NEW YORK UNIVERSITY
STERN SCHOOL OF BUSINESS
FINANCE DEPARTMENT

Working Paper Series, 1997

Common Factors in Mutual Fund Returns

Elton, Edwin J., Martin J. Gruber and Christopher R. Blake

FIN-97-4

COMMON FACTORS IN MUTUAL FUND RETURNS

Edwin J. Elton
New York University

Martin J. Gruber
New York University

Christopher R. Blake
Fordham University

First Draft: June 1996
Third Draft: March 1997

We are grateful to Micropal for supplying data used in this study. This research was supported in part by a grant from Fordham University's Graduate School of Business Administration.

A great deal of the literature in financial economics contains the assumption that returns are a linear function of a set of observable or unobservable factors. The specification of the variables in the linear process (known as the return-generating process) is one of the key issues in finance today.

The return-generating process is an important building block in asset pricing models, portfolio optimization, risk management models, mutual fund evaluation, and event studies. For many purposes (such as in developing asset pricing models and evaluating mutual fund performance), it is important to separate systematic from non-systematic factors. There have been numerous attempts to examine the number and type of systematic factors in equity returns.

Approaches to identifying the return-generating process include purely statistical models such as those of Dhrymes, Friend and Gultekin (1984), Cho, Elton and Gruber (1984), Roll and Ross (1980), and Lehmann and Modest (1988), and models that a priori specify and test a set of fundamental factors and/or portfolios such as those of Chen, Roll and Ross (1986), Fama and French (1992, 1993), Burmeister, et al. (1986, 1987) and Berry, et al. (1988).[1] In addition, tests of return-generating processes sometimes involve using individual securities' returns as the units of observation and sometimes involve testing on portfolios of securities.[2]

The purpose of this study is to determine systematic factors by examining mutual fund returns. There are two reasons why it might be informative to work with mutual fund returns in

---

[1] Another version referred to as conditional asset pricing uses the returns on a set of *a priori* selected portfolios multiplied by lagged instruments. See Hansen and Jagannathan (1991), Cochrane (1996), and Ferson and Schadt (1996).

[2] See Connor and Korajczyk (1986) as an example of using individual securities and Fama and French (1993) for a recent example of tests on portfolios.

addition to either security returns or portfolios of security returns constructed on a mechanical basis.

The first reason arises from modern portfolio theory. One important implication of modern portfolio theory is that, given a belief about systematic factors, an investor should select an exposure (beta) to each factor, a level of expected risk-adjusted return (alpha) and a level of residual risk (residual variance).[3] The mutual fund industry has an incentive to offer an array of exposures to systematic factors in order to span investors' differing objective functions. If mutual funds all choose similar sensitivities to a factor, a mutual fund deviating from the norm should attract substantial investor interest and cash inflow. Thus, investors' objectives and mutual funds' incentives should result in a spread of sensitivities to factors viewed as systematic by investors. Therefore, mutual funds provide a logical way to obtain portfolios which have a spread on the characteristics of interest to investors while smoothing much of the noise inherent when a model is fitted to individual security returns.[4] Finally, employing mutual fund data, rather than forming portfolios of securities directly, avoids having to pre-specify the relevant characteristics and should lead to better separation of the sensitivities than forming portfolios on the basis of a proxy characteristic (e.g. size) chosen to obtain separation.

A second reason arises when we concern ourselves with the study of mutual fund

---

[3] See, for example, Elton and Gruber (1992) or Fama (1996).

[4] There is one complication that arises in using mutual fund data to test for systematic influences: the tendency of mutual funds to hold stocks which are held by other funds may affect our results. We test and adjust for these influences. The procedure developed to do this is interesting in its own right. It shows how to estimate the effect of common holdings for any return-generating process and its impact on diversification.

performance. To examine mutual fund performance in a meaningful way, one needs to specify a return-generating process. Much of the literature uses an assumed return-generating process to evaluate mutual funds. This approach started with Jensen's (1966) early work (which assumed a one-index return-generating process) up through the work of Elton, Gruber and Blake (1996a), which assumed a four-index model. As shown in Elton, Gruber, Das and Hlavka (1993), the choice of the return-generating process can affect the performance attributed to management. What better way to find these systematic influences that affect mutual funds than by studying mutual funds themselves?

Finally, studying mutual funds gives us a convenient holdout sample. We test the return-generating process developed from mutual fund data on common stocks formed into industry portfolios (using SIC codes). This allows us to see whether our results are unique to actively managed portfolios or have wider implications for capital markets. The use of mutual funds has some disadvantages. First, in common with other grouping procedures, there are fewer cross-sectional observations than would be available using individual securities. However, our sample size is much larger than those employed by others who have used portfolios. Second, there is a possibility that some sort of common dynamic behavior (such as herding) could lead to misidentifying a factor. Third, common holdings across funds might lead to misidentifying factors. These concerns are explored in some detail later in the paper. In particular, we deal with the problem of common holdings explicitly, and we include a section testing our model on common stocks. This sample is free from the potential problem of common holdings and the impact of of dynamic trading strategies.

I. SAMPLE

We use data on mutual funds and indexes in this study. All mutual fund data were supplied by Micropal.[5] We initially selected all mutual funds that existed as of January 1979 and that had monthly return data through December 1993. Such a sample clearly has survivorship bias, but for the purposes of determining the important factors affecting returns, survivorship bias should be unimportant.[6] There were 351 funds (excluding money market and municipal bond funds) that existed in the Micropal database over this period. From this data set we eliminated bond, option, precious metal, international, and index funds. We selected this set of plain-vanilla stock funds because we are searching for factors (indexes) that are important in explaining the returns on common stocks. (We might well have uncovered other factors if we had included funds which held foreign stocks or only bonds, but that would go well beyond the intended scope of this study. Our sample design is analogous to that of most of the empirical literature on APT, which does not include stocks traded on foreign exchanges or bonds in forming portfolios for testing purposes.) This left us with a set of 267 funds. We divided these funds into three 89-fund subsamples (group A, group B and group C). Having multiple samples allows us to test the robustness of results and in particular to see whether results derived from one set of data are generally applicable.

The three subsamples were selected so that each subsample had the same number of funds with a given objective and so that funds from any fund family (e.g. Fidelity) were spread evenly across the three groups. Having the same number of funds with a given objective in each group

---

[5] The accuracy of the data is discussed in Elton, Gruber and Blake (1996b). Micropal is a successor firm to Investment Company Data, Inc. (ICDI).

[6] A problem would result only if funds which did not survive were the only funds affected by some factor so

makes the subsamples reflective of the overall distribution of objectives. Spreading any fund family across all groups helps to ensure that the factors we pick up are not associated with a fund family's style. By using three separate samples we are able to examine whether the results are statistically significant in each subsample as well as the total sample, and whether the ordering of models is robust across samples.

The indexes we use fall into two groups: those that are publicly available, such as the S&P 500 index, and those that we constructed from other publicly available data bases. We will discuss the detailed construction and characteristics of these indexes in later sections.

II. ANALYSIS

A multi-index model that captures all of the relevant influences that explain why securities move together should have a diagonal residual variance/covariance or correlation matrix and should have non-zero betas on each index for many assets. Testing whether betas on each index are statistically different from zero is standard in the literature and will be one of the tests we perform. However, testing that the residual correlations are zero needs more elaboration.

The correlation between security returns is generally positive for all pairs of assets. However, once the market index is removed, some pairwise correlations of residuals are usually negative, while some are positive. The average residual correlation could be zero, while each individual residual correlation could be large in absolute value and thus very different from zero. Furthermore, as we explain below, adding a systematic index that significantly explains returns can cause the average residual correlation to increase or remain constant. However, if the index

---

that this factor would not be uncovered in this study.

added is a systematic influence, it should cause the average absolute value of the residual correlations to be closer to zero, and in fact, if the true model is uncovered, all of the off-diagonal elements of the residual correlation matrix should be exactly zero.

We illustrate this with a simple example employing covariances. Assume returns are generated by a two-index model. For any pair of securities i and j, let the residual covariance of returns from a one-index model using only the first index be $Cov(u_i\, u_j)$, let the residual covariance from the two-index model be $Cov(e_i\, e_j)$, let the variance of the second index be σ^2_I, and let the betas on the second index be β_{i2} and β_{j2}. Then

$$Cov(u_i u_j) = \beta_{i2}\beta_{j2}\sigma_I^2 + Cov(e_i e_j)$$

First, note that if $\beta_{i2}\ \beta_{j2}\ \sigma^2_I$ has the same sign as $Cov(u_i\, u_j)$, then in general $Cov(e_i\, e_j)$ is closer to zero than $Cov(u_i\, u_j)$, though, in cases where $Cov(u_i\, u_j)$ is negative, $Cov(e_i\, e_j)$ will actually be a larger number.[7] In addition, for multi-index models it is normal for some β_2's to be positive and for some to be negative. Summing both sides, we have

$$\sum_i \sum_{\substack{j \\ j\neq i}} Cov(u_i u_j) = \sum_i \sum_{\substack{j \\ j\neq i}} \beta_{i2}\beta_{j2}\sigma_I^2 + \sum_i \sum_{\substack{j \\ j\neq i}} Cov(e_i e_j)$$

To demonstrate why employing average residual variances might not be appropriate, consider the feasible case where $\sum_i \sum_{\substack{j \\ j\neq i}} \beta_{i2}\beta_{j2}\sigma_I^2$ equals zero. From the above equation, the average residual covariance must remain constant even though many and perhaps all of the residual covariances are closer to zero. It is even possible that, after adding a second index, the average

[7] If the sign is the same but $\beta_{i2}\beta_{j2}\sigma_I^2$ is greater than $2\,Cov(u_i\, u_j)$, the second index will increase $Cov(e_i\, e_j)$.

residual covariance increases, while each residual covariance is closer to zero. This same effect holds for residual correlations.

How might we measure closeness to zero when some observations are positive and some negative? One measure is the average absolute value of the residual correlations. This has the property that if a model results in residual correlations closer to zero, it is reflected in the metric.[8]

If we find that one model produces on average lower absolute value correlations than does another model, and if this difference is statistically significant, we have evidence that this model is a superior explanation of the return-generating process.[9] We also examine the entire distribution of the absolute value residual correlations between funds for each model. If, in addition to having a lower average, a model has fewer large absolute value correlations for each of a number of preselected absolute values, we have further evidence of that model's superiority.

The last set of tests we perform to judge significance involves an examination of the time series equation relating the return on each fund to the returns on the indexes for each of our models. More specifically, in considering a specific index we examine if we have a larger number of funds than expected which have sensitivities to the new index that are different from zero at a statistically significant level. If the ranking on all these tests is the same, we have strong evidence

---

[8] There are a number of direct tests that a variance/covariance matrix is diagonal. These are in the factor analysis literature. The best known is the Bartlet test, which uses the square of the residual correlation to eliminate the sign problem rather than the absolute value of the residual correlation (the measure we use). However, even the proponents of these tests suggest they are at best a rough guide. See for example Lawley and Maxwell (1963).

[9] We also report the average residual correlation (as opposed to the average absolute value residual correlation) for each model. While we believe this to be a less relevant measure, the reader can verify that, except for one case when comparing results from factor models where alternative numbers of factors are extracted and utilized, this measure leads to the same ranking of models as does the absolute value measure.

that one model outperforms the other as a candidate for the return-generating process.

In what follows, we will use the average absolute value correlation of residuals, the distribution of absolute value correlations, and the significance of betas as tests for differentiating between alternative return-generating processes..

A. The Base Model

Table 1 and Table 2 present these results for a number of different models. Let's start by examining Panel A of Table 1. Panel A presents results based on the absolute values of the correlations between residuals for three models — a zero-index model (where the "residuals" are simply the excess returns, i.e., the fund returns minus the CRSP SBBI 30-day T-bill rate), a standard one-index model, and a four-index model (called the "base" model) that we have employed in previous research.[10]

The one-index model uses the excess return on the CRSP SBBI S&P 500 total return index as the single index. The base four-index model adds to the S&P 500 index (measured in excess-return form): (1) a bond market index (a par-weighted combination of the Lehman Brothers aggregate bond index and the Blume/Keim high-yield bond index in excess-return form); (2) a small-cap minus large-cap index (the average of the Prudential-Bache small-cap growth and value indexes minus the average of their large-cap growth and value indexes); (3) a growth minus value index (the average of the Prudential-Bache large-, mid- and small-cap growth indexes minus the average of their large-, mid- and small-cap value indexes).

[10] We examined different combinations of the indexes from our four-index model in two- and three-index versions and found that they did not perform as well as the four-index model. Hence, we do not report them.

The S&P 500 total return index and the small minus large index were selected because they have been shown to be related to security returns in a number of studies. Growth minus value was selected as the third index in our base model. Both growth minus value and market-to-book have been shown to be importantly related to security returns. However, because growth minus value is highly correlated with returns on portfolios separated by their market-to-book ratios, we included only one of the indexes. Shortly we will test both market-to-book and growth minus value to try to get insight into which is the more fundamental index. A bond index was included because many mutual funds that have "common stock" as their objective often hold bonds in their portfolios. While a bond index may or may not be important in explaining return patterns on common stocks, it is clearly important when bonds can be included in the portfolios under study. In a latter section we will show it is also important in explaining security returns.

We use the four-index model as our base model. We performed (but do not report) tests to see if the four-index model outperforms models using any combination of subsets of the indexes. The four-index model reduces residual correlation at a statistically significant level from a model employing any possible combination of these indexes taken two or three at a time. Later in this paper we consider alternative specifications for one of these indexes.

As shown in Panel A of Table 1, the mean absolute value of the residual correlations (as well as the average residual correlation) becomes smaller as we move from the zero-index to the one-index model to the base four-index model. The means are statistically different using a simple t test. Of more importance, each model with more indexes dominates the models with fewer indexes using first-order stochastic dominance. First order stochastic dominance examines the cumulative distribution. If, for each threshold value, one model always has a smaller number of absolute value residual correlations than does another model, first order stochastic dominance exists.[11] The logic is that under any weighting scheme the model with fewer large absolute value residual correlations will be superior as long as large residual correlations are weighted more heavily than small. This is a particularly powerful test, since it does not depend on the choice of a weighting scheme of correlations (e.g., squared correlations or average correlation).

Finally, examining Table 2 shows that the beta on the S&P index and the small minus large index is significant at the 5% level for almost all funds, while the sensitivities on the growth minus value index and bond index are significant for 74% and 47% of the funds respectively.

Since the four-index model outperforms the models with fewer indexes for each of our groups, the question remains as to whether we can find a fifth index which is important in the return-generating process. Before we turn to this question, let us examine the spread in sensitivities of the funds in our sample to each of the indexes in our model. As discussed in the introduction, a justification for using mutual fund data is a belief that we will get a substantial spread on sensitivities. To judge this, we need a comparison group. Since size is often used as a criterion for forming portfolios to test return-generating processes, we selected the size deciles from the monthly CRSP Stock Indices file as an alternative set of portfolios with which to judge the dispersion of sensitivities. Table 3 presents the standard deviation of sensitivities and the difference in the 20th and 80th deciles for our full sample of mutual funds as well as for the CRSP

[11] For reasons of space, we show only the distribution for a selected number of threshold values in Table 1.

deciles.[12] It is clear from this table that mutual funds not only show dispersion on the sensitivities to each index, but also (with one exception) they show more dispersion than the CRSP size deciles. The one exception is the sensitivity to the size index. This is logical, since the CRSP deciles were selected to maximize dispersion on this index. However, even here the mutual funds show a high degree of dispersion consistent with the dispersion of mutual fund sensitivities on other indexes. Mutual funds are presenting investors with alternative sensitivities to the indexes in our model, and mutual funds thus present a meaningful way to examine a return-generating process.[13]

B. A Fifth Index

In this section we explore whether a fifth index should be added to the base model. The first candidate we examine for a fifth index is derived from the data itself. For each group we performed a maximum-likelihood factor analysis on the residuals from the base (four-index) model and extracted the one-factor solution. For any group, this represents the best index that can be found for explaining the residual covariances for that group. However, the factor will pick up influences that may be unique to the group from which it is extracted, as well as more general systematic influences. To eliminate the effect of unique influences, the factor derived from group A was used to explain the correlation in group B, the factor from B to explain C, and the factor from C to explain A.

---

[12] We make comparisons using standard deviation and percentiles, since these measures are not affected by sample size and the sample sizes are different for the groups being compared.

[13] Table 3 also shows that mutual funds have more dispersion with respect to a fifth index (a mutual fund growth index) which is introduced in the next section.

As shown in panel B of Table 1, when the fifth index is extracted via factor analysis, there is a very large improvement in both the average and the average absolute value of the residual correlation estimates. The results show stochastic dominance over the four-index model, and the difference in the average absolute values of the correlation coefficients is significant at the .01 level. In addition, when we examine Table 2, more than 71% of the funds have a significant sensitivity (at the 5% level) on the fifth factor.

The question remains as to whether we can find an alternative fifth index that works as well or better than the factor approximation and that has an economic meaning. We tried three approaches. One involved forming other portfolios of stocks. The second involved using a sentiment index derived from closed end funds. The third involved using data from mutual funds (as opposed to common stocks).

For the first of these approaches we formed portfolios of stocks that represented sectors of the economy. We examined a financial sector, a utility sector, a high tech sector, metals stocks, foreign stocks, and a natural resource sector. These were selected because of their inclusion in other studies. None of the five-index models using a sector factor outperformed the four-index base model at a significant level, and all were outperformed by the five-index model containing a factor (base model plus factor) at the .01 level of significance. In the interest of space, we do not show these results in the tables.

The second approach is based on a sentiment index developed by Lee, Schleifer and Thaler (1991). They have argued that the change in the discounts of closed end mutual funds reflects investor sentiment, and that this index systematically affects stock returns and should be priced. Since our time frame differs from theirs, we need to construct our own version of their index. We followed the exact procedure they described except that we include new funds six months after issuance and delete funds three months before they disappear, because of the well-documented effect on returns of these two events. We regressed the sentiment index on our base model, and used the residuals as our fifth index.[14] Examining both the value and the absolute value of the correlations in Table 1, Panel B shows that sentiment does not add to the explanation of the correlation when added as a fifth index. As shown in Table 2, when we examine the number of significant betas for our fund sample, we observe only a few more significant betas than one would expect by chance. Sentiment does not seem to be a reasonable fifth index.

Having failed to find a sector index or a sentiment index which performed adequately, we decided to examine an index which represented mutual fund returns themselves rather than security returns. We selected the Morningstar growth mutual fund index (an equally weighted index of the funds Morningstar classifies as growth) as our fifth index. This growth index was selected because this is a large category of funds and because residuals from the four-index model are smaller for income funds than they are for growth and aggressive growth funds. We reformulated the Morningstar index by regressing it against the four-index base model and used the residuals as our mutual fund growth index, which we refer to as MGO.

The first fact to notice from Panel B of Table 1 is that the introduction of the mutual fund growth index (MGO) results in a model which outperforms the four-index base model (the difference in average absolute values is statistically significant level at the .01 level). Furthermore,

[14] Sentiment was related to size and value growth. Sentiment may be an explanation for either of these indexes entering. In this paper we are testing whether it has any independent influence.

from Table 2, 72% of the betas are significantly different from zero. When we compare the five-index model using the mutual fund growth index to the model using the factor as the fifth index, we find that the results are virtually indistinguishable.[15] No technique shows stochastic dominance over the other in any of the three samples, the mean differences in absolute values of correlation coefficients are not statistically different, and the number of betas that are significant is identical. Furthermore, the mutual fund growth index is highly correlated with the factor extracted by factor analysis. The correlation between the factor scores and the growth index is .86, .88 and .82 for the three samples we examined. Since a mutual fund growth index improves results and is economically identifiable, it is worthwhile to try to understand why it enters and what its relationship is to the growth-minus-value index in our base four-index model.

The base four-index model includes the difference between a growth and a value index as one of the indexes. One possibility for the improvement of adding the mutual fund growth index is that when we use the difference in growth and value we are implicitly assuming they are equally important. Perhaps the five-index model leads to improvement because an unequal combination of the Prudential Bache growth and value indexes represents the factor affecting returns. This can be tested by adding either the Prudential-Bache value or growth index to our base four-index model. When we did this, the results were not improved over the base four-index model. A second possibility is the base model could be improved by a better formulation of the growth-value variable.

---

[15] It would not be surprising to find the factor related to an equally weighted index of all mutual finds, given the nature of factor analysis. However, it is surprising to find that the factor is so highly related to an equally weighted combination of a subset of the funds which constitutes about 40% of our sample and which has particular characteristics

There are two generic types of value and growth indexes. One type classifies firms into portfolios on the basis of high or low values on a multiple set of characteristics such as earnings price ratios, forecasted earnings growth, dividend yields, etc. (e.g., Prudential Bache and Wilshire use this approach). The second type uses a single variable, market-to-book ratio, to divide firms into portfolios (e.g., Barra, Fama and French and Russell use this approach). All of the indexes mentioned above are formulated as the return on a portfolio of stocks which have high (or low) values on the firm characteristics specified above. Which type performs better is tested in Panel C of Table 1, where we compare the residual correlations using Pru-Bache indexes with those of two indexes that classify by market-to-book ratio (Barra and Russell).[16] The four-index model using the Pru-Bache indexes results in a smaller average absolute value of residual correlations (and average correlations) than do the four-index models using market-to-book ratios. These differences are statistically significant using a simple t test and using stochastic dominance tests. Thus the Prudential Bache multi-criteria growth index performs better than one that classifies firms solely by their market-to-book ratios.

Why else could adding an index of growth funds improve the ability to explain the correlation structure? There are two additional explanations. First, the index representing growth in the base model could be imperfectly measured. Potentially, both market-to-book and growth-value may be proxying for a more fundamental factor. Second, many mutual funds have a large

---

different from the other funds.

[16] In each alternative four-index model, we use the difference between the returns on two portfolios as a fourth index. Barra separates the firms in the S&P 500 into two groups (growth and value) based on their market-to-book ratios. For the Russell growth-value index, we use the Russell 1000 value and growth indexes; these are constructed by splitting the largest 1000 stocks into two groups by market-to-book ratios.

number of common holdings, and it is possible that we are simply picking up a factor that captures common holdings. We now turn to an examination of whether either of these explanations might be correct. We start by examining the impact of common holdings on residual correlations.

C. Estimating the Effect of Common Holdings

At this point, the question remains as to how much of the correlation between mutual fund returns is due to common holdings as opposed to systematic factors affecting stock prices. Many mutual funds tend to hold the same stocks, and obviously the correlation between the parts of their holdings which are in common is perfect (equal to one). Unless factors in the return-generating process capture the returns on a portfolio of these common holdings, common holdings will explain part of the residual correlation between funds.

To see this, assume for the moment that the residual variance for each stock i ($Var(e_i)$) is the same, i.e., $Var(e_i) = CVAR$ for all i, and that the residual covariance between each pair of stocks i and j ($Cov(e_i e_j)$) is the same, i.e., $Cov(e_i e_j) = CCOV$ for all i and j. Then the covariance of the residual returns between two funds A and B ($Cov(e_A e_B)$) can be represented as:

$$Cov(e_A e_B) = \sum_{i \in S} X_{Ai} X_{Bi} CVAR + \sum_{i} \sum_{\substack{j \\ j \neq i}} X_{Ai} X_{Bj} CCOV \qquad (1)$$

where S is the set of stocks held in common and X_{Ai} represents the proportion invested in stock i by fund A.

Recognizing that the set notation can be dropped since $X_{Ai}\ X_{Bi} = 0$ for i not in S and rearranging equation (1) by combining summations and adding and subtracting $\sum_i X_{Ai} X_{Bi} CCOV$

yields:

$$Cov(e_A e_B) = \sum_i X_{Ai} X_{Bi}[CVAR - CCOV] + \sum_i \sum_j X_{Ai} X_{Bj} CCOV \quad (2)$$

The first term in the right-hand side of equation (2) represents the marginal impact of common holdings on the residual covariance of a pair of funds. Now funds might not be 100 percent invested in common stocks. Equations (1) and (2) embody the assumption that all residual fund covariance comes from holding stocks.[17] We can scale equation (2) by the percentage of stock held by each fund:

$$\frac{Cov(e_A e_B)}{\sum_i X_{Ai} \sum_j X_{Bj}} = CCOV + [CVAR - CCOV] \times \left[\frac{\sum_i X_{Ai} X_{Bi}}{\sum_i X_{Ai} \sum_j X_{Bj}} \right] \quad (3)$$

Equation (3) can then be estimated as a cross-sectional linear regression model:

$$\frac{Cov(e_A e_B)}{\sum_i X_{Ai} \sum_j X_{Bj}} = \gamma_0 + \gamma_1 \left[\frac{\sum_i X_{Ai} X_{Bi}}{\sum_i X_{Ai} \sum_j X_{Bj}} \right] + \eta_{AB} \quad (4)$$

where η_{AB} is a random error term.

Estimates of the common residual covariance between stocks (*CCOV*) and the stocks' common residual variance (*CVAR*) from any model can be computed directly from the regression estimates of γ_0 and γ_1 in equation (4).

---

[17] While this assumption is not strictly true, it should be a very good approximation. As shown in Blake, Elton and Gruber (1993), once the effect of a general bond return index is removed, correlations between the residuals of bond portfolios are small, and we have removed a bond index effect here. In addition, the variances of the residuals for bond portfolios after a bond index effect is removed are quite small relative to those for stock portfolios.

The above analysis assumes that the residual covariance between each pair of stocks is the same and that the residual variance of all stocks is the same. However, we might expect the residual variance and covariance for stocks held by aggressive growth funds to be different from the residual variance and covariance for stocks held by income funds. We divided our sample funds into three types — aggressive growth (AG), growth (G), and income (I) — and allowed the estimates of stock residual risk and covariance to be different for each type of fund.

Equation (4) can be used to estimate the residual variance of the returns on stocks and the residual covariance between the stocks in any fund type or any two fund types. To estimate this equation, we need to know the composition of the portfolio of each fund to which it is fit. Because of the difficulty of obtaining composition data, we obtained data only as of one date, December 1992, and for the funds in only one of our subsamples (group A). To the extent that the percentages of stocks held in common on this date were not representative of the whole of our sample period, this should bias the results against finding that common holdings help explain the covariance between funds. The source of our composition data was Morningstar. Group A contained 89 funds. Eleven of these funds were eliminated because we could not obtain composition data, or because their investment policy was outside the three groups described above, or because the data on fund composition contained inconsistencies or was incomplete. We call this restructured version of group A "group D."

Before we turn to our analysis using equation (4), let us examine the amounts of common holdings in group D. Table 4 shows the distribution of common holdings within and between types of funds with different policies. The median common holdings range from 1.3% to 5.6%, depending on the sample. For four of the samples, the 75th percentile is greater than 5%. This is substantial, given the residual correlations we are observing.

Table 5 shows the estimates of covariances and variances of individual securities for each fund type and across fund types obtained from the regression employing equation (4). The estimates in Panel A using the residuals from our base four-index model are consistent with what we expect to find and, in most cases, are statistically significantly different from zero.[18] The estimated variances of the residuals is highest for the stocks held by aggressive growth funds, next highest for stocks held by growth funds and much lower for stocks held by income funds. The estimated residual variances of stocks held by two different types of funds show a similar pattern. E.g., the stocks held by both aggressive growth and growth funds have a higher estimated residual variance. The ordering of the residual risk seems rational, as we would expect aggressive growth funds to have stocks with higher residual risk than those held by growth funds and we expect the lowest residual variance from those held by income funds. The covariances estimated in Panel A of Table 5 also fit a reasonable pattern, with the highest covariance found between stocks held by aggressive growth funds and the next highest between stocks held by both aggressive growth and growth funds. Examining the estimates for the five-index MGO model shown in Panel B, we find broadly similar results with one exception being that the estimated covariances are not statistically

[18] As a check on the regression estimate, we examined equation (3) only for those few funds which had common holdings close to zero. For those funds, within any fund type or pair of fund types, the average scaled residual covariance between the funds should be equal to the residual covariance between stocks held by the funds. The results from this procedure produced estimates of residual covariances consistent with those obtained from the regression procedure described above. For example, for the four-index model the estimate from pairs of funds of type AG was .6464 for the regression and .5547 for zero holdings; for pairs of types AG and G it was .2535 for the regression and .2676 for zero holdings; for pairs of types AG and I it was .1196 for the regression and .1277 for zero holdings. This consistency gives us additional confidence in the regression estimates.

different from zero. We shall discuss the implication of this in a later section.

D. Is It Common Holdings or Another Factor?

The analysis in section B indicated that the MGO index contributed to explaining the covariance between securities. In this section, we present further evidence on why the MGO index is important. The question we shall address is to what extent is the MGO index picking up common holdings and to what extent is it an independent influence. We present two approaches to examining the effect of common holdings. The first examines the relationship of the residual covariance between funds predicted by both the MGO index and common holdings. The second looks at the absolute value of the correlation between residuals from a four-index and a five-index model with and without removing the estimated effect of common holdings.

The residual covariance between two funds A and B due to the MGO index is:

$$\beta_{A,\text{MGO}}\, \beta_{B,\text{MGO}}\, Var(\text{MGO}) \tag{5}$$

We can correlate this estimate of the residual covariance due to the fifth index (MGO) with the residual covariance due to common holdings after four indexes have been removed to see how closely correlated they are. To do this we return to equation (2). Using estimates of variance and covariance of security residuals between stocks from Panel A of Table 5, along with actual common holdings, we calculate the residual covariance between any two funds due to common holdings (using the first term in the right-hand side of equation (2)). We then calculate the estimate due to MGO from equation (5). We then compute the correlation between these two estimates across all pairwise combinations of funds. When we do this we see that the adjusted R^2 between residual correlation due to the MGO index and that explained by common holdings is

.22. Thus, the MGO Index comes in, in part, because it picks up the effect of common holdings.[19] However, the MGO Index may contain independent information and common holdings may contain independent information. To test this we must examine the impact of common holdings on the residual correlations from the four-index and five-index models.

In Table 6 we present the distribution of absolute values of residual correlation coefficients for different return-generating processes with and without an adjustment for common holdings. Let's start by examining the residual correlation distribution from the four-index model shown in Panel A of Table 6. The unadjusted distribution is shown in the column labeled "Base," and the adjusted distribution is shown in the column labeled "Regression Estimate." Adjusting for common holdings lowers the average absolute value correlation from .130 to .119. Not only is the difference statistically significant, but the adjustment shows stochastic dominance across the cells in the table. Common holdings have a major impact on reducing the absolute values of correlations of residuals from the four-index model.

When we examine the residual correlation from a five-index model (adding MGO) without any adjustment for common holdings (shown in the column labeled "MGO" in Panel B of Table 6), we find that the average absolute value of the residual correlations is reduced below that found for the four-index model with adjustment for common holdings. The results are statistically significant at the .01 level and show stochastic dominance. Clearly, then, the MGO index introduces information not captured by common holdings. The question remains as to whether common holdings add additional information when the five-index model is used. The estimate of

---

[19] Similar results are obtained when the fifth index is constructed from a one-factor factor analysis of the residuals from the four-index model. In this case the R^2 is .13 rather than .22.

the funds' residual covariance due to common holdings is subtracted from the funds' actual residual covariance to adjust for the effect of common holdings.[20] Examining the column labeled "Regression Estimate" in Panel B of Table 6, we see that employing the correction for common holdings reduces the absolute correlation between residuals. While the difference is statistically significant, the distribution does not show stochastic dominance. However, using the correction for common holdings reduces the largest absolute values of residual correlations.

These results are consistent with those presented earlier in this section. When a four-index model is used, an adjustment for common holdings improves results markedly. When a fifth index is introduced, adjustment for common holdings reduces the residual correlation only a small amount. The introduction of an index based on actual mutual fund performance captures a large part (but not all) of the improvement introduced by explicitly taking common holdings into effect. The fifth index partially accounts for common holdings and partially adds a different influence.

E. Are Five Factors Enough?

In this section we present evidence which suggests that our five-factor fundamental model is a sufficient description of the return-generating process. Our evidence consists of four parts: 1) an examination of cross-sectional estimates of covariance; 2) an examination of how much of the remaining covariance is due to common holdings; 3) forecasts of correlations based on the sufficiency of the model; and 4) a comparison of the fundamental model with multi-index models

---

[20] The effect of common holdings is estimated by using the first term in the right-hand side of equation (2) along with estimates of residual variance and covariance of securities from panel B of Table 5 and the percentages of stocks in common. This estimate of the funds' residual covariance due to common holdings is subtracted from the funds'

based on factor analysis.

Covariance between common stock residuals comes about because of randomness and because of omitted factors. If the model captured all common influences, then the covariances of the residuals between stocks would average zero, covariances between stock residuals due to randomness would tend to cancel out in a portfolio, and the only reason funds would have a residual covariance would be common holdings. When we examine Table 5, Panel B, all estimates of the residual covariance estimates between securities for the five-index model are not statistically different from zero except for two cases: the aggressive growth and the aggressive growth and income samples. In these two cases they are significant, but with opposite signs. This is a strong indication that five indexes capture all common influences.

A second way to see if five factors are enough is to compare the unexplained covariance between funds assuming zero covariances between *securities* (all common influences are captured) with one that assumes they are non-zero. We calculated two estimates of covariance, one assuming the estimates shown in Panel B of Table 5, and one assuming that the best estimate of the covariance between securities is zero.[21] The results are shown in Panel B of Table 6 under the headings "Regression Estimate" and "Zero-Cov Estimate." The estimates from the five-index growth model are more accurate when zero covariance between security residuals is assumed. The results show stochastic dominance and are significant at the .01 level. Thus, assuming the covariance between securities is zero and that all common influences are captured by the five-

---

actual residual covariance to adjust for the effect of common holdings.

[21] There is still covariance between funds due to common holdings, and this is taken into account.

index growth model leads to the better estimates.

Another way to look at the data in Table 5 is to ask how much of the forecasted residual covariance is due to common holdings and how much is due to any remaining covariance between individual securities. With the five-index model, on average the proportion of the forecasted residual covariance due to common holdings is 60 times that due to the residual covariance between securities. For the group with the largest estimate for the residual covariance between securities, aggressive growth with other aggressive growth, the proportion of the estimated residual covariance due to common holdings is 80%. This is strong evidence that five indexes are enough to capture common influences.

A final way to examine if we have captured the important common factors is to compare our results with those from a purely statistical extraction of factors. We performed a maximum-likelihood factor analysis on the variance-covariance matrix of the funds' excess returns for each of our three samples. We extracted one to eight factors. If these are truly common, then the factors extracted from one sample should explain the structure of returns for a different sample. Thus sample A factors were used as indexes in a model for sample B, sample B for C and C for A. Table 7 reports the distribution of the absolute values of residual correlations when we used as indexes the factors derived from the maximum-likelihood factor analysis. The average absolute values of the residual correlations decrease as we move from the one-factor model to the eight-factor model. However, there is no longer stochastic dominance as we move from the four- to five-factor model for sample A and from the five- to six-factor model for sample C. This evidence is consistent with the presence of four, or possibly five, statistical factors.

If we compare our prespecified factors to the statistical factors, we see that our five prespecified factors do slightly better than the statistically extracted four-factor solution although the differences are not statistically significant.[22]

The average absolute value residual correlation from our five-factor prespecified model (base plus MGO) is the same or less than that of the four-factor model from the factor analysis. Furthermore, when we regress each of the four factors from the statistical factor model on the five prespecified factors, the four statistical factors are highly related to our five prespecified factors (see Table 8).[23] In addition, all of the five prespecified factors are significantly related to the statistical factors at the .01 level. Since four statistical factors capture the bulk of the residual covariances between funds, and since our five prespecified indexes perform as well or better, the five prespecified indexes seem to capture all common influences.

In this section we have presented evidence that four factors (or at most five) derived from maximum-likelihood factor analysis seem to capture the covariance between securities. In addition, our five-index fundamental model (base plus MGO) does at least as good a job of explaining covariances as does the four-index factor model. The performance is close, which is not surprising since each of the four factors is highly correlated with the five indexes in our fundamental model. When the base plus growth five-index model is used, the estimate of the residual covariance between securities for most groups is not statistically significantly different from zero. Almost all of the estimated residual covariance between funds comes from common

[22] The comparison is biased in favor of the statistical model because we fit and test the model in the same time period.

[23] The same results hold if we reverse the process. Each of the five fundamental factors is highly related to the

holdings, and assuming that covariances between security residuals from the five- index model are all zero produces better estimates than assuming that they are at their estimated levels.

III. TESTS ON PASSIVE PORTFOLIOS

We have established that a fifth index which is based on mutual fund data, whether developed from a factor analysis of the residuals from the four-index model or from an index of growth mutual funds, adds to the explanation of both returns and the correlation structure of returns when examining mutual funds. We have also seen that part of the explanatory power of the fifth index is due to its relationship with common holdings between mutual funds.

The question remains as to whether we have developed an index which is only relevant to mutual fund returns, perhaps because of some attribute of common holdings or dynamic behavior, or we have discovered a more fundamental factor in the general return-generating process.

To test this we now test the models we have been examining on passive portfolios of common stocks. We selected a random sample of 483 stocks from the CRSP tape and then divided this sample into 28 portfolios by using two-digit SIC codes. The design of the sample means that it can act as a holdout sample, that active management cannot influence the results, and that all firms are only in one SIC classification and thus the results cannot be impacted by common holdings.

In Table 9 we present the percentages of statistically significant regression coefficients across all 28 portfolios for each model. In Table 10 we present the distribution of the absolute residual

---

four statistical factors.

correlations for our base four-index model and three five-index models. Notice that the five-index model using the sentiment index underperforms the base four-index model, whether one uses the average correlation coefficient, the average absolute value of the correlation coefficient, or the stochastic dominance of absolute coefficients as a criterion.

When we examine the two candidates for the fifth index that were developed from mutual fund data, we find a large improvement in results compared to the four-index model. Adding either the factor index or the mutual fund growth index results in a decrease in the average correlation coefficient and in the average absolute correlation coefficient, and both models exhibit distributions of the absolute correlation coefficient which dominate the distribution for the four-index model. Furthermore, performing a *t* test on the pairwise differences in absolute correlation coefficients shows that both models outperform the four-index model at the 5% level of significance.

From Table 9, which reports the percentages of betas on each index that are significantly different from zero in the time series regression of the return for each portfolio on the indexes (estimates of the return-generating process), we see analogous results. The sentiment index is significant 7.14% of the time. We would expect it to be significant 5% of the time at the .05 level of significance being used. On the other hand, the factor index is significant 42.9% of the time and the mutual fund growth index is significant 35.7% of the time, much higher values than we would expect on the basis of chance.

The two indexes developed using mutual fund data have applicability well beyond the study of mutual funds.

IV. IMPLICATIONS OF COMMON HOLDINGS FOR INVESTOR BEHAVIOR

One of the implications of this paper is that common holdings are an important source of risk for investors. Mutual funds add stocks on the basis of their own analysts' recommendations or upon the recommendations of outside analysts. Internal analysts will be influenced by outside analysts. Thus the recommendations concerning which stocks to purchase have a great deal of commonality across funds. This should lead to stocks being held in common, with more common holdings across funds with similar objectives. Investors, believing they are getting diversification by buying a portfolio of mutual funds, will get less than anticipated because of common holdings. Even a very sophisticated investor who examines a fund's loadings on factors (indexes) will get less diversification than expected because of common holdings.

V. CONCLUSION

In this paper we have examined several alternative models of the return-generating process. We have chosen to test the models on mutual fund data because these data lead to a natural differentiation on important influences while damping out random influences and because the indexes uncovered by this research have a natural advantage in measuring mutual fund performance. The research reveals that:

1. A four-index model based on passive portfolios of common stocks with different characteristics explains a great deal of the correlation between mutual funds. All of the four indexes have been used in some form in previous papers, although the particular form we use has not previously been tested against alternative specifications.
2. A value-growth index based on firm fundamentals is better in explaining covariance than an

index based on market-to-book values.

3. Using factor analysis, there is very strong evidence that, after removing the effects from four indexes, a fifth index (a factor constructed from the residuals) has strong explanatory power. This is true whether the importance of the fifth index is judged by examining the return-generating process itself or by examining its ability to explain the correlation of the residuals from a four-index model.
4. A fifth index representing an equally weighted portfolio of mutual funds with growth as a stated objective performs almost identically with the factor index mentioned above. This is due in part to the fact that the correlation between the two models is extremely high. It also means that we can identify the fifth factor.
5. Some of the residual correlation between mutual funds is due to common holdings. We have presented a methodology for removing this influence. This in itself is important for studying the effect of forming portfolios of mutual funds.
6. We find that part of the performance of our fifth index (whether formed by factor analysis or a portfolio of growth mutual funds) is accounted for by common holdings, but part is due to capturing a unique influence.
7. Finally, we test our five-index model on passive portfolios of common stocks. this serves several purposes. It allows us to test the performance of the fifth index where its importance is not obfuscated by common holdings. It acts as a holdout sample. It allows us to test the fifth index on a sample where the results cannot be caused by some element of active management. The strength of the fifth index is again demonstrated.

BIBLIOGRAPHY

Berry, Michael, Burmeister, Edwin and McElroy, Marjorie, 1988, "Sorting Out Risks Using Known APT Factors," *Financial Analysis Journal,* 44:2, 29-42.

Blake, Christopher R., Elton, Edwin J., and Gruber, Martin J., 1993, "The Performance of Bond Mutual Funds," *Journal of Business,* 66:3, 371-403.

Burmeister, Edwin and McElroy, Marjorie, 1987, "APT and Multifactor Asset Pricing Models with Measured and Unobserved Factors: Theoretical and Econometric Issues," Discussion Paper, Department of Economics, University of Virginia and Duke University.

Burmeister, Edwin and Wall, Kent, 1986, "The Arbitrage Pricing Theory and Macroeconomic Factor Measures," *The Financial Review,* 21:1, 1-20.

Chen, Nai-fu, Roll, Richard and Ross, Stephen, 1986, "Economic Forces and the Stock Market," *Journal of Business,* 59:3, 383-404.

Cho, D. Chinhyung, Elton, Edwin J., and Gruber, Martin J., 1984, "On the Robustness of the Roll and Ross Arbitrage Pricing Theory," *Journal of Financial and Quantitative Analysis,* 19:1, 1-10.

Cochrane, John, 1996, "A Cross-Sectional Test of a Production-Based Asset Pricing Model," *Journal of Political Economy*.

Connor, Gregory, and Korajczyk, Robert A., 1986, "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15:3, 373-394.

Dhrymes, Pheobus J., Friend, Irwin, and Gultekin, N. Bulent, 1984, "A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Theory," *The Journal of Finance*, 39:2, 323-346.

Elton, Edwin J. and Gruber, Martin J, 1994, "Multi-Index Models Using Simultaneous Estimation of All Parameters." In *A Practitioner's Guide to Factor Models*, The Research Foundation of the Institute of Chartered Financial Analysts, Charlottesville, VA, 31-58.

Elton, Edwin J. and Gruber, Martin J., 1992, "Portfolio Analysis With a Non-Normal Multi-Index Return-Generating Process," *Review of Quantitative Finance and Accounting*, 2:1, 5-17.

Elton, Edwin J., Gruber, Martin J., and Blake, Christopher R., 1996a, "The Persistence of Risk-Adjusted Mutual Fund Performance," *Journal of Business*, 69:2, 133-157.

Elton, Edwin J., Gruber, Martin J., and Blake, Christopher R., 1996b, "Survivorship Bias and Mutual Fund Performance," *Review of Financial Studies*, 9:4, 1097-1120.

Elton, Edwin, Gruber, M., Das, S., and Hlavka, M., 1993, "Efficiency With Costly Information: A Reinterpretation of Evidence for Managed Portfolios," *Review of Financial Studies*, 6:1, 1-22.

Elton, Edwin J., Gruber, Martin J., and Rentzler, Joel, 1983, "The Arbitrage Pricing Model and Returns on Assets Under Uncertain Inflation, *The Journal of Finance*, 38:2, 525-538.

Fama, Eugene, 1993, "Multifactor Portfolio Efficiency and Multifactor Asset Pricing," University of Chicago, working paper.

Fama, Eugene and French, Kenneth, 1993, "Common Risk Factors in the Returns on Bonds and Stocks," *Journal of Financial Economics*, 33 (February), 3-53.

Fama, Eugene and French, Kenneth, 1992, "The Cross Section of Expected Stock Returns," *The Journal of Finance*, 47:2, 427-466.

Ferson, Wayne and Schadt, Rudi, 1996, "Measuring Fund Strategy and Performance in Changing Economic Conditions," forthcoming, *The Journal of Finance*.

Hansen, Lars P. and Jagannathan, Ravi, 1991, "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225-262.

Jensen, Michael, 1966, "Risk, the Pricing of Capital Assets and the Evaluation of Investment Portfolios," *Journal of Business*, 42:2, 167-247.

Lawley, D.N. and Maxwell, M.A., 1963, *Factor Analysis as a Statistical Method.* Butterworths, London.

Lee, Charles, Schleifer, Andrei & Thaler, Richard, "Investor Sentiment and the Closed-End Fund Puzzle," *Journal of Finance* 1991, v. 46, 75-110.

Lehmann, Bruce and Modest, David, 1988, "The Empirical Foundations of the Arbitrage Pricing Theory." *Journal of Financial Economics,* 21:2, 213-254.

Roll, R. And Ross, S. A., 1980, "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance*, 35:5, 1073-1103.

TABLE 1

COMPARISON OF ALTERNATIVE RETURN-GENERATING PROCESSES

PANEL A

	Zero-Index Model			&P One-Index Model			Base Four-Index Model		
Threshold Value	A	B	C	A	B	C	A	B	C
0.010	3916	3916	3916	3818	3827	3812	3720	3705	3707
0.025	3916	3916	3916	3665	3681	3653	3406	3419	3400
0.050	3916	3916	3916	3394	3432	3394	2902	2948	2887
0.100	3916	3916	3916	2872	2983	2908	1996	2034	2075
0.150	3916	3916	3916	2348	2545	2388	1279	1285	1318
0.200	3916	3915	3916	1915	2141	2008	812	765	802
0.250	3916	3909	3916	1528	1748	1630	454	404	455
Average Abs. Val. Correlation	0.814	0.831	0.840	0.228	0.244	0.235	0.125	0.123	0.126
Average Correlation	0.814	0.831	0.840	0.193	0.218	0.213	0.065	0.078	0.082

PANEL B
(alternative five-index models; base model plus index shown)

	Factor			Sentiment			MGO		
Threshold Value	A	B	C	A	B	C	A	B	C
0.010	3690	3676	3665	3719	3732	3701	3676	3680	3672
0.025	3339	3302	3292	3406	3433	3384	3300	3293	3333
0.050	2741	2687	2744	2909	2961	2914	2771	2682	2746
0.100	1767	1700	1749	1995	2026	2079	1750	1682	1734
0.150	1052	979	948	1271	1286	1311	1033	967	960
0.200	529	507	475	804	765	799	540	493	488
0.250	261	235	217	455	398	460	262	243	225
Average Abs. Val. Correlation	0.107	0.104	0.103	0.125	0.123	0.126	0.107	0.103	0.104
Average Correlation	0.023	0.026	0.003	0.065	0.078	0.081	0.014	0.017	0.019

PANEL C
(alternative four-index models, using alternative growth minus value indexes)

	Base			Barra			Russell		
Threshold Value	A	B	C	A	B	C	A	B	C
0.010	3720	3705	3707	3707	3744	3732	3703	3732	3713
0.025	3406	3419	3400	3415	3448	3465	3382	3425	3424
0.050	2902	2948	2887	2947	3004	3024	2871	2959	2969
0.100	1996	2034	2075	2110	2150	2219	2037	2059	2132
0.150	1279	1285	1318	1445	1430	1576	1375	1359	1439
0.200	812	765	802	967	915	1054	881	822	907
0.250	454	404	455	606	554	660	533	479	526
Average Abs. Val. Correlation	0.125	0.123	0.126	0.137	0.136	0.144	0.129	0.128	0.133
Average Correlation	0.065	0.078	0.082	0.077	0.087	0.098	0.072	0.083	0.089

Notes: Table shows number of absolute values of pairwise residual correlations greater than threshold values for groups A, B and C (@ 89 funds).
Zero-index model uses unadjusted excess returns (over 30-day T-bill rate) of the sample funds; one-index model uses the excess return on the S&P 500 index; base four-index model uses the excess return of the S&P 500 index, the excess return of a composite bond index, the average of the Pru-Bache small-cap indexes minus the average of the Pru-Bache large-cap indexes, and the average of the Pru-Bache growth indexes minus the average of the Pru-Bache value indexes.
The alternative five-index models use the base four-index model plus an additional index as follows:
"factor" uses a single factor extracted from the residuals of the sample funds' excess returns regressed on the base four-index model, where group A uses a factor extracted from group C, group B uses a factor from group A, and group C uses a factor from group B;
"sentiment" uses the change in the discount of a value-weighted portfolio of closed-end funds, orthogonalized to the base four-index model;
"MGO" uses the the excess return of the Morningstar Growth fund index, orthogonalized to the base 4-index model.
The alternative four-index models use alternative growth-value indexes as follows:
"base" uses a growth minus value index obtained from Pru-Bache indexes as described above;
"barra" uses a growth minus value index obtained from BARRA/S&P indexes;
"russell" uses a growth minus value index obtained from Russell 1000 indexes.

TABLE 2

PERCENTAGE OF REGRESSION COEFFICIENTS SIGNIFICANT AT THE 5% LEVEL (267 FUNDS)

Model	S&P 500	Small-Large	Growth-Value	Bond	Fifth Index
Base	100.00%	87.27%	73.78%	46.82%	
Factor	100.00%	88.01%	75.28%	49.44%	71.54%
Sentiment	100.00%	87.27%	73.78%	46.44%	9.36%
MGO	100.00%	87.64%	74.91%	49.06%	71.54%

Notes: Table shows percentage of regression coefficients that are different from zero at 5% level for 267 funds when a time-series regression is run on the returns for each fund against the variables identified in the body of the table.

Base model is four-index model using the excess return of the S&P 500 index, the excess return of a composite bond index, the average of the Pru-Bache small-cap indexes minus the average of the Pru-Bache large-cap indexes, and the average of the Pru-Bache growth indexes minus the average of the Pru-Bache value indexes.

The alternative five-index models use the base four-index model plus an additional index as follows:
"factor" uses a single factor extracted from the residuals of the sample funds' excess returns regressed on the base four-index model, where group A uses a factor extracted from group C, group B uses a factor from group A, and group C uses a factor from group B;
"sentiment" uses the change in the discount of a value-weighted portfolio of closed-end funds, orthogonalized to the base four-index model;
"MGO" uses the the excess return of the Morningstar Growth fund index, orthogonalized to the base 4-index model.

TABLE 3

DISTRIBUTION OF BETAS FROM FIVE-INDEX GROWTH MODEL

index	from 267 sample funds					from CRSP deciles				
	XSP500	XAGGHY	PRUSMLG	PRUGRVL	MGO	XSP500	XAGGHY	PRUSMLG	PRUGRVL	MGO
20th percentile	0.685	0.017	0.094	-0.096	0.215	0.931	-0.057	0.253	-0.327	0.042
median	0.836	0.103	0.231	0.152	0.592	0.961	0.041	0.711	-0.082	0.714
80th percentile	0.936	0.221	0.392	0.480	1.081	0.973	0.118	0.999	-0.041	0.866
standard deviation	0.173	0.172	0.201	0.346	0.526	0.022	0.074	0.423	0.170	0.382

Notes: "XSP500" is the excess return (over the 30-day T-bill rate) of the S&P 500; "XAGGHY" is the excess return of a composite bond index; "PRUSMLG" is the average of the Pru-Bache small-cap indexes minus the average of the Pru-Bache large-cap indexes; "PRUGRVL" is the average of the Pru-Bache growth indexes minus the average of the Pru-Bache value indexes; "MGO" is the residual of the Morningstar Growth fund index obtained from the base 4-index model (which consists of the prior four indexes).

Betas are estimated via time-series regressions of the excess returns of either funds or stock deciles on a five-index model using the indexes shown.

TABLE 4

DISTRIBUTION OF PERCENTAGES OF COMMON HOLDINGS

Fund Types	proportion with 0%	median %	25th percentile	75th percentile	highest %
AG, AG	8.89%	2.85%	1.88%	5.28%	8.17%
G, G	5.71%	4.57%	3.13%	6.22%	12.02%
I, I	0.57%	5.59%	3.95%	6.73%	11.44%
AG, G	17.71%	2.71%	1.13%	4.53%	10.53%
AG, I	33.03%	1.34%	0.00%	2.99%	6.64%
G, I	4.24%	4.25%	2.89%	5.75%	16.12%

Note: "AG" = aggressive growth funds; "G" = long-term growth funds;
"I" = income, balanced, and growth and income funds.
The percentages shown are obtained from Morningstar composition data for group D (78 funds from group A).
For the last four columns, the percentages shown are based on the square root of the sum of the scaled products of the fractions of securities held in common between pairs of funds (i.e., the square root of the last term in equation (3) in the text).

TABLE 5

ESTIMATES OF VARIANCES AND COVARIANCES OF RESIDUALS FOR INDIVIDUAL SECURITIES FROM EQUATION 4

PANEL A
RESIDUALS OBTAINED FROM 4-INDEX BASE MODEL

Fund Types	Covariance (intercept)	t	Slope	t	Estimated Variance
AG, AG	0.6464	3.61	158.78	2.24	159.42
G, G	0.1160	2.76	99.98	8.01	100.10
I, I	0.1343	3.88	11.50	1.29	11.63
AG, G	0.2535	5.22	176.13	8.37	176.38
AG, I	0.1196	3.24	91.37	2.66	91.49
G, I	0.1843	8.18	13.30	1.75	13.48

PANEL B
RESIDUALS OBTAINED FROM 5-INDEX MGO (GROWTH) MODEL

Fund Types	Covariance (intercept)	t	Slope	t	Estimated Variance
AG, AG	0.3199	2.07	58.26	0.95	58.58
G, G	-0.0614	-1.77	35.43	3.44	35.37
I, I	0.0003	0.01	24.91	3.15	24.91
AG, G	-0.0332	-0.82	79.40	4.49	79.37
AG, I	-0.0991	-3.03	90.36	2.96	90.26
G, I	-0.0322	-1.67	29.34	4.50	29.31

Notes: "AG" = aggressive growth funds; "G" = long-term growth funds; "I" = income, balanced, and growth and income funds.

The estimated variance is obtained by adding the intercept to the slope.

Equation 4 is:

$$\frac{Cov(e_A e_B)}{\sum_i X_{Ai} \sum_j X_{Bj}} = \gamma_0 + \gamma_1 \left[\frac{\sum_i X_{Ai} X_{Bi}}{\sum_i X_{Ai} \sum_j X_{Bj}} \right] + \eta_{AB}$$

for any two funds A and B, where X is the weight that a fund has in a stock and e is the error term for a fund from a given model.

TABLE 6

THE EFFECT OF COMMON HOLDINGS

PANEL A

Base 4-index model without and with adjustment for common holdings

Threshold Value	Base	Regression Estimate
0.010	2858	2853
0.025	2614	2589
0.050	2254	2186
0.100	1590	1475
0.150	1076	942
0.200	692	549
0.250	391	288
Average Abs. Val. Correlation	0.130	0.119
Average Correlation	0.079	0.047

PANEL B

MGO 5-index model without and with alternative adjustments for common holdings

Threshold Value	MGO	Regression Estimate	Zero-Cov Estimate
0.010	2806	2808	2806
0.025	2522	2551	2547
0.050	2133	2140	2120
0.100	1379	1361	1347
0.150	833	822	810
0.200	444	414	399
0.250	218	199	195
Average Abs. Val. Correlation	0.110	0.108	0.107
Average Correlation	0.017	-0.013	-0.007

Notes: Table shows number of absolute values of pairwise residual correlations greater than threshold values for group D (78 funds from group A)

Base four-index model uses the excess return of the S&P 500 index, the excess return of a composite bond index, the average of the Pru-Bache small-cap indexes minus the average of the Pru-Bache large-cap indexes, and the average of the Pru-Bache growth indexes minus the average of the Pru-Bache value indexes.

MGO five-index model uses the base four-index model plus the the excess return of the Morningstar Growth fund index, orthogonalized to the base 4-index model.

TABLE 7

COMPARISON OF ALTERNATIVE FACTOR MODELS

Threshold Value	1-Factor A	1-Factor B	1-Factor C	2-Factor A	2-Factor B	2-Factor C	3-Factor A	3-Factor B	3-Factor C	4-Factor A	4-Factor B	4-Factor C
0.010	3784	3757	3747	3697	3719	3699	3685	3685	3688	3651	3674	3675
0.025	3566	3538	3540	3377	3372	3384	3359	3350	3319	3317	3278	3307
0.050	3206	3146	3154	2887	2830	2862	2819	2798	2766	2721	2712	2714
0.100	2454	2416	2419	2013	1860	1988	1917	1821	1827	1765	1672	1744
0.150	1805	1776	1798	1308	1141	1217	1196	1094	1089	993	973	1011
0.200	1297	1232	1271	769	679	676	686	601	568	532	498	504
0.250	902	832	874	416	347	338	362	309	261	256	237	232
Average Abs. Val. Correlation	0.168	0.160	0.162	0.124	0.115	0.117	0.117	0.112	0.110	0.107	0.104	0.105
Average Correlation	0.008	0.009	0.010	0.009	0.011	0.012	0.007	0.013	0.013	0.005	0.009	0.012

Threshold Value	5-Factor A	5-Factor B	5-Factor C	6-Factor A	6-Factor B	6-Factor C	7-Factor A	7-Factor B	7-Factor C	8-Factor A	8-Factor B	8-Factor C
0.010	3666	3628	3640	3680	3623	3632	3645	3604	3595	3625	3626	3600
0.025	3271	3227	3205	3292	3214	3223	3258	3185	3168	3243	3194	3140
0.050	2690	2591	2555	2665	2562	2529	2602	2510	2451	2540	2474	2465
0.100	1650	1491	1511	1602	1434	1506	1512	1362	1384	1431	1292	1353
0.150	904	781	774	809	707	735	767	641	623	697	593	605
0.200	477	356	335	372	323	325	324	254	236	296	234	227
0.250	228	145	144	154	134	128	124	102	79	107	87	76
Average Abs. Val. Correlation	0.102	0.093	0.092	0.096	0.090	0.091	0.092	0.086	0.085	0.089	0.084	0.084
Average Correlation	0.006	0.010	0.007	0.006	0.011	0.006	0.005	0.012	0.007	0.005	0.013	0.008

Notes: Table shows number of absolute values of pairwise residual correlations greater than threshold values for groups A, B and C (@ 89 funds). Factors are extracted from excess returns of each group. The factors are then used as indexes for time series regressions to obtain residual correlations. group A uses factors extracted from group C; group B uses factors from group A; group C uses factors from group B.

TABLE 8

ADJUSTED R2'S OF FACTORS (FROM FOUR-FACTOR SOLUTION) REGRESSED ON VARIABLES FROM THE BASE MODEL PLUS MGO

factor	group A	group B	group C
1	1.00	1.00	1.00
2	0.89	0.90	0.86
3	0.76	0.59	0.72
4	0.58	0.67	0.11

Note: The factors are obtained from the excess returns of the funds in each group.

TABLE 9

PERCENTAGE OF REGRESSION COEFFICIENTS SIGNIFICANT AT THE 5% LEVEL (28 SIC PORTFOLIOS OF STOCKS)

Model	S&P 500	Small-Large	Growth-Value	Bond	Fifth Index
Base	100.00%	92.86%	28.57%	32.14%	
Factor	100.00%	92.86%	28.57%	32.14%	42.86%
Sentiment	100.00%	92.86%	28.57%	28.57%	7.14%
MGO	100.00%	92.86%	28.57%	32.14%	35.71%

Notes: Table shows percentage of regression coefficients that are different from zero at 5% level for 28 SIC portfolios of stocks when a time-series regression is run on the returns for each fund against the variables identified in the body of the table.

Base model is four-index model using the excess return of the S&P 500 index, the excess return of a composite bond index, the average of the Pru-Bache small-cap indexes minus the average of the Pru-Bache large-cap indexes, and the average of the Pru-Bache growth indexes minus the average of the Pru-Bache value indexes.

The alternative five-index models use the base four-index model plus an additional index as follows:
"factor" uses a single factor extracted from the residuals of the sample funds' excess returns regressed on the base four-index model, where group A uses a factor extracted from group C, group B uses a factor from group A, and group C uses a factor from group B;
"sentiment" uses the change in the discount of a value-weighted portfolio of closed-end funds, orthogonalized to the base four-index model;
"MGO" uses the the excess return of the Morningstar Growth fund index, orthogonalized to the base 4-index model.

TABLE 10

COMPARISON OF FOUR-INDEX BASE MODEL AND ALTERNATIVE FIVE-INDEX MODELS FOR SIC GROUPS

Threshold Value	Base	Factor	Sentiment	MGO
0.010	360	354	360	360
0.025	331	327	329	329
0.050	285	277	283	279
0.100	205	198	206	198
0.150	137	136	141	133
0.200	88	83	87	84
0.250	40	37	41	36
Average Abs. Val. Correlation	0.131	0.128	0.132	0.128
Average Correlation	0.055	0.044	0.055	0.045

Notes: Table shows number of absolute values of pairwise residual correlations greater than threshold values for a sample of 28 stock portfolios, where each portfolio is an equally weighted portfolio of stocks grouped by two-digit SIC code.
The alternative five-index models use the base four-index model plus an additional index as follows:
"factor" uses a single factor extracted from the residuals of the funds in group A, where the residuals were obtained by regressing the excess returns of the funds in group A on the base four-index model;
"sentiment" uses the change in the discount of a value-weighted portfolio of closed-end funds, orthogonalized to the base four-index model;
"MGO" uses the the excess return of the Morningstar Growth fund index, orthogonalized to the base four-index model.