Where Have All The Models Gone?

Michael Even, Executive Vice President
Michael McElroy, Vice President
Independence Investment Associates, Inc.

Summary Of Research.

With a lack of new valuation modeling ideas available, quantitative investors are forced to try and squeeze more information from their existing models. This study outlines the issues faced in a research project having the goal of reconfiguring our existing valuation system to reflect stock-specific information.

New valuation ideas are few and far between these days. Broadly speaking, there are three types of valuation models -- those that identify "value," such as dividend discount models (DDMs) or P/E models, those that identify "growth" (momentum and "technical" models,) and models that use macroeconomic factors to pick stocks based on stock sensitivity to these factors. Model vendors purport to have models with a new "twist", but testing shows that these new ideas can be classified into one of the three above groups. So what is a quantitative manager to do in order to avoid models becoming "stale?"

Different quantitative managers have different methods of putting models together. One common method is to use a multi-factor valuation model in which the weights chosen for each constituent model are chosen to attain the highest alpha level for a given level of composite model volatility. Once this set of weights is determined, it is then applied to models across a universe of stocks in order to compute a composite ranking. These weights are recomputed periodically to reflect both the historical efficacy of each of the models as well as the uniqueness of the models. Conceptually, this structure can be thought of as a "portfolio" of models whose holdings of each model are chosen to achieve the highest utility possible.

An alternative method would be to find the best set of models for each stock (or cluster of stocks) and use those weights to compute the composite model. This idea poses several questions -- how stable are the clusters? What models to choose as the starting universe? And, how does one combine cluster information into a composite model? Using these questions as a guide, we embarked on comparing results from our current experience with a new weighting scheme for valuation models.

Introduction to Our Current Process.

Valuation models are really simple things -- each represents a question asked of each stock on a list, with the responses ranking the stocks from most attractive to least attractive. An example of one question is "what is your P/E ratio?" The lower it is, the more attractive. However, a low P/E can become lower. Another question (model) could be earnings momentum -- "does this stock show improving quarterly percentage changes in its earnings (which could be adjusted for seasonality, economic conditions, etc.)?" What can be scary about momentum is that it often gives its strongest buy signals near the tops of ranges.

Philosophically, investors want to own stocks which are "cheap" and show strong earnings momentum, or growth. At Independence, we have eleven valuation models to identify these characteristics, and weight these models in the composite model according to their predictive ability and their correlations with other models. These valuation models must be theoretically sound (such as a DDM), logical (for instance, low P/E), and common-sensical (earnings surprise or momentum.) The models should be unique, meaning that the responses to one question should have low correlation with responses to other questions. Finally, the models should be robust, spanning the entire list of stocks, although one hypothesis of this project is that there could be "sub-sector" models.

Investment managers are always looking for unique ways to identify attractive stocks. However, new ideas are hard to come by -- most of them seem to fall into one of the above buckets. While one could conjure up some new idea (for instance, in years when the NFC wins the Super Bowl, value stocks do well) which has very low -- perhaps negative -- correlation with existing models, the new idea might not have predictive ability; more importantly, this idea might not reflect other investor's opinions at all, which is really the intent of a good valuation idea. Slide #6 shows the correlations between ranked lists generated with "value" models (identified by "V") and growth models (identified by "G.") One salient point sticks out -- value models seem highly correlated with value models, and growth models with growth models, and the cross-model correlations are the ones that seem low. Given the eleven models and their correlations shown here, what is the hope of finding some unique model that doesn't use some of this information?

Model correlation is only one piece of the puzzle -- model predictive ability is the other, and is the really critical one, since these are the buy/sell decisions being made in a portfolio. Some measures of model predictive ability are the information coefficient or "IC" (the rank correlation of an a-priori list of stocks with its a posteriori actual returns), the standard deviation of this information coefficient, or the "batting average" (the number of positive IC months divided by the total number of months in the sample.) Each of these measures represents some way to quantify the effectiveness of a model.

Our Experience Identifying New Model Ideas.

At Independence we are always keeping our eyes open for "new" valuation modeling ideas. Unfortunately, there aren't a whole lot of "new" ideas floating around in the literature or being peddled by vendors when these ideas are scrutinized under a microscope. Internally, we have investigated modified DDMs, enhancements to earnings revision models, and cash flow/balance sheet models over the past three years, and have had a tough time finding anything new. Why would this be? One could believe a so-called "Simple Market Hypothesis," or SMH, which says that only a few items matter in valuation, such as earnings, and all other types of models simply represent different manifestations of this data.

For example, we have looked at valuation models based on sales (e.g., a sales-to-price model), yet have found that these are highly linked to earnings-type models, primarily due to the fact that sales become earnings after applying a margin multiplier to them. Another "interesting" idea we heard about is a "POP" model for cellular stocks which values a cellular firm's earnings as a function of how many "POPs" (its capacity) it owns or has rights to. Even here, the purist would say that a good analyst should be able to determine the earning power of these POP assets, which when input to traditional models should yield fair value measures for these stocks just as for any other stock.

An alternative to the SMH could be the "RIH," or Real Investor Hypothesis, which says that valuation model ideas matter only after a theory is widely accepted. Are DDMs universal truths, or not? How did investors under the Buttonwood tree (the precursor to the NYSE) value stocks -- with DDMs, or with some other scheme accepted at the time? While empirical testing of new model ideas continues, it seems the "theory well" (academic and financial literature) has been dry in the past few years. So it seems investment managers are mired in the "million monkeys on a million typewriters" phenomenon -- if one gives a million monkeys a million typewriters, eventually they will generate all the great works of Shakespeare by chance. A similar effect can be seen in the investment management community, where by giving managers enough data, all possible "tweaks" of existing valuation ideas will eventually be derived by chance. So what is an investment manager to do for new ideas?

New Ways of Combining Information.

One should not throw up their hands in frustration just yet. Existing valuation models seem to work over the long run. While a new idea would be great, how about investigating using the information provided by existing valuation models more intelligently? This approach was what we did in our research.

Using multiple valuation models, there are several ways to incorporate the information. The simplest way is to equal weight the models. The drawback to this approach is it does not take advantage of the really good models, but treats all equally. A second approach is to apply optimized weights to the various valuation models, weights which reflect both the efficacy of the models as well as the correlation between the models. A third approach one could try would be to use some sort of non-linear scheme, either derived from a non-linear specification, a neural network or some sort of chaos-theoretical algorithm. The problems with this approach are in specifying the appropriate non-linear form for the models, given that these models dynamically change. Using neural networks or chaos theory often will not tell one what the specification is, which can be a problem for both performance attribution and client communication, often borders on data-mining, and have little relevant theory as to why the methods should work. While each investor's choice of weighting scheme is unique, we feel that optimized weights make the most sense logically. As can be seen from the graph on slide #12, optimized weights do better than equal weights in our experience.

Our current system uses "vertical cross sectional" data to determine the optimum weights -- that is, a set of weights is determined which gives us the highest return for level of risk across our entire list of stocks in a given time period. An opposite way to approach the problem would be to use "horizontal cross sectional" data (stock-by-stock) over all time periods in question. This technique begs the question of what type(s) of models to use, since there possibly could be a different set of valuation models for each stock that are most applicable. The project we discuss here attempts to find better overall explanatory power through a combination of horizontal and vertical cross sections -- vertical in the sense that we use the models whose efficacy across our entire list of stocks over time has been the highest, and horizontal in the sense that we might find it better to apply different weighting schemes to different stocks or clusters of stocks within our list depending on the characteristics of the cluster(s).

Motivation of Study.

The motivation for this study was to check whether our composite valuation model should be constructed differently than by our current process. Currently, our universe is viewed as one homogeneous cluster of stocks, with a single set of weights applied to the constituent models across all stocks (a "vertical" cross section) in a given time period. An alternative way to view our universe is as a universe composed of unique "clusters" of stocks, each cluster being best explained by different combinations (weightings) of models.

In order to validate our hypothesis, we will choose the most relevant models from the "vertical" (time) perspective, and use these to explain a time series of ranked stock alphas stock-by-stock using multiple regression by stock across time (a "horizontal" cross section.) In this paper "vertical cross section" will refer to looking across all stocks in a single time period, while "horizontal cross section" will refer to looking at one stock across multiple time periods. Vertical cross sectional data (the model ICs) will be used to determine the models we will use in the analysis -- the models which have shown efficacy over time and have the most diverse theoretical underpinnings-- while horizontal cross sections are used to try and explain the individual stocks' alphas with these models across multiple time periods. The study is approached in this way as to avoid "data mining" -- using only horizontal cross-sectional information, without the added information that the IC data gives, allows for the potential to have models that don't make sense enter the analysis.

In order to be able to determine whether valid conclusions and results are drawn, the data set will be broken into two parts -- an in-sample set for study, and an out-of-sample set for validation of results. Additionally, stability of results will be checked through "bootstrapping" of the data, as well as looking at results after omitting observations -- that is, changing the data set around to find out if the results are invariant to data ordering and number of observations.

We would not expect ahead of time to get any significant results using only the stock-by stock data; rather, significance should manifest itself through clustering. The expectation of this study is that using vertical and horizontal cross-sectional data should lead to the identification of some homogeneous clusters (utilities, value, growth, financial, cyclic, or perhaps others.) By identifying commonalty between clusters, we can perhaps develop a new weighting scheme which weights these clusters of stocks (with each cluster having a particular, unique set of model weights), rather than each individual name using the same weight for each model.

Methodology of The Study.

The purpose of the study was to use multiple regression to explain a stock's alpha (its risk-adjusted extra normal market return) as a function of individual model alphas mentioned above across the number of time periods for each stock. In order to do this, it was first necessary to get the stocks' monthly alphas. This was derived using the following formula:

$$\alpha_{s,t} = AR_{s,t} - TR_{s,t} = AR_{s,t} - [R_{f,t} + \beta_{s,t}*(R_{m,t} - R_{f,t})]$$

where the expression in brackets is the "theoretical" return expected for the riskiness of the stock (the CAPM expected return) in month t, and $AR_{s,t}$ is the actual total return for stock S in month t. The source for each stock's beta estimates was the BARRA FRMS database. The monthly risk-free returns came from Bloomberg, and the SP500 monthly rate of return was used as the market rate of return each month. Each stock's alpha was then calculated for each time period using this data.

The decision about how many models to use in the analysis was very important. As mentioned above, although we had 11 models to choose from, each model chosen represented one degree of freedom, which could potentially eat up a good amount of the limited in-sample data we had. Additionally, since some of the models were highly correlated, it did not make sense to choose all of the possible models, since that would use redundant information. Another consideration was to ensure adequate representation between both value and growth models. A final consideration was to choose models with strong theoretical underpinnings. Sensitive to these issues, we chose the top six models over the 109-month period (as identified by their ICs) to use in the study. Three "value" models were chosen (DDM, P/E type models), as were three "momentum" models. Because these were the models that, for the most part, made up our composite model over the 9+ year period, it allowed us to check any findings against reality to see if we had improved on our existing model-weighting system. An alternative methodology would have been to include all potential models (rather than just the most predictive) in our analysis. This would have gone against the initial hypothesis of our study, though, and so we decided against it.

Month-end valuation data spanning the period 3/85 through 3/94 were used. This data provided "alphas" for each valuation model by ticker. Missing data was represented by 0.00, representing zero "alpha" for a given stock-model combination in a given time period, and comprising about 13% of the sample. This database had *up to* 109 monthly observations for each ticker. Once the database was built and checked for accuracy, two data sets were created from it -- a working (in sample) data set with 58 months of data from the odd-numbered years 1985, 1987, 1989, 1991, and 1993, and an out-of-sample data set with the other 51 months of data from the even-numbered years. The data set was split so that any results garnered from the working data set could be applied to another data set without creating any "look-ahead" bias, or having to wait several months to gather an out-of-sample data set. The in-sample data set contained 718 tickers.

Once the data were prepared, regressions were run by ticker. As mentioned above, having six independent variables (the three value models and three growth models) in each period required that there were enough monthly observations available to make the regression results significant. How much data is enough? I used a somewhat arbitrary cut-off level as 30 months of data -- any ticker with less than 30 months of data was excluded from the regression results. In addition, any stock which did not have valuation model data for each of the six models in at least one month of the 30 month minimum requirement was excluded. As it turned out, 318 of the 718 tickers mentioned above made it over these two hurdles.

The form of the regression for each ticker "s" was the following:

$$\alpha_s = b_{1,s}(G1_s) + b_{2,s}(G2_s) + b_{3,s}(G3_s) + b_{4,s}(V1_s) + b_{5,s}(V2_s) + b_{6,s}(V3_s) + \varepsilon_s$$

where the independent variable was the actual stock alpha and the dependent variables were the stock's model alphas (V1, V2, V3, G1, G2 and G3) in each month. The coefficients ($b_{1,s},\ldots,b_{6,s}$) for each stock s were determined by using the monthly valuation data for each ticker. Slide #19 shows the average regression statistics and some selected stock regression statistics for the set of 318 tickers fitted, along with the regression R^2 values.

Stability of Model Weights.

In order to determine the stability of these coefficients, the "bootstrapping" methodology was employed. This technique uses an existing distribution (which in this case is the set of monthly alphas and model alphas) to build "similar" distributions, which are then analyzed. If the results obtained with the original sample and the "bootstrapped" samples are similar, then stability of the analysis is assumed. The mechanics of the technique are as follows. For each ticker, ten alternative data sets were created by taking the original 58 (or however many existed) months of data by ticker and doing substitution "with replacement" of selected observations within the data set. For each of the ten alternative data sets, one-sixth (1/6) of the observations were replaced using the following algorithm: Two random numbers were selected, the first representing the month-end observation which would be substituted for in the new series, and the second the month-end data which would be substituted in its place. This does not just "mix up" the observations in the original data set, since this would not change the regression coefficients at all; rather, it changes the original data set in about 16% of the occurrences. Using this algorithm, it is possible to have a monthly observation repeated twice or more than twice in the alternative distributions.

After ten alternative distributions were created for each ticker, the same regression analysis was done, which generated ten regression coefficients for each model by ticker. The correlation of these sets of regression coefficients were then checked, and turned out to be very high While different runs gave different results, these correlations were in the range of 0.89 to 0.99 for the models. The importance of this was that our results do not turn out to be dependent on actual history -- had history been different (as the different bootstrapping runs indicated) we still basically get the same results for the regression coefficients.

A second test done was to omit observations and then check for regression coefficient correlation. For each ticker, one-seventh (1/7) of the observations were dropped (randomly throughout the sample), and the regressions were re-run. The correlation between the original regression coefficients and the newly-computed coefficients using the omitted observation data sets again turned out to be very high, in the range of 0.94 to 0.99. The joint results of these two simple tests gave comfort with the stability of the regression data.

Cluster Analysis of Regression Coefficients.

Once the regression coefficients were calculated and checked for stability, cluster analysis was used on the data set in hopes of finding "similarity" between stocks based on these coefficients. Clustering was done on all models' coefficients together, the value models' coefficients alone, the growth models' coefficients alone and the R^2 values of the regressions. Before going into detail on the cluster results, it might be helpful to discuss clustering techniques and how they can identify similarity among data.

Cluster Analysis Primer

Cluster analysis is a technique for assigning observations into unknown groups. It differs from other classification techniques (such as discriminant analysis) in that the number and characteristics of the clusters are derived from the data, and are usually not known a priori. **Cluster analysis is highly empirical, meaning that different clustering methods can lead to very different groupings** (both in the number of clusters formed, as well as in their characteristics.) Most clustering routines form clusters based on the notion of distance as an indication of "closeness" of observations. Since different units can lead to widely different distance measures (and hence different clusters), it is normal practice to standardize data sets prior to sending them into clustering routines so that all data is scaled appropriately ("apples-to-apples" comparisons.)

Two common clustering algorithms are known as hierarchical clustering, and K-means clustering. In hierarchical clustering, the process starts with N clusters (N represents the number of data points to be clustered.) Successive steps try to combine these N clusters down into a smaller number of clusters based on the distance between the clusters (or their centroids, if a cluster contains more than one observation.) Taking this algorithm to its maximum degree, it agglomerates all observations into one cluster. The hierarchical procedure can be misleading in certain situations -- for instance, an undesirable early combination of observations does not "disaggregate" itself later in the procedure, but rather cascades throughout the analysis, often leading to suspicious results. For this reason, hierarchical methods are often performed multiple times, with potentially suspicious data points removed. Other hierarchical algorithms are centroid, nearest neighbor, Kth nearest neighbor, complete linkage and Ward's method.

The second type of clustering algorithm is known as "K-means" clustering. This algorithm determines cluster membership by initially dividing the data into K clusters. For each observation in each cluster, the distance between it and the centroids of all clusters is computed. If the minimum distance is to its own cluster's centroid, the observation is left in the cluster; however, if the minimum distance is between it and the centroid of another cluster, the observation is reassigned to that cluster. This procedure continues for all observations until each is assigned to its minimum distance centroid. **The interesting fact about this algorithm is the fact that the initial (and perhaps final) results depend on the order of observations in the data set being clustered.**

Some issues to be aware of when using clustering routines: (1) Unless there is considerable separation among the inherent groups, cluster analysis does not provide realistic results; (2) Clustering is sensitive to outliers; (3) Most routines are not scale-invariant (insensitive to the scale of the variables being clustered), so standardization of the data is a good idea, and (4) splitting the data set into two halves, clustering each half (the same idea as in-sample vs. out-of-sample testing,) and obtaining similar clusters within the two halves of the data lends credence to the results.

As can be inferred from above, cluster analysis is both an art and a science. The "K-means" algorithm was used in this analysis. The analysis proceeded in the following manner. Using the 318 tickers' *standardized* regression coefficients (for each of the six valuation models), the first pass allowed the clustering routine (SAS FASTCLUS) to put the data into as many as 20 clusters. From these initial clusters, any having less than (the arbitrarily-chosen number of) ten observations were dropped as potentially being "outliers" (although perhaps these clusters are where the real information in the analysis lies.) The pared-down set of clusters was then re-fed into the clustering routine as the "seed" or starting configuration for the next pass, where again the routine was given the latitude to try and find up to twenty clusters.

The clustering results were not very enlightening, as the clustering routine tended to choose only three clusters for each "run." The portfolios represented by these clusters are described briefly in slide #23. Overall, these results were not too promising, nor did they meet our expectations.

With these results, a "re-thinking" of the goal of the regressions was done. What should be predicted -- the stock's actual alpha, or the *rank* of the alpha? Likewise, should we use the actual model alphas in the regressions, or should we use the *ranks* of the model alphas? Unranked data has the potential problem of outliers affecting the analysis, though the data used for model alphas was normalized and truncated which should have dampened some of these effects. More importantly, as investors we care about being relatively correct, not absolutely correct. While one can conjecture all sorts of reasons to either use or not use ranked data, the decision was made to do the same analysis described above, except to use the ranks of the alphas rather than the actual alphas. The form of the "ranked" regression for each ticker "s" was the following:

$$\alpha_s = b_{1,s}(G1_s) + b_{2,s}(G2_s) + b_{3,s}(G3_s) + b_{4,s}(V1_s) + b_{5,s}(V2_s) + b_{6,s}(V3_s) + \varepsilon_s$$

where the independent variables were the rank of the model alphas and the dependent variables were the ranks of the stocks' alphas in each month. The coefficients ($b_{1,s},\ldots,b_{6,s}$) were determined by using the ranks of the monthly valuation data for each ticker. Ranking was done from smallest to largest alpha in a given month for both the stock alphas and the valuation model alphas. Slide #24 shows both the average results for the entire universe of 318 tickers as well as some selected regression coefficients and R^2 values for specific stocks. Slide #25 shows that clustering using ranks provided more differentiated results, and slide #26 shows some representative SAS clustering output (yippee!) The next step with this data was to send the clusters through some portfolio analysis software.

__Analysis of Clusters.__

As stated above, it is easy to speculate as to what these clusters represent. The next step in the analysis was to run these clusters through the BARRA PORCH system to characterize these portfolios. This analysis yielded some very interesting results. As can be seen from the table in slide #27, each of the seven clusters (these are the clusters formed from cluster analysis on all regression coefficients together; I have not included the results for clusters based on only value, only growth, or regression R^2 values) has some distinguishing characteristics, although there is not quite enough information here to draw any solid conclusions.

The table shows the number of assets in each cluster, and the relative over/under weighting versus the SP500 for the BARRA style factors (capitalization-weighted). The interpretation of the above data for cluster #1 is that it contains 18 assets, consists of larger stocks (due to size factor over weighting) and has a "value" tilt to it (due to growth factor underweight and yield overweight.)

__Conclusions & Next Steps.__

This research project still has not borne any fruit. There are many ways to interpret the data, and along the way there will be many questions as to the statistical methodology used. The initial intent of the research project was to get some results and begin testing a new methodology for weighting models. This does not seem as clear-cut after seeing the ambiguous results of the study. Several questions arise from the project, though: (i) If we had conclusive findings, how do we implement a new strategy -- by optimizing the model weights for each cluster separately? (ii) How stable are these clusters over time? This might be able to be answered by looking at the out-of-sample data set and seeing if we get clusters with similar risk characteristics; (iii) Should certain clusters of stocks be excluded completely because of their lack of information -- that is, assigned a "cluster alpha" of zero? (iv) Are there "holes" in using vertical cross-sectional data for certain groups of stocks in our universe, "holes" which we can look to fill by altering our weighting methodology?

While there are many questions, we feel we have opened up a new and interesting area for research in the weighting of composite valuation models. The bugaboo with this kind of research is always the same -- scarcity of data. We hope that as the data sets get larger and our hypotheses become sharper, some tangible results will come out of this research.

Game Plan

Defining The Elusive "Great" Model

Any NEW Ideas Anyone? None That We Can Find

Other Options - The Dangers And Possibilities

A Brief Synopsis Of Our Work

Going Forward From Here?

Defining The "Great" Model

Defining ...

Theoretical

Unique

Robust

Defining The "Great" Model

Theoretically Sound

Value Models - The Easy Part
Dividend Discount Model (The One And Only?)

Momentum/Growth Models - Pushing Theoretical Definitions
Spread Of Information Theories
The "Other Investor" Explanation

Macro-Economic Models - Substituting Understanding For Theory

Unique Value

Original Idea

Correlation Collaboration

Despite Originality, Reflecting Investor Opinions

	G	G	V	G	G	G	V	V	V	G	G
G	1.000										
G	0.182	1.000									
V	0.285	-0.131	1.000								
G	0.180	0.288	0.011	1.000							
G	0.108	0.207	-0.024	0.451	1.000						
G	0.059	0.184	-0.255	0.005	-0.022	1.000					
V	0.273	-0.162	0.939	-0.020	-0.041	-0.268	1.000				
V	0.071	-0.193	0.332	-0.096	-0.098	-0.068	0.314	1.000			
V	0.310	-0.143	0.827	-0.056	-0.060	-0.111	0.824	0.330	1.000		
G	0.190	0.384	-0.013	0.558	0.449	0.037	-0.035	-0.142	-0.067	1.000	
G	0.170	0.477	0.024	0.330	0.241	0.093	-0.006	-0.046	-0.079	0.429	1.000

Defining The "Great" Model

Robust Results

- Testing & Success
- Understanding Periods Of Failure
- Broad Universe Application

	Avg. I.C.	Std. Dev.	Batting Avg.%
Model 1	0.063	0.097	80.00
Model 2	0.052	0.096	71.21
Model 3	0.049	0.111	70.00
Model 4	0.049	0.117	65.71
Model 5	0.034	0.115	67.14
Model 6	0.026	0.096	62.86
Model 7	0.024	0.082	68.57
YLD36	0.023	0.082	62.86
YLD12	0.021	0.124	52.86
YLD24	0.020	0.105	55.71
Model 8	0.016	0.154	58.57
Model 9	0.010	0.137	61.43
YLD48	0.005	0.085	54.29
Model 10	0.000	0.116	52.86

Any New "Great" Models?

Three Years Of Dismal Experience

- Modified DDM
- Data Mining The Earning Revision
- Anyone For Balance Sheets???
  - Cash Flow Models
  - Asset Value
  - Earning Potential Analysis (Financing Capacity)
- Non-Earnings Information

Exceptions

- Real Time/Trading Models

Any New "Great" Models?

Why???

The Simple Market Hypothesis: Only A Few Things Really Count - Everything Else Measures Those Same Things

Sales vs. Earnings
The "POP" Model

The Real Investor Hypothesis: Discounted Cash Flows Mattered Only AFTER The Theory Was Widely Accepted

Empirical Tests Continue - The Theory Well Has Been Dry

The Million Monkeys On A Million Typewriters Hypothesis: The Data Mine is Mined Out

Other Options: If Mohamad Cannot Find A New Mountain, Maybe The Old Mountain Can Come To Mohamad Faster.

Combining Information - The Options:

Equal Weighted

Optimized

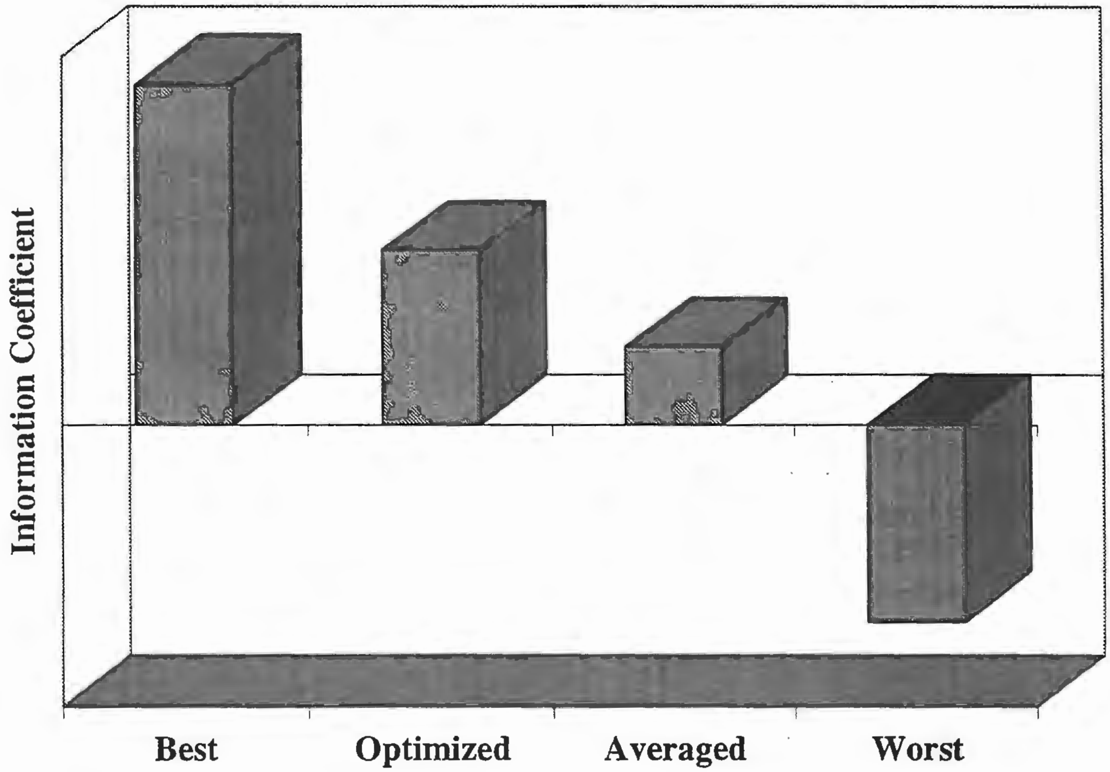
Non-Linear

Combining Information - Our Experience:

Optimization Works.
How We Do It At IIA (Relevant To Later Discussions)

Non-Linear Dangerous
Data Mining
Little Relevant Theory
BUT IT IS NEW!!

Cybercode Predictive Accuracy - Assuming Various Submodel Combinations



Other Options: If Mohamad Cannot Find A New Mountain, Maybe The Old Mountain Can Come To Mohamad Faster.

Limiting Data Mining By Combining Horizontal
And Vertical Cross Sections

Cross Section By Month

- Choose Relevant Models
- Define Goal Of Prediction - Monthly Rank

Cross Section By Stock

- Look For New Patterns

Independence Designs A New Test

Choose The "Most Relevant" Models

Best Historical Performance

Most Diverse Theoretical Underpinning

Define Relative Rank As A Prediction Goal

Split Your Data: Test vs. Out Of Sample

Create A "Best Fit" Regression Model To Explain
"Excess Returns" For Individual Stocks Across Time
(As Opposed To For The Whole Universe As A Time Slice)

$$\alpha_{i,t+1} = f(V1_{it}, V2_{it}, V3_{it}, G1_{it}, G2_{it}, G3_{it})$$

Carefully Test Stability Of Results:
"The Northfield" Bootstrap Model
"Rolling Period" Analysis

Expect Little To No Statistical Significance From
Individual Stock Runs

Analyze The Regression Results To Look For Patterns

Cluster Analysis Of Stocks By Model Weights

Hypothesis: Three To Four Groups Reflecting Growth,
Value, Cyclicals And (least likely) Financials Or Utilities/Energy

Cluster Stocks By Regression Fit:
Hypothesis: Two General Groups -
Financial/Utilities/Energy With Bad Fit And Everything Else

Testing Methodology

- Attempt to explain stock's future extra-market return using current model information

$$\alpha_{i,t+1} = f\left(V1_{it}, V2_{it}, V3_{it}, G1_{it}, G2_{it}, G3_{it}\right)$$

- Determine stability of solution

- Find any common characteristics of stocks within clusters

Data Set Characteristics

- Month-end valuation model data spanning period 3/85 through 3/94 (109 months)

- In-sample, out-of-sample data sets created

- Six valuation models selected based on historical predictive ability
    - Three "value" (DDM, P/E-based models)
    - Three "growth" (earnings revision, analyst opinion)

- 740 names in data set, reduced to 318 names when data point threshold imposed

- Missing data: approximately 13% -- assigned alpha of zero

Step 1: Regression of Alpha against models by ticker

Regression form (i=security index, t=time index):

$$\alpha_{i,t+1} = f(V1_{it}, V2_{it}, V3_{it}, G1_{it}, G2_{it}, G3_{it})$$

TICKER	**G1**	**V1**	**G2**	**V2**	**V3**	**G3**	**RSQ**
MOT	0.03	-0.01	-0.01	-0.02	0.01	-0.01	0.07
MRK	0.02	0.00	0.02	-0.01	0.01	0.00	0.10
MS	0.01	0.09	0.02	-0.06	-0.02	-0.00	0.15
N	0.06	0.04	-0.03	-0.04	-0.01	0.01	0.29
NCB	-0.30	-0.35	0.18	-0.22	0.78	-0.02	0.97
NCR	0.24	-0.37	-0.65	0.10	1.18	0.18	0.87
NEM	0.00	-0.02	0.00	0.01	0.00	-0.01	0.05
Overall	**0.0047**	**0.025**	**0.0006**	**-0.022**	**0.0097**	**0.0018**	**0.154**

How Stable are Regression Coefficients?

- "Bootstrapped" the individual ticker data series
    - Created 10 new time series by doing selection with replacement
    - Re-ran regressions using these series, correlated coefficients against original
    - Correlations very high, ranging from 0.89 - 0.99
- Dropped observations from each ticker's data series
    - Threw out 1/7 of observations, re-ran regressions and correlations
    - Again, very high correlations with original regression results (0.94-0.99)

Step 2: Cluster Analysis of Regression Results

- Cluster analysis -- both an "art" and a "science"
- Attempts to find "commonality" among data points by looking at separation distances
- Many "bells and whistles" with routines -- decisions to be made:
    - How many clusters are right?
    - How to prepare data for analysis (standardize or not?)
    - What distance is appropriate (in multiple dimensions...)
    - Which variables to cluster together?

Clustering Choices and Procedure

- Four clustering "runs" done:

    (1) Cluster regression coefficients of all models

    (2) Cluster only the "value" models' regression coefficients

    (3) Cluster only the "growth" models' regression coefficients

    (4) Cluster the regression R^2 values (the stocks we predict well)

- SAS clustering procedure
    - Standardize data
    - Let the routine choose as many clusters as appropriate
    - Eliminate clusters with less than ten names in them
    - Re-cluster, using the previous trimmed output as seed input

Clustering Results (not very enlightening)

- Routines tended to choose only three clusters for each"run"
    - Very diversified portfolios, not any apparent sector/style tilts
    - Cluster Results by "style"
        - All models: 2 clusters (11, 307 names); first cluster large cap value(?)
        - Growth models: 1 cluster!
        - Value models: 3 clusters (40, 253, 25 names); last cluster "value"
        - R-squared: 7 clusters, all diversified

Back Into The Mine

- Should we regress predictive model alphas vs. stock alphas, or perhaps use **rank**?
    - Rank tends to reduce effects of outliers in data (if an issue at all...)
    - Scaling of model alphas, stock alphas should not be important
    - As investors, we care about being relatively correct, not absolutely correct
- What happens when we use rank?

TICKER	G1	V1	G2	V2	V3	G3	RSQ
MOB	-0.07	0.16	0.58	0.22	0.04	0.10	0.89
MORR	0.09	-0.14	0.07	-0.14	0.32	-0.01	0.58
MOT	0.25	0.75	0.43	-0.08	0.04	-0.04	0.85
MRK	0.11	-0.87	0.19	0.61	0.27	0.43	0.90
MS	0.24	1.02	-0.09	-0.20	-0.49	0.15	0.85
MSFT	0.47	-1.54	-0.08	1.46	0.33	0.14	0.80
MTC	0.27	-0.07	0.41	-1.05	1.31	0.33	0.83
N	0.80	0.54	0.29	-0.07	-0.69	0.33	0.77
NCB	-0.07	0.07	0.03	0.10	-0.04	0.03	0.61
NCR	0.07	0.42	0.21	-0.24	-0.40	0.03	0.65
Overall	**0.14**	**0.04**	**0.23**	**0.20**	**0.24**	**0.12**	**0.82**

Re-clustering Results and Observations

- Clusters using rank data provided more differentiation among stocks
  - Using all variables -- 7 clusters formed (16-125 tickers)
  - Using only "growth" models -- 6 clusters (21-119 tickers)
  - Using only "value" models -- 7 clusters (7-98 tickers)
  - Using model R-squared -- 9 clusters (22-53 tickers)
- Next Step: Look at make-up of clusters, find out what characteristics each contains
  - Cluster output
  - BARRA risk analysis

Some Clustering Output

- Representative output for clustering regression coefficients of all models

Cluster	Frequency	RMS Std Deviation	Max Distance from Seed to Observation	Nearest Cluster	Distance Between Cluster Centroids
1	21	0.6992	4.1103	4	1.7905
2	43	0.6345	4.4200	4	1.6220
3	77	0.8344	10.6098	4	1.4284
4	125	0.4393	2.8070	3	1.4284
5	16	1.1755	8.2353	4	2.0314
6	30	0.6174	3.1747	4	1.9040
7	16	2.4833	21.9658	2	2.7550

- Cluster Means

	G1	V1	G2	V2	V3	G3
1	-0.95944	0.02069	-0.53317	0.66653	-0.47858	1.32101
2	-0.12778	-0.60148	-0.50101	1.06965	0.25193	-0.00384
3	0.67591	0.16903	-0.00559	-0.13909	-0.19603	-0.93561
4	-0.00246	0.17620	-0.17074	-0.25319	-0.00305	0.29020
5	-0.26480	0.81244	0.01855	-1.04322	-0.77948	1.83619
6	-0.67111	0.00132	1.37837	-0.10922	-0.32487	-0.49950
7	-0.10782	-1.41561	0.80404	0.14588	2.30691	-0.38768

Composition of Clusters

- Barra risk analysis on clusters yielded some interesting(?) results

  over (+) or under (-) weightings vs. SP500

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Var Mkts	-0.08	-0.15	0.14	-0.12	-0.10	-0.34	-0.09
Success	-0.29	-0.09	0.09	0.03	-0.29	-0.13	-0.14
Size	0.11	0.22	0.27	0.23	-0.29	0.10	-0.66
Trad Act	-0.26	0.07	0.27	-0.18	-0.20	-0.16	-0.16
Growth	-0.39	0.03	0.05	0.01	-0.07	-0.38	0.09
E/P	0.07	0.04	-0.11	-0.05	0.08	-0.01	0.05
B/P	-0.02	-0.31	0.25	-0.11	-0.13	0.37	-0.23
Earn Var	-0.23	-0.23	0.40	-0.14	-0.04	0.00	-0.19
Fin Lev	-0.10	-0.10	-0.08	0.16	-0.09	-0.02	-0.03
For Inc	-0.22	0.20	0.43	0.30	0.14	-0.01	0.50
Lab Int	-0.18	0.28	-0.09	-0.14	0.04	-0.26	1.11
Yield	0.47	-0.09	-0.01	-0.04	0.04	0.54	-0.25
Locap	0.00	0.00	0.00	0.00	0.00	0.00	0.00
#	18	42	75	125	15	28	15
	"Value"	"Growthy"	?	?	?	"Value"	"Small Growth"

Conclusions

Based on Results Begin Testing A New Methodology for Weighting Models

Optimize Weights For Separate Stock Groups Separately?

Exclude Certain Groups Completely? (Automatic Zero Alpha)

Find "Holes" In Current Methodology For Certain Groups
(Which May Improve Performance Of Other Groups Eventually)

Testing With The "Reserved" Data