CHOOSING A MULTIPLE FACTOR MODEL

BY BARR ROSENBERG

(Reprinted with permission from Investment Management Review)

Barr Rosenberg, Ph.D., is Managing Partner of Rosenberg Institutional Equity Management, Orinda, CA. A preliminary version of this paper was given at the Berkeley Program in Finance, September 1987.

Multiple factor modeling is an exercise in matrix algebra. To understand it fully, you must see a geometric image of the problem. Therefore an unusual kind of training is required for would-be model builders. But for model users, it is enough to understand the opportunities available. And there are a tremendous number of opportunities.

A LINEAR MULTIPLE FACTOR MODEL

A multiple factor model is a model of investment returns. A linear factor model can be expressed as:

\[ \text{Return}(n,t) = \text{Exposure}1(n,t) \times \text{Factor}1(t) + \text{Exposure}2(n,t) \times \text{Factor}2(t) + \ldots + \text{Exposure}k(n,t) \times \text{Factor}k(t) + \text{Specific return} \]

\[ n = 1, \ldots, N \text{ Companies} \]

\[ t = 1, \ldots, T \text{ Time Periods} \]

The return is a dependent variable, representing an asset (n) at a date (t). That return is determined by multiplying the asset’s exposure to the first factor by the factor value, and repeating this for the K factors in the model. At the end, you add the asset’s specific return — that is, something idiosyncratic to the stock.

You may only be interested in how well the model predicts risk. But other criteria may also be important. You may want the factors to have economic meaning, so that you can forecast them. You may want to know what the factors are, or what the exposures are. You may want to understand the
asset’s exposures in terms of fundamentals, so that you can judge their validity.

These things may help you in your investment planning. To the extent that they do, you will want to rely on a model that offers more than just predictive accuracy. In order to appreciate the process of model building, consider six alternative styles, as listed in Table I.

**TABLE I**
Comparison of Multiple Factor Methodologies

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Factor Values</th>
<th>Exposure Values</th>
<th>Factor Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>JUDGEMENTAL FACTOR ANALYSIS</td>
<td>Concommitants</td>
<td>Fundamentals</td>
<td>Judgemental</td>
</tr>
<tr>
<td></td>
<td>Returns</td>
<td>Returns</td>
<td>Returns</td>
</tr>
<tr>
<td>FUNDAMENTAL CENTROIDS</td>
<td>Fundamental &amp;</td>
<td>Returns</td>
<td>Returns</td>
</tr>
<tr>
<td></td>
<td>Returns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONCOMMITANT FACTORS</td>
<td>Concommitants</td>
<td>Returns</td>
<td>Concommitants</td>
</tr>
<tr>
<td>CROSS-SECTIONAL REGRESSION</td>
<td>Returns</td>
<td>Fundamentals</td>
<td>Returns</td>
</tr>
<tr>
<td>MAXIMUM LIKELIHOOD</td>
<td>Concommitants</td>
<td>Fundamentals</td>
<td>Concommitants</td>
</tr>
<tr>
<td></td>
<td>&amp; Returns</td>
<td>&amp; Returns</td>
<td>&amp; Returns</td>
</tr>
</tbody>
</table>

**A JUDGEMENTAL APPROACH**

A fundamentalist might start by saying, at the beginning of 1987, “There are only two factors in the world: the stock market and the bond market”. To represent these factors, you might use the S&P 500 and the LBKL index. The fundamentalist might forecast how the stock factor would be positive, and the bond factor negative.

The fundamentalist might know that utility stocks behave like bonds, because their earnings are regulated. Shares in financial intermediaries behave like bonds, because their financial liabilities are generally shorter in duration than their assets.
In short, these stocks are exposed to both stock and bond factors. So the fundamentalist would decide to buy stock, avoiding these industries. In this example, the factors and exposures are both arrived at purely from economic judgement.

This is the kind of multiple factor model building that investment committees used to do with great confidence. But there are other, more sensitive approaches, which rely on databases of asset returns. With a large number of companies and time periods, there is a great deal of data. The challenge is to use the data and this general formula to develop a useful model.

**FACTOR ANALYSIS WITHOUT FUNDAMENTALS**

The polar opposite of the judgemental approach is to perform factor analysis of returns without fundamentals of any kind. Here, you use a database of investment returns — and that is all the data you have.

Once you have a historical panel of returns, you can easily calculate a correlation matrix. In fact, you might think that this completes the task because a correlation matrix is sufficient for a risk model. But you may want to do a factor analysis to make it more useful.

Why might you do this? Well, if you have 6,000 stocks, you have a total of 18 million correlations. That's a lot of correlations to store. You want to conceptualize risk differently, so you do not have to store so many things. Also, there may be some underlying correlations that you think are more real than others. A factor analysis will help you isolate these correlations.

You may reduce the correlation matrix, of dimensions 6,000, to about 20 underlying factors that account for most of the correlations among stocks. Because the factors are widespread among many stocks, you can have more confidence in their accuracy than you would have in the individual correlation coefficients.

Now you have 20 factors, isolated and ready to work with. You might then build 20 portfolios, calling them
“pooled funds” — each one designed to match the corresponding factors. You could then build “portfolios of portfolios” to do your hedging with regard to any risk exposure toward those factors. In this way, the small number of factors is reflected in the investment vehicle.

The other approaches to multiple factor models are mixtures of this pure data method and the pure judgement method. We will look at them one at a time.

GROUP AROUND A CENTER

The essence of this approach is to use fundamental knowledge to obtain your factors. Then you regress the stock returns on these factors to obtain exposures.

You start by saying, “I know what the interesting things are in the market” These might be market sectors, small stock, large stock, different industries, indexes, or something else. You then represent each one of these “interesting things” by a portfolio of like stocks.

Next, you calculate the return of each of these portfolios in each time period. These portfolio returns are the factor values, defined on the basis of fundamental relevance. Now you have a small number of time-series of factor values for small stocks, large stocks, industries, and so forth. Then you take the time-series return of each stock and regress it on this small number of time-series, in order to obtain the factor exposures of the stock. In short, you are calculating the exposures of the stock's returns to your “interesting things.”

Let's say some stocks are like bonds. One of your “interesting things” might be the utility index. And if small stocks behave differently, then another “interesting thing” would be a small stock index.

On this example, you are using fundamentals to group your assets. Then you calculate the return for each of those groups. You might call them “centroids” — or, of course, factors. But “centroids” is more evocative because data will group around it.

The use of data in this method is clear: You regress each stock return on the centroids to calculate exposures. The
judgemental approach is evident in the initial choice of factors. The bottom line is: you are trying only those factors which you think are meaningful.

THE FLIP SIDE

What if you say, “We know what determines the market valuation of assets. It has to do with people's ideas about market value.” And these ideas are in turn, related to industries, size, quality of earnings and other fundamental aspects of stocks. Alternately, valuation may depend on a technical concept: past relative strength.

These characteristics measure how stocks respond to trends in market valuation. In other words, these characteristics are the stocks' exposure to those trends. Therefore, if you know the key determinants of valuations, you then know the exposures!

Starting with known exposures (a risk index, for example), you calculate the factor values through a cross-sectional regression of investment returns. This is a very model-oriented approach. You might even say it's a fairly arrogant approach — because you assert that you know what is going to determine exposure to all the factors. Then, you are just using regression techniques to find out what those factors are. This is the opposite of the previous method: the data is used to determine the factors in cross-sectional regression.

A FACTOR YOU CAN FORECAST

Here you start out by looking for factors that are interesting. What's an interesting factor? It is one that, ideally, you should be able to forecast or at least understand.

For example, the exchange rate or the index of industrial production might be used. Many economists are trying to forecast these series. Therefore, if factors exist and you can indeed forecast them, you can use the exposures to those factors in an intelligent way.
If you want to start out with this “wish list” of interesting factors, you would then use their observed values as “concommitant variables.” These are outside variables that do not refer to the stocks themselves, but which play a role in defining the factor model. These become your factor values.

You might say, for example, “The change in the index of industrial production is a factor value. I’m telling you it is! That being true, I can do a time series regression on my stock of that factor value — to find out the stock's exposure to industrial production.”

**A MAXIMUM LIKELIHOOD**

The sixth and final approach to multiple factor modeling is a technique that, potentially can combine all of the other five approaches. You may throw in some judgment, some fundamentals for your exposures, some concommitant variables for factors, and so forth.

This method mixes data with judgement. As such, you will notice that the “maximum likelihood” approach will begin to take on an empirical, almost Bayesian flavor. Its great advantage is that it allows you to bring many different dimensions to bear upon the use of a factor model.

**THE THRILL OF SCIENTIFIC CONFIRMATION**

You may recall that BARRA was thrilled when it found that its factor values were related to the economy. Why? Because BARRA started out thinking it knew exposures — then it found factors.

Conversely, if you start out with fundamentals for your factors (e.g., using centroids, or the macro economy), then it is thrilling to you when the model makes sense in terms of the companies. You may, for example, confirm that utilities do indeed respond more to interest rates than do growth stocks.

Both approaches start out using limited information and then find confirmation in the fact that the other side of the
two-part relationship also makes sense. This is part of scientific confirmation.

When you start out thinking that you know both aspects, that is, you know something about factors and something about exposures, you can begin to build models that way. The strength of the various methodologies is best illustrated by discussing some examples.

**PRIMERICA**

If you want to perform a long historical study, you are faced with the problem that Primerica is no longer the company that it used to be. It has gone from containers to finance. Therefore, its factor exposures are not the same as they once were.

The pure data approach requires that the factor exposures be the same over time. And so it cannot deal with the fact that Primerica was once in a different industry than now. The concommitant factor approach also faces a problem here. If you regress Primerica on interest rates, you must develop a long historical record to gather accurate numbers. But, with Primerica, you are mixing two different sets of exposures — their present interest rate exposure (high) with their past interest rate exposure (low).

Using fundamentals for exposures helps you here because they indicate that Primerica changed industries. You may then classify it differently. These same considerations apply to all new stocks as the historical record is unable to supply information on them. In which case, you must either use judgement to make the initial classification or else rely on fundamentals for exposures.

**COMMUNICATIONS SATELLITE CORPORATION (CSC)**

As you may know, this is a company incorporated in the District of Columbia. A prudent fundamentalist might get his data from a highly qualified source like Compustat. Until recently, however, CSC was labelled by a Compustat statistician as being in the Country of Columbia. In short, CSC appeared as a South American Company.
If you are using company fundamentals to do your groupings, you are not going to do very well on this one. The point is: question the accuracy of your fundamentals. They might be wrong, or totally irrelevant. A method that relies only on returns to calculate exposures will not be misled here.
THE AT&T BREAKUP

This was an interesting episode for one important, statistical reason: weighting. Most statistical methods are implemented by standard computer programs. Implicit in these standard programs is a particular weighting of the data. And this has a very strong influence on results.

Suppose you are building a risk model of large liquid stocks, perhaps to guide an arbitrage strategy for index futures. Your factor model would look very different if you had one AT&T, rather than one AT&T plus seven local Bell Systems companies. Telephone companies have moved from being a relatively unimportant industry (with three big companies) to being a dominant one (with ten big companies) including GTE & ITT. If you run a factor analysis of large companies today, you will come up with a telephone factor. Prior to the AT&T breakup, you would not have done so. This is not so much an economic change as it is a change in weighting the telephone industry.

Handling these weighting questions is generally a matter of using fundamentals. If you fail to address the weighting issue, your results will be strongly influenced by that omission.

AT&T VERSUS ALLNET

This is the issue of size. According to fundamentals, you might say that AT&T and ALLNET are very much alike. But, clearly, most of us would weigh AT&T differently when thinking about the US market.

The problem is that most factor models will implicitly weigh them equally. In fact, some approaches would actually weigh ALLNET higher because its risk is greater.

THREE MILE ISLAND

If you used a standard sample and estimated a risk model before the accident at Three Mile Island, you would
have been incapable of telling the difference between the gas utility industry and the electric utility industry.

Since Three Mile Island, of course, there is no trouble telling the difference. In fact, the day after the accident, there was not a portfolio manager in America who had not figured out what the difference was. This is the kind of obvious fundamental fact you want your model to respond to.

Some models respond well. Those are models where the factors have a micro-economic meaning. They let you track the factors back to the micro-economic genesis of the model. Three Mile Island was an industry event. It had a nice micro-economic flavor. If you could have increased the variance of the electric utility factor the next day, you would have quickly corrected your model. However, if your model had been built purely from returns and macro factors, there would have been no way to adjust.

In general, you want to devise a model that is realistic for the future. In pure factor analysis, you must wait until after events have unfolded, so it is a good idea to have micro-economic input to your factors.

**THE MONETARISTS**

This is the flip side of the micro economic event. It used to be that bonds were safe — that is, they did not move up and down a lot. Then, along came the monetarists. As the result of economic policy, interest rates fluctuated and interest rate risk rose.

That made the correlations change. Correlations between stocks and bonds went up. Total risk levels changed, as the risk of bonds went up. Interest rate factors became more important in the stock market. The effects spread across all industries. In short, an essentially macro-economic event changed the nature of investment risk.

If your model has factors tied to the macro-economy, you may make the adjustments required. You would say, “I know that the interest rate factor has a bigger variance than it used to.” On the other hand, a model having factors without a macro-economic link would be difficult to adjust. Integrating this example with the prior one, it is clear that you have to handle both sides: the micro-economic links and the
macro-economic links. To do this ideally calls for the “maximum likelihood” methodology, although other models can be adjusted by ad hoc methods.

**GM AND IBM**

This is a classic case of sampling error from random coincidence. Here are two giant firms that lately have not done as well as some smaller firms in their industries, respectively Ford & Digital. You would think that GM and IBM are correlated by chance. Because of sampling error, however, this correlation might show up in a risk model. You would be believing in a pattern that was purely accidental.

Sampling error shows up in all these methods. There is not enough data to avoid it. And there is a lot of randomness in historical returns. So you would like your methodology to be, at the least, not severely damaged by randomness. Therefore, it is desirable to use a method in which sampling error is curtailed by fundamental judgement.

**BAT AND BRITISH PETROLEUM**

For the counterargument, however, consider the problem of different investor constituencies. BAT seems to follow the US market. British Petroleum seems to follow the UK market. Yet both are British companies, and it would be natural to put them together in one fundamental group.

It might be difficult to know, based on a fundamental model, that these two companies are not so strongly correlated with each other. Here reliance on the data is the better course, as naive judgements can lead us astray.

So we see two opposite tendencies. Sampling error makes us want to rely on what we know about the stock or fundamentals. Whereas the possibility of making mistakes in what we know makes us want to rely on the data.

**CONCLUSION**

Multiple factor models are necessarily complex, because they attempt to quantify broad trends throughout the market. They are diverse approaches to model-building, each
with its own strengths and weaknesses. The viewpoints considered here attempt to make sense of this fascinating field.