogeneity should improve our ability to find significant relationships. Factors that explain stock returns often vary across economic sectors. Analysis of style performance should suggest that the driving forces of relative return may vary greatly between different groups.

For example, univariate stock selection systems should show that earnings momentum measures are better predictors of technology relative stock performance than are value measures. And, it is reasonable that relative value plays a much more important role in predicting performance of financial stocks.

One way to group stocks is to sort them on return correlation. We can go through the formal exercise of forming clusters based on historical return similarities. Another approach is to accept a subjective classification scheme such as S&P industries or Russell sectors, and so on. In this case, we have simply accepted the Russell designations for "technology."

Other sectors are health care, consumer discretionary, consumer staples, integrated oils, other energy, materials and processing, producer durables, auto and transportation, financial services, utilities, and miscellaneous.[2]

THE INPUTS

The application we present is designed to improve the screening rules for selection of buy candidates versus sell candidates, using technology stocks as the sample. The degree of fineness of the dependent variable has considerable bearing on the model. Ranking stocks on a complete continuum from top performance to lowest performance is probably too tall an order for any quantitative approach. Our more modest purpose is to assist the portfolio manager in reducing the number of stocks for analysis. Alternatively, the purpose might be to create a subset of buy candidates to overweight in a risk-controlled index enhancement portfolio.

First, we measure the total return of each stock for each month in history. The database begins in 1992 and extends through the end of October 1999. Next, we compute the relative return of each stock in each month, by subtracting the median stock return for the month in the sample from each individual stock return. This allows us to categorize the level of outperformance (or underperformance) for each stock in each month. Obviously, one stock might have a month of superior (inferior) returns in a declining market and/or a month of inferior (superior) returns in a rising market.

The objective is quite simple—to sort the above-median performers from the below-median performers. That is, the dependent variable is binary and splits the sample in half. Each observation is unique to the specific stock, the specific month, and the distribution of the cross-sectional returns for the specific month. Therefore, half of the observations are tagged "outperform," and half are tagged "underperform."

EXHIBIT 2
INPUT VARIABLES

Variable	Definition
SALES-PRICE	Sales-to-price-ratio
CFLOW-PRICE	Cash flow-to-price ratio
EPS-PRICE	Forward-12-mo. consensus forecasted earnings-to-price ratio
ROA	% Change in ROA
EPS-MOM	12-week change in consensus forecasted earnings
PRICE MOM	Previous 1-mo. stock returns

We select a small set of independent variables. The variables in Exhibit 2 are taken from a large set of univariate factors that are the most popular among money managers. They span valuation, fundamental profitability, consensus expectations for earnings, and price momentum criteria. The specific six variables are chosen because we find they each have some relevant explanatory power as univariate predictors.

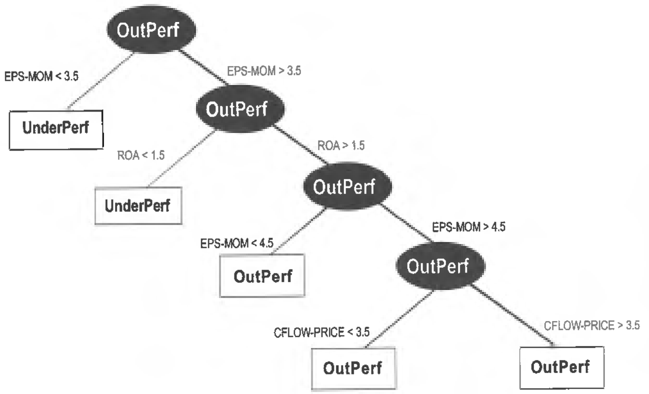
A key consideration is the specification of each factor. After calculating the various ratios and momentum factors, each variable is quintiled for each month. Placing the variables into five categories, as opposed to finer partitioning, provides a more stable estimation structure. Quintiles are more consistent with popular screening techniques. In addition, if CART estimates the model using more precise data, such as continuous variables, the structure is vulnerable to overfitting.

Overfitting is illusory because, on the one hand, it appears to explain more of history. On the other hand, a model that is overspecified lacks predictive stability. It also lacks intuition in interpretation of the tree.

TECHNOLOGY STOCK MODEL: A STATIC TREE

Our first approach in estimating a technology stock selection model is a simplistic version that assumes a high level of stability in the interactions between relative performance and input variables. In this model, we divide our data set into two groups of approximately equal size samples: 1) observations from February 1993 through December 1995, and 2) those from January 1996 through October 1999. The first group serves as the "in-sample"

EXHIBIT 3
STATIC TREE MODEL



data—observations that we use to estimate a tree model. The second group serves as the "out-of-sample" data—samples that we use to evaluate the performance of the estimated tree model in classifying cases that it has never seen before. An out-of-sample performance test provides a gauge for the predictive ability of the model. As there is only one tree model that spans the whole period, we term this method a static tree approach.

Exhibit 3 is the classification tree from the model. The tree is populated down to five terminal nodes, with four levels of partitioning.

At the top of the tree, the primary variable is EPS-MOM (the 12-week change in consensus forecasted earnings). The partitioning algorithm has determined that, of all possible variables, estimate revision is the single most powerful factor in explaining winners versus losers on a historical monthly basis. Because we have quintiled the variables, there are four possible splits: 1) between quintiles 1 and 2; 2) between quintiles 2 and 3; 3) between quintiles 3 and 4; or 4) between quintiles 4 and 5. (Quintile 5 are stocks with the highest readings on the variable, and quintile 1 are stocks with the lowest readings.)

The model in Exhibit 3 first splits between quintiles 3 and 4. This means that the maximum order of the outcomes is to divide stocks into two categories: the top two quintiles on earnings estimate revision versus the bottom three quintiles. The RPA is not done at this point, however. There is further improvement of historical classification, with secondary splits on the right-hand side of the tree.

When the model branches to the right, the RPA is repeated for the subgroup of samples that end on the right. ROA and a splitting threshold of 1.5 are determined to form the variable-threshold pair that most differentiates outperformers from underperformers. If stocks are in the upper two quintiles of EPS-MOM, then there is a good likelihood that they will outperform in the next month. In addition, having conditioned on high earnings revision, the momentum in ROA has incremental predictive power. In particular, high estimate revision—if also associated with the top four quintiles of ROA momentum— will lead to even greater likelihood of outperformance.

This is an intuitively satisfying result because it means that the market discriminates between firms with analysts' optimism that have supporting fundamental stability or improvement versus those that have deteriorating ROA readings (lowest quintile). Although the model continues to split farther down the tree, note that these additional splits have no bearing on the final classification of the stock, as each of the terminal nodes corresponds to the same outcome, "outperform."

CART has confirmed our intuition, and allowed for a much richer specification than simple linear categorical screening. In addition, the model has optimally determined all the relevant partitions based on data from 1992 to 1995.

The actionable output of the tree is quite simple. Every stock has characteristics associated with one of the five nodes in each month. If the node is labeled outperform, each stock in the node takes on that prediction. The odds of correct classification vary from node to node.

How can we trust the tree for prediction accuracy? First, the tree should have a structure that fits our prior intuition. Second, we can estimate the tree over various periods and/or samples. Third, we can devise a backtest.

Exhibit 4 is an out-of-sample test of the model. It is a very simple approach. Each month we form two portfolios. One portfolio is an equal weighting of the outperform stocks (long), and the second is an equal weighting of the underperform stocks (short). The bars in Exhibit 4 are the monthly return spreads between two portfolios. The top panel shows the spreads between the long portfolio and the overall universe. The bottom panel shows the return spreads between the long portfolio and the short portfolio. So we are simply using the tree estimated with data from the first three years, and checking to see whether it can be useful in subsequent periods.

Exhibit 5 shows the wealth curves associated with the long versus short test. The long minus short port-

EXHIBIT 4
STATIC MODEL:
MONTHLY PERFORMANCE COMPARISONS

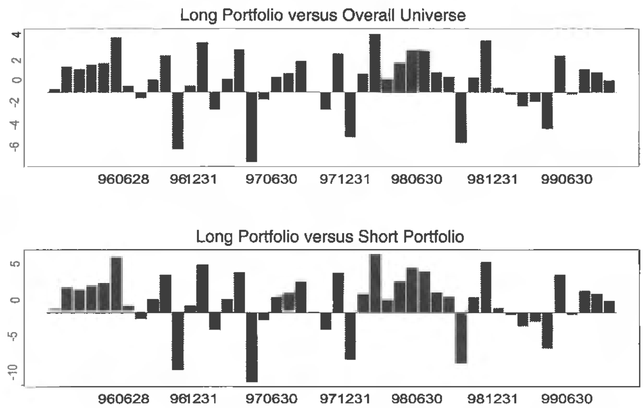


EXHIBIT 5
STATIC MODEL: PORTFOLIO WEALTH CURVES

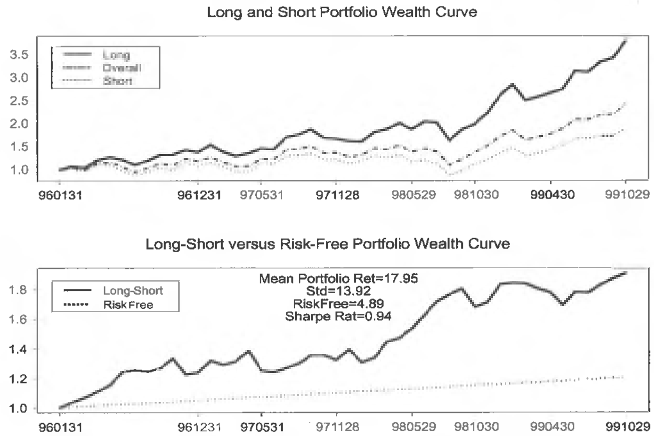


EXHIBIT 6
STANDARD T-TEST OF MONTHLY EXCESS RETURNS

	Outperform Portfolio	Underperform Portfolio
Estimated Mean Excess Returns	0.92	-0.48
t-stat	2.49	1.69
p-value	0.01	0.09
	Outperform Minus Underperform Portfolio	
Estimated Mean Difference	1.40	
t-stat	2.94	
p-value	0.00	

EXHIBIT 7
WILCOXON RANKED TEST OF MONTHLY EXCESS RETURNS

	Outperform Portfolio	Underperform Portfolio
Estimated Mean Excess Returns	0.92	-0.48
Signed-Rank Normal Statistic with Correction Z	3.00	-2.41
p-value	0.00	0.02
	Outperform Minus Underperform Portfolio	
Estimated Mean Difference	1.40	
Rank-Sum Normal Statistic with Correction Z	3.83	
p-value	0.00	

folio has a mean return of approximately 17.95% annually with a standard deviation of approximately 13.92%. We have not adjusted for transaction costs.

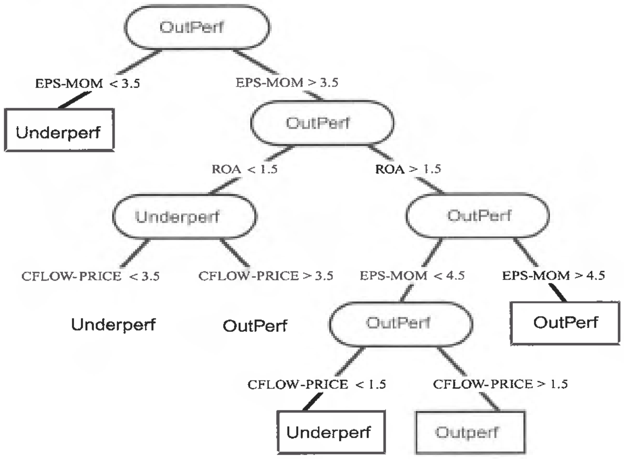
As an additional test, we also evaluate the excess returns distribution of the long and the short portfolios as constructed by the model. Exhibit 6 gives results of a standard T-test on the excess returns distribution of the portfolios. As the excess returns distribution may not be normal, the Wilcoxon ranked test in Exhibit 7 may be a better measure of statistical significance.

On a monthly basis, the stocks classified as outperform have an average excess return of 1.40% over those classified as underperform. This monthly difference is statistically significant at the 95% confidence interval.

TECHNOLOGY STOCK MODEL: AN EVOLVING TREE

The static model we have described demonstrates some stability in the tree structure. Namely, a tree estimated for the 1992–1995 period has some predictive power for the subsequent 1996-1999 period. As an alternative to this static model, we might see whether dynamic re-estimation each month can create better out-of-sample predictions.

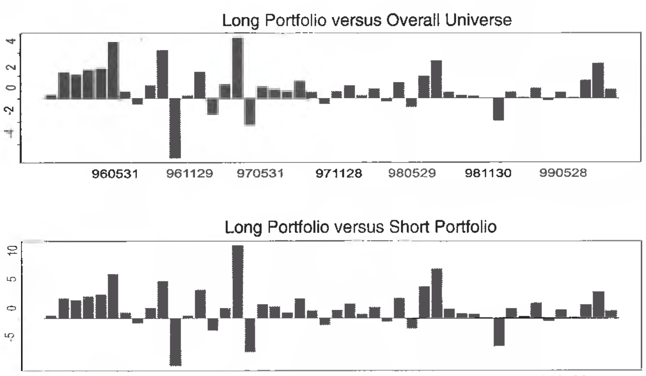
EXHIBIT 8
THE LATEST OF THE EVOLVING TREE (AS OF OCTOBER 1999)



Definition of Terms:

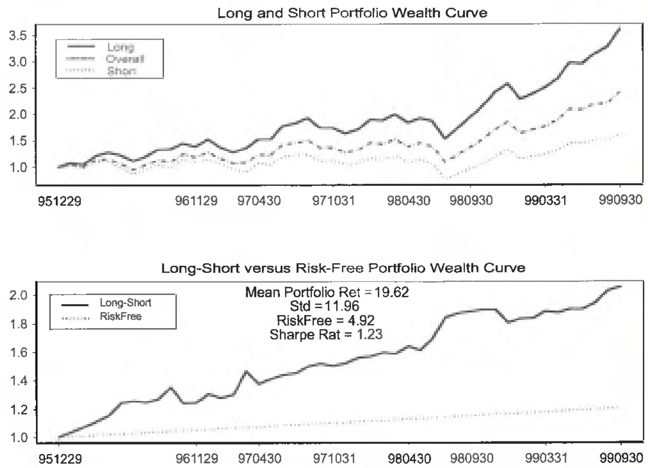
EPS-MOM	12-Week Change in Consensus Forecasted Earnings
ROA	% Change in ROA
CFLOW-PRICE	Cash Flow-to-Price Ratio

EXHIBIT 9
EVOLVING MODEL: MONTHLY PERFORMANCE COMPARISONS



Our second approach re-estimates a tree every month using all available data from the beginning to the current month. Starting at the end of December 1995, the model uses data from February 1993 through December 1995 to estimate a tree model, and the resulting model is then used to classify stocks for January 1996. For each ensuing month, new samples from the latest month are added to the model estimation process, and new forecasts are made based on the model.

EXHIBIT 10
EVOLVING MODEL: PORTFOLIO WEALTH CURVES



Clearly, a dynamic approach could result in a different tree model for every single month. But, from our experience, the tree structures tend to be relatively stable over the short term, with only occasional subtle changes from month to month. Small short-term changes, however, could add up to a more significant structural shift over the longer term. For example, the tree model for June 1999 could be significantly different from that for June 1996. Hence, we call the technique the "evolving tree approach."

There are various advantages to the evolving tree approach. The tree estimation process requires a lot of data in order to be statistically significant. In an evolving tree approach, progressively more data are used for the estimation. Logically, the evolving nature of the model makes better sense. It allows for a gradual shift in market dynamics by adapting and incorporating changes into the tree structure over time.

Exhibit 8 shows the latest tree produced by the model as of October 1999. Comparing the latest tree with that of the tree as of January 1996 (shown in Exhibit 3), one notices that, overall, there has not been a significant change in the tree structure. The model still favors stocks with good earnings momentum, as long as they have reasonable improvement in profitability (change in ROA). It is interesting to note, however, that the valuation measure (cash flow-to-price) has become more of a deciding factor.

EXHIBIT 11
STANDARD T-TEST OF PORTFOLIOS' MONTHLY EXCESS RETURNS

	Outperform Portfolio	Underperform Portfolio
Estimated Mean Excess Returns	0.75	-0.72
t-stat	2.39	2.21
p-value	0.02	0.03
	Outperform Minus Underperform Portfolio	
Estimated Mean Difference	1.47	
t-stat	3.25	
p-value	0.00	

EXHIBIT 12
WILCOXON RANKED TEST OF PORTFOLIOS' MONTHLY EXCESS RETURNS

	Outperform Portfolio	Underperform Portfolio
Estimated Mean Excess Returns	0.75	-0.72
Signed-Rank Normal Statistic with Correction Z	2.79	3.03
p-value	0.00	0.00
	Outperform Minus Underperform Portfolio	
Estimated Mean Difference	1.47	
Rank-Sum Normal Statistic with Correction Z	4.10	
p-value	0.00	

The model has turned negative on high-earnings momentum stocks that sport expensive valuation (lowest quintile in cash flow-to-price ratio). In other words, the model has adopted a more value-driven strategy, as opposed to the predominantly earnings momentum-driven strategies that have dominated in the previous few years.

Exhibit 9 shows a backtest of the evolving tree model. Each month (starting with January 1996), we use the estimated tree from CART to predict the outperformers and underperformers of the universe for the next month. Equal-weighted long and short portfolios are constructed from the outperform and underperform classes of stocks, respectively, and the actual performance of the two portfolios is shown in Exhibit 9. The top panel shows the return spreads between the long portfolio and the overall technology universe, and the bottom panel shows the spreads between the long portfolio and the short portfolio.

In the top panel of Exhibit 10, we track the wealth curves of the long, short, and overall portfolio over the out-of-sample period. In the bottom panel, we evaluate the performance of a long-short strategy based on the model. Over the out-of-sample backtest period, the mean annual returns of the long-short portfolio are 19.62%, with a standard deviation of 11.96%.

EXHIBIT 13
PERFORMANCE COMPARISONS

	Outperform Portfolio Excess Returns		Underperform Portfolio Excess Returns		Mean Difference of Excess Returns		Long-Short Strategy	
Model	**Mean**	**T-Stat**	**Mean**	**T-Stat**	**Mean**	**T-Stat**	**Ann. Ret**	**Sharpe**
Static Tree Model	0.91	2.49	-0.48	1.69	1.40	2.94	17.95	0.94
Evolving Tree Model	0.75	2.39	-0.72	2.21	1.47	3.25	19.22	1.15
EPS-MOM	0.62	2.01	-0.63	1.89	1.25	2.76	14.62	0.60
ROA	0.12	0.38	-0.12	0.37	0.24	0.53	1.85	0.27
CFLOW-PRICE	0.18	0.62	-0.19	0.55	0.37	0.82	6.00	0.07
Mean of EPS-MOM, ROA, and CFLOW-PRICE	0.28	0.93	-0.32	0.92	0.60	1.31	7.65	0.23

EXHIBIT 14
CONSISTENT OUTPERFORMERS ACCORDING TO MODEL FORECASTS

		Quintile Rank (1 = lowest, 5 = highest)			
Ticker	**Company**	**EPS-MOM**	**ROA**	**CFLOW-PRICE**	**3-M Returns**
ADPT	Adaptec Inc	4	5	4	15.79%
ADBE	Adobe Systems Inc	4	4	4	63.20%
AMCC	Applied Micro Circuit	4	4	2	65.51%
EMC	EMC Corp	4	3	3	20.84%
INTU	Intuit Inc	5	5	5	6.88%
NSOL	Network Solutions	4	2	2	90.41%
TXN	Texas Instruments	4	5	3	24.74%
COMS	3Com	4	3	5	20.22%
UIS	Unisys Corp	4	3	2	-44.72%
AVX	AVX Corp	5	2	1	27.68%
BEAS	BEA Systems	4	5	2	86.31%
CNXT	Conexant Systems	5	5	3	48.49%
MSFT	Microsoft	5	3	4	7.85%
PMCS	PMC-Sierra Inc	5	4	1	20.45%
QLGC	Qlogic Corp	5	5	1	24.80%
QCOM	Qualcomm Inc	5	4	2	42.69%
SUNW	Sun Microsystems Inc	5	3	3	55.91%
VRSN	Verisign Inc	5	5	1	66.59%
XLNX	Xilinx Inc	4	4	3	26.05%
LLTC	Linear Technology	5	2	4	14.13%
SFE	Safeguard Scientific	5	1	5	31.44%
Mean Return					**34.06%**

We perform the same statistical tests on the monthly excess returns distributions (Exhibits 11 and 12). The means of excess returns for the long portfolio and the short portfolio are found to be statistically significant using both methods of statistical testing.

The results of the dynamic model show a marginal improvement over the static model. The monthly return differential is 1.47%. In addition, the underperform group is now clearly inferior to the overall universe.

In this particular application, the monthly portfolio turnover averages 22% (or 264% annually). The turnover of any strategy is influenced by many factors, including the rebalance frequency (e.g., monthly versus quarterly) and the stock selection criteria employed by the decision tree. For example, if the tree is dominated by valuation criteria (which tend to be slower moving), presumably the turnover will be lower than if the process is more dependent on momentum factors (such as estimate revision trend).

The objective of the analysis is to improve on traditional methods of stock screening (that is, uncover attractive opportunities) rather put forward a disciplined investment process, which would explicitly control for turnover.

PERFORMANCE EVALUATION

How does the performance of the decision tree models compare with simple stock screening and ranking systems? To answer this question, we construct several screening strategies and evaluate their performance. Exhibit 13 summarizes the results. For comparison, we construct three single-factor models based on the same variables selected by the decision tree model in Exhibit 8. Each of these models ranks stocks on a monthly basis depending on their respective stock attributes; stocks in the top half (for example, above-median EPS-MOM) constitute the outperform portfolio, and stocks in the bottom half (for example, below-median EPS-MOM) constitute the underperform portfolio. We also construct a model that uses mean of the decile rankings of all three attributes (EPS-MOM, ROA, and CFLOW-PRICE) as a composite rank for stock selection.

Exhibit 13 shows that both decision tree models produce significantly better Sharpe ratios than do the simple ranking models. Other than the outperform portfolio of the EPS-MOM screen, the excess returns of all other portfolios as constructed using the simple ranking approaches prove to be statistically insignificant. Not surprisingly, of all the models, the evolving tree model performs the best, based on Sharpe ratio and t-statistic comparisons.

EXHIBIT 15
CONSISTENT UNDERPERFORMERS ACCORDING TO MODEL FORECASTS

		Quintile Rank (1 = lowest, 5 = highest)			
Ticker	**Company**	**EPS-MOM**	**ROA**	**CFLOW-PRICE**	**3-M Returns**
ADPT	Adaptec Inc	4	5	4	15.79%
CDN	Cadence Design Sys	1	1	1	43.91%
GIC	General Instrument	3	4	4	18.60%
SEG	Seagate Technology	1	5	5	9.54%
CTXS	Citrix Systems Inc	3	4	1	23.19%
ADCT	ADC Telecommunication	3	1	3	7.21%
CPQ	Compaq Computer	1	5	4	-20.46%
ITWO	I2 Technologies Inc	1	4	1	156.51%
LHSG	LHS Group Inc	3	4	2	-20.84%
SPOT	Panamsat Corp	3	2	1	6.90%
PSFT	PeopleSoft Inc	1	1	1	10.16%
ROK	Rockwell Intl	2	5	4	-17.22%
SANM	Sanmina Corp	2	1	3	37.90%
SE	Sterling Commerce	1	3	2	-10.60%
SBL	Symbol Technologies	2	2	4	1.16%
SNPS	Synopsys Inc	2	3	3	3.53%
VTSS	Vitesse Semiconductor	3	2	3	43.80%
Mean Return					**18.33**

ILLUSTRATION

The bottom line of the tree-based screen is a list of buy or sell candidates. To illustrate this, we present two short lists of stocks. In Exhibits 14 and 15, we list a few of the stocks that are consistently ranked as outperform and underperform, respectively, in the three months of Q3:1999 using the evolving tree model.

ENDNOTES

[1]See Chan, Jegadeesh, and Lakonishok [1995, 1996]; Fama and French [1992, 1996]; Fama and MacBeth [1973]; Hawawini and Keim [1995]; and Lakonishok, Shleifer, and Vishny [1994].

[2]One can easily argue that sector definitions are arbitrary and/or too broad. Nevertheless, it is a model design decision that the researcher must make. One extension for refinement may be to use cluster analyses in conjunction with industry designations. In addition, the specification of the input variables can help to reduce heterogeneity in the structure. We have found that, within broad sectors, valuation measures that are relative to industry groups are sometimes more predictive than nominal valuation measures. For example, we may define value as P/E or alternatively as the ratio of the stock's P/E to the P/E of the respective industry group.

REFERENCES

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees.* New York: Chapman & Hall, 1984.

Chan, Louis K.C., Narasimhan Jegadeesh, and Josef Lakonishok. "Momentum Strategies." *The Journal of Finance*, 51, 5 (March 1996), pp. 1681-1713.

—— "Size and Book-to-Market Factors in Earnings and Returns." *The Journal of Finance,* 50, 1 (March 1995), pp. 131-155.

Fama, Eugene E., and Kenneth R. French. "The Cross-Section of Expected Stock Returns." *The Journal of Finance,* 47, 2 (June 1992), pp. 427-465.

——."Multifactor Explanations of Asset Pricing Anomalies." *The Journal of Finance,* 51, 1 (March 1996), pp. 55-84.

Fama, Eugene E., and James D. MacBeth. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy*, 81, 3 (May/June 1973), pp. 607-636.

Hawawini, Gabriel, and Donald B. Keim. "On the Predictability of Common Stock Returns: WorldWide Evidence." In R. Jarrow, V. Maksimovic, and W.T. Ziemba, eds., *Handbooks in OR & MS*, Vol. 9. Amsterdam: Elsevier, 1995, pp. 497-544.

Kao, Duen-Li, and Robert D. Shumaker. "Equity Style Timing." *Financial Analysts Journal*, January-February 1999.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny. "Contrarian Investment, Extrapolation and Risk." *The Journal of Finance*, 49, 5 (December 1994), pp. 1541-1578.

Sorensen, Eric H., Joseph J. Mezrich, and Keith L. Miller. "Asset Allocation—The Cart Before the Bourse." Salomon Brothers, June 1996.

——. "Decoupling of U.S. Stocks and Bonds." Salomon Brothers, September 1994.

——. "A New Technique for Tactical Asset Allocation." In F.J. Fabozzi, ed., *Active Equity Portfolio Management*, Ch. 12. New Hope, PA: FJF Associates, 1998.