Title

A Comparison of Single and Multifactor Portfolio Performance Methodologies (formerly WP #13-83)

Authors
Chen, Nai-fu
Copeland, Thomas E.
Mayers, David

Publication Date
1987-02-01

#2-86

A COMPARISON OF SINGLE AND MULTIFACTOR PORTFOLIO PERFORMANCE METHODOLOGIES

September 1985
Revised February 1987

Nai-Fu Chen *

Thomas E. Copeland

David Mayers

* Graduate School of Business
University of Chicago, Chicago, IL
60637

Graduate School of Management
University of California, Los Angeles, CA
90024

Forthcoming JFQA

A Comparison of Single and Multifactor Portfolio Performance Methodologies

February 1987

Nai-Fu Chen
Thomas E. Copeland
David Mayers

Graduate School of Business
University of Chicago, Chicago IL 60637

Graduate School of Management
University of California at Los Angeles 90024

We benefited from comments and suggestions of Eugene Fama, Michael Hanssens, Richard Roll, Myron Scholes and George Tiao. Our thanks to Glenn Graves, Professor of Mathematical Methods, Graduate School of Management, UCLA, for lending us the use of his mathematical programming system. An earlier version of this paper appeared in Stock Market Regularities, Dimson, E., editor, Cambridge University Press, 1986.

Abstract

A comparison of single and multifactor portfolio performance methodologies using Value Line and size-ranked portfolios indicates that although both methodologies provide unbiased estimates of portfolio performance, there are systematic differences in the power of the two methodologies. The predictive power of the multifactor methodology is superior for well- diversified portfolios but inferior for less diversified portfolios.

1 Introduction

Efficient methods to measure the economic significance of new information arrival are of great interest to financial economists. In their seminal paper, Fama, Fisher, Jensen, and Roll (1969) used the residuals from the single factor market model to measure the unanticipated returns of securities:

$$R_{it} = \hat{a}_i + \hat{b}_i R_{mt} + \epsilon_{it}, \tag{1}$$

where R_{it} and R_{mt} are the returns of security i and a market index at time t, $\hat{a}_i$ and $\hat{b}_i$ are asset-specific time-stationary parameters, and ϵ_{it} is the market model residual for asset i at time t. Thus, the normal relation between the return of asset i and the return of the market is described by $\hat{a}_i + \hat{b}_i R_{mt}$, which is the conditional expected return of asset i given the realization of R_{mt}. The residual, ϵ_{it}, contains the conditional unanticipated return, including the effect of new information.

The advantage of using the conditional residuals comes from removing the unanticipated "systematic" term, $\hat{b}_i(R_{mt} - E(R_{mt}))$, from the residuals. Copeland and Mayers (1982) found that using the residuals from a future benchmark market model methodology (conditional expected returns) produced more powerful tests than using the residuals from the mean return (unconditional expected return) in their study of the information content of Value Line recommendations. Their findings differ from that of Brown and Warner's (1980) simulation study which indicated about equal power between the two choices. A likely reason for the difference is that more efficient market model parameter estimates were obtained in the Copeland-Mayers study using portfolio rates of return as opposed to individual security rates of return as in Brown-Warner.

In this study, we examine the question of whether efficiency is further improved using a multifactor market model for the conditional expected returns. The rationale behind this approach is derived from the empirical regularity that there is more than one source of common covariation among asset returns (see King(1966) or Roll and Ross (1980)). Our approach is quite apart from the equilibrium consideration of Ross' Arbitrage Pricing Theory, which was the focus of the Roll and Ross study. Here, we are merely trying to determine whether extracting more common covariation from the residuals improves the efficiency of measuring the economic impact of new information.

The use of market model residuals always provides unbiased estimates of the unanticipated returns provided that the model parameters are stationary. For example, if we use the equally-weighted stock market index in a single factor market model, we estimate $\hat{a}_i, \hat{b}_i$ in equation (1) in the benchmark period. If the model also holds in the test period with the same a_i and b_i, then the conditional residuals, ϵ_{it}, will be unbiased estimates of the unanticipated returns, regardless of whether the true underlying process is single or multi factor, and regardless of whether CAPM or APT or some other model is the correct asset pricing model. The result follows from

$$\begin{aligned} \epsilon_{it} &= R_{it} - E_t(R_{it}|\hat{a}_i, \hat{b}_i, R_{mt}) \\ &= R_{it} - [\hat{a}_i + \hat{b}_i R_{mt}]. \end{aligned} \tag{2}$$

Note that we need not make assumptions relating $E(R_{it})$ to any pricing model. The consistency of $\hat{\epsilon}_{it}$ follows from the stationarity assumptions about the market model parameters.

Similarly, a sufficient condition for the consistency of the residuals from a multifactor market model is that the parameters are stationary. If both the parameters for a single and a multiple factor model are stationary, then both models will give consistent estimates of the residuals, but their variances may be different.

If the multifactor market model is correct and stationary, and if the associated parameters are known, then we would expect that the multifactor model will purge more common covariation from the residuals. However, the multifactor model requires more parameter estimates. If one uses out-of-sample conditional forecast errors, it is not obvious that the multifactor market model will dominate the traditional single factor market model even if the parameters are time stationary. To facilitate the comparison, we use two data sets. The first is the Value Line data set as in Copeland and Mayers (1982) and the second is a set of five size-ranked portfolios. As all other aspects of our research methodology are held constant, the only difference in the observed residuals comes solely from the single and multifactor specification of the return generating process.

Section 2 of our study provides brief descriptions of the data and our rate of return calculations. Section 3 outlines our experiment and provides a comparison of the single and multifactor models using the future benchmark procedure. We find that the two models yield measures of abnormal returns that are similar. However, there is evidence of systematic differences between the mean square errors of the single and multifactor models. Section 4 summarizes the results.

2 Data Description and Rate of Return Calculations

Our Value Line data base was obtained from The Value Line Investment Survey (Weekly Summary of Advices and Index) which lists the Value Line rankings for the set of securities contained in their universe. [1] These rankings are entitled, "Probable Market Performance, Next 12 months," and consist of Roman numerals I through V assigned to each firm. [2] The announcements of these rankings may constitute information events. [3] Commencing with the November 26, 1965 Survey, Value Line performance rankings were obtained at intervals at least 26 weeks apart for a total of 24 holding periods for all firms that are also contained on the CRSP daily rate of return file. [4] Thus, our study covers a 12 year history. [5]

We also constructed 5 size-ranked portfolios for each of the 24 holding periods. Portfolio 1 represents the smallest quintile of firms on the CRSP daily tape and portfolio 5 the largest. Market values for all NYSE and AMEX listed companies were computed by multiplying the number of shares outstanding on the Value Line ranking date by the price per share at the end of the previous month. There is no reason to believe that firm size on a given Value Line ranking date constitutes an information event. However, size-ranked portfolios possess certain empirical regularities and may prove interesting for distinguishing between single and multifactor market model estimates of residual returns.

[1] Since April of 1965, Value Line has published performance predictions using their present ranking system. They rank stocks from 1 to 5 with 1 being the most favorable. Currently, they are the world's largest (based on number of subscriptions) published advisory service, employing over 200 people. Security rankings result from a complex filter rule which utilizes four criteria (1) the earnings and price rank of each security relative to all others, (2) a price momentum factor (3) year-to-year relative changes in quarterly earnings, and (4) an earnings "surprise" factor. Roughly 53% of the stocks are ranked 3rd, 18% are ranked 2nd and 4th, and 6% are ranked 1st or 5th.

[2] Value Line indicates in a pamphlet entitled Investing in Common Stocks that for rank I stocks, "Expect the best price performance relative to the other stocks covered in the survey." Similarly, for rank V one should expect the poorest.

[3] See Copeland and Mayers (1982) or Stickel (1985) for evidence concerning the information content of Value Line rankings.

[4] The CRSP (University of Chicago Center for Research in Security Prices) file contains daily returns for all New York and American Stock Exchange listed securities, since July of 1962.

[5] Exact ranking dates are in Copeland and Mayers (1982). The intervals are occasionally irregularly spaced because the early data was originally collected for a slightly different experiment. The intervals always contain at least 26 weeks. There are 3 occasions when the intervals are greater than 26 weeks.

The daily CRSP rate of return file was converted to a weekly file using a Friday close to Friday close return interval. [6] The rates of return are adjusted, by CRSP, for dividends, splits, etc.. A weekly equally weighted rate of return index, R_{mw}, of all CRSP listed securities was constructed as a single factor market index. Justification for this index is provided in Brown and Warner (1980).

We also constructed five weekly factor return indices for the multifactor model; using Chen's (1983) procedure which is outlined in the appendix. We form well diversified mimicking portfolios, one for each factor, which have high sensivity to the kth factor and zero sensitivity to all other factors. Chen's procedure first estimates factor loadings for ten factors using the Jöreskog asymptotic maximum liklihood procedure. Then five mimicking portfolios are formed. The weekly rates of return , $R_{1w}...R_{5w}$, on these five mimicking portfolios are our factor return indices. Our choice of five factors (rather than some larger number) is arbitrary, but the procedure guarantees that the factor loadings are not biased due to misspecification (so long as there are no more than ten factors). We estimated the mimicking portfolio returns separately for the first 12 holding periods and for the last 12 periods in order to allow for nonstationarity in the factor returns. [7] Test periods are defined as the 26 weeks following each Value Line recommendation date and benchmark periods as the 26 weeks following each test period. [8] We evaluate weekly rebalanced, equally weighted portfolios of Value Line securities, and equally weighted portfolios based on firm size. There are five portfolios in each set. Both sets are reconstructed at the beginning of each of the 24 test periods. For each

[6] Of some concern is the timing of Value Line activities and when the clients actually have the Value Line recommendation. The construction of our weekly rate of return file assumes investors buy or sell at the closing price on the date of the recommendation, which is always a Friday. Value Line staggers mailings so they will arrive on the recommendation date. The recommendations are actually printed over the weekend preceeding the ranking date. Thus, Value Line analysis is completed the week prior to the week that the recommendation arrives and recommendations are one week old when received.

[7] Mimicking portfolio weights were also estimated from the entire 12 year variance-covariance matrix. The results, which are reported in footnote 14, lead us to conjecture that nonstationarity may be a problem.

[8] Copeland and Mayers (1982) refer to this as the future benchmark technique. Picking a future benchmark period avoids selection bias problems associated with using historic benchmarks to evaluate managed portfolios. However, it has problems of its own. For example, future benchmarks may be biased if the manager's predictive ability extends into the benchmark period. These problems (as well as others, e.g. the effects of statistical dependencies on significance tests) are discussed in detail in Copeland and Mayers(1982).

portfolio ($p = 1, ..., 5$) the weekly raw rate of return is defined as

$$R_{pt} = ln(1 + \sum_{j=1}^{N_t} \frac{R_{jt}}{N_t}) \quad (3)$$

Here R_{jt} is the weekly rate of return for security j in the portfolio of interest during week t. N is subscripted by t to note the possibility of delisting of listing. [9] Table 1 gives the raw rates of return for each Value Line and size-ranked portfolio averaged across all 24 test periods and cumulated from week 1 of the test period to week 26. [10] Table 1 also provides the average number of securities per portfolio.

Table 1 – Average 26 Week Cumulative Raw Returns for 24 HoldingPeriods from November 26, 1965 to February 3, 1978; and Average Number of Securities per Portfolio

	Portfolio Number				
	1	2	3	4	5
Value Line	.0738	.0651	.0410	.0270	.0037
Avg. No.	95	273	521	270	91
Size Ranked	.0964	.0672	.0495	.0406	.0265
Avg. No.	406	406	406	406	406

Note that the 26 week average cumulative returns are positive for all 10 portfolios. Also, the difference between the gross returns on the largest and smallest size-ranked portfolios 6.99%, is roughly the same as the difference in returns between Value Line portfolios 1 and 5, 7.01%.[11]

[9]The amount of listing and delisting was minor. For a discussion of their possible impact on measurement of Value Line performance, see Copeland and Mayers (1982, pp. 27-28). In addition, the possibility of delisting explains why we formed portfolios before measuring abnormal returns rather than measuring individual security abnormal returns and then averaging them. It is impossible to obtain a future benchmark return for a delisted security. Hence, selection bias would be introduced if we were to purge delisted securities when using a procedure based on individual security benchmarks.

[10]Copeland and Mayers repeated all of their performance experiments in non-log form also. The results were practically identical to the log form. Thus, we do not replicate our experiment in the log form.

[11]Copeland and Mayers (1982) found the Value Line portfolio 1 and 5 firms, not significantly different in size.

3 Abnormal Performance Using the Market Model Methodology –

3.1 Description of Methodologies –

The specifics of the market model methodology are given in the following formulae. Benchmark rates of return are estimated for each week, w, of a given test period, p. The single factor market model benchmark return is

$$B1_{pw} = \hat{a}_p + \hat{b}_p R_{mw}, \quad w = 1, ...26. \tag{4}$$

The coefficients $\hat{a}_p$ and $\hat{b}_p$ are estimated from the simple market model regression,

$$R_{pt} = a_p + b_p R_{mt} + \epsilon_{pt}, \quad t = 27, ..., 52, \tag{5}$$

estimated over the future benchmark period, i.e., the 26 weeks following the test period. [12]

The five factor market model benchmark return is

$$B5_{pw} = \hat{\alpha}_p + \hat{b}_{1p} R_{1w} + \hat{b}_{2p} R_{2w} + \hat{b}_{3p} R_{3w} + \hat{b}_{4p} R_{4w} + \hat{b}_{5p} R_{5w} \tag{6}$$

where the intercept and slope terms are estimated from the multiple regression

$$R_{pt} = \alpha_p + b_{1p} R_{1t} + b_{2p} R_{2t} + b_{3p} R_{3t} + b_{4p} R_{4t} + b_{5p} R_{5t} + \eta_{pt} \tag{7}$$

again estimated over the future benchmark period.

For each Value Line portfolio and size-ranked portfolio, we then estimate the cumulative excess returns for period p and the average cumulative excess rates of return,

$$CR_p = \sum_{w=1}^{26} (R_{pw} - B_{pw}) \tag{8}$$

$$ACR = \frac{1}{24} \sum_{p=1}^{24} CR_p \tag{9}$$

using both benchmark models, and test for whether the cumulative excess rate of return performance is different from zero. The standard deviation for this test is the usual unbiased estimator calculated with the period-by-period cumulative test period excess returns. [13]

[12] Coefficients were also estimated over a 52 week benchmark period, following the test period. Results for the Value Line portfolios are reported in footnote 15.

[13] The formula is

$$SD(CR) = \sqrt{\frac{\sum (CR - \overline{CR})^2}{24 - 1}}$$

3.2 Abnormal Performance Results –

Table 2 and Figure 1 contain the abnormal performance results. The table presents the average cumulative residuals (ACR's) and t-statistics for the five Value Line portfolios (Panel A) and the five size-ranked portfolios (Panel B) for both the single-factor and the five-factor market models over the entire 24 period history. The figure contains plots of cumulative abnormal performance for weeks relative to the Value Line ranking date using both models.

The results for the single and multifactor market models are practically indistinguishable. For the Value Line portfolios (Panel A) the 3.05% ACR reported for the single-factor model for portfolio 5 is marginally significant at the 0.05 level. The 2.89% ACR reported for the five-factor model for portfolio 5 is marginally insignificant. [14] Moreover, the differences between the portfolio 1 and 5 ACR's are almost identical for the two market models; 3.38% for the single-factor model and 3.42% for the five-factor model. [15]

As expected, the results for the size-ranked portfolios (Panel B) show no statistically significant performance at all. The largest t-statistic is only -1.2051, and the difference in ACR's between the smallest and largest size portfolios is only 1.33%

[14] Recall that the mimicking portfolio returns for the five-factor model in Table 2 were estimated separately for the first 12 and the second 12 holding periods in order to allow for possibile non-stationarity. The results given below are estimated using an identical procedure except that the mimicking portfolio returns were estimated over the entire 12 year period, not allowing for non-stationarity.

Value Line Portfolio Residuals					
	1	2	3	4	5
ACR	0.0193	0.0218	0.0117	0.0084	-0.0105
AR	0.0007	0.0008	0.0004	0.0003	-0.0004
t-stat	1.4580	1.8696	1.1853	0.8438	-0.8528
Size-ranked Portfolio Residuals					
	1	2	3	4	5
ACR	0.0243	0.0186	0.0141	0.0103	0.0112
AR	0.0009	0.0007	0.0005	0.0004	0.0004
t-stat	1.2261	1.5939	1.0946	1.0413	1.2414

The difference between portfolios 1 and 5 remains about the same but the ACR's for all portfolios have shifted upward.

[15] We also repeated the experiment using 52-week rather than 26 week benchmarks. In this case, the difference between portfolios 1 and 5 for the Value Line data set was 4.49% for the single factor model and 3.25% for the five-factor model. This was the greatest difference which we found and it was not large.

Table 2 - Abnormal performance for 24 ranking dates from November 26, 1965 to February 3, 1978; single-factor and five factor market model benchmarks, estimated over a 26 week benchmark period

Panel A: Value Line Abnormal Returns[a]

Market Model		Portfolio Number				
		1	2	3	4	5
Single factor:	ACR	0.0033	0.0035	-0.0057	-0.0112	-0.0305
	AR	0.0001	0.0001	-0.0002	-0.0004	-0.0012
	t-stat	0.2545	0.3801	-0.5935	-1.10074	-2.1258*
Five factor:	ACR	0.0053	0.0058	-0.0005	-0.0071	-0.0289
	AR	0.0002	0.0002	-0.0000	-0.0003	-0.0012
	t-stat	0.3606	0.6704	-0.0699	-0.8054	-1.9640

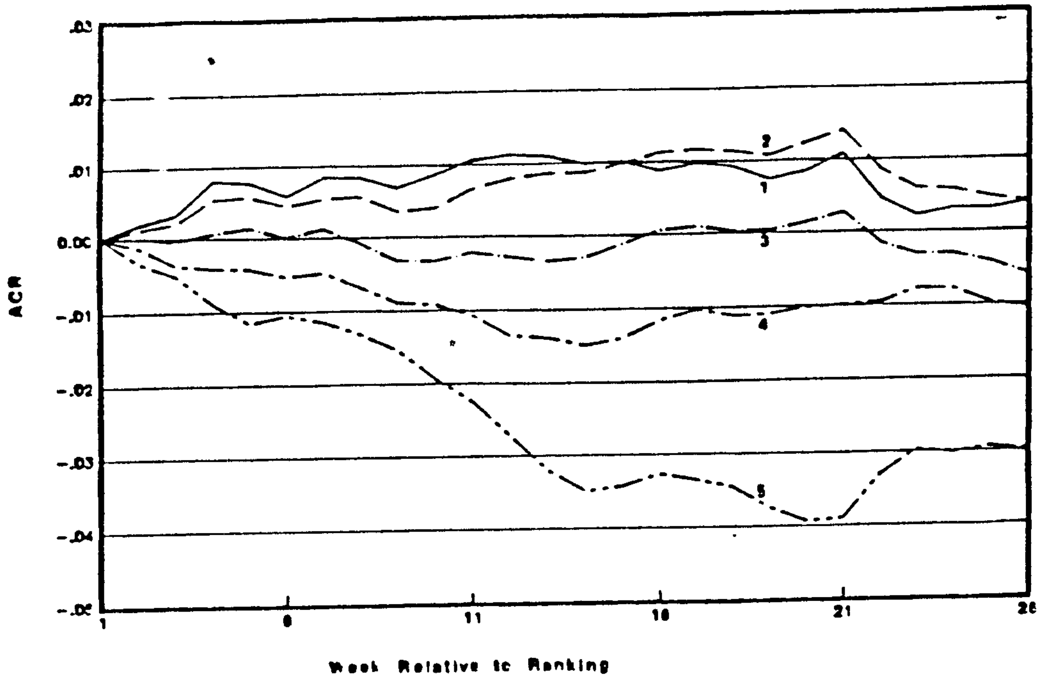
[a]Portfolio numbers correspond to the Value Line rankings of the securities in the portfolio. ACR is the average cumulative 26-week test period residual return, AR the average weekly residual return and t-stat the t-statistic under the null hypothesis that ACR = 0. Degrees of freedom for t-stats are 23. Thus, any t-stat greater than 2.069 is significant at the 0.05 level, and is marked with an asterisk.

Panel B: Size-ranked Portfolio Abnormal Returns[b]

Market model		Portfolio Number				
		1	2	3	4	5
Single factor:	ACR	0.0031	0.0027	-0.0020	-0.0113	-0.0098
	AR	0.0001	0.0001	-0.0001	-0.0004	-0.0004
	t-stat	0.2272	0.4997	-0.3665	-1.2051	-0.6850
Five factor:	ACR	0.0088	0.0062	0.0004	-0.0065	-0.0045
	AR	0.0003	0.0002	-0.0000	-0.0003	-0.0002
	t-stat	0.5320	0.5916	0.0396	-0.8659	-0.4879

[b]Portfolio numbers correspond to the size rankings of the securities in the portfolio. ACR is the average cumulative 26-week test period residual return, AR the average weekly residual return and t-stat the t-statistic under the null hypothesis that ACR = 0. Degrees of freedom for t-stats are 23. Thus, any t-stat greater than 2.069 is significant at the 0.05 level.

Panel A: Value Line Portfolio ACR's, Single Factor Market Model



Panel B: Value Line Portfolio ACR's, Five Factor Market Model

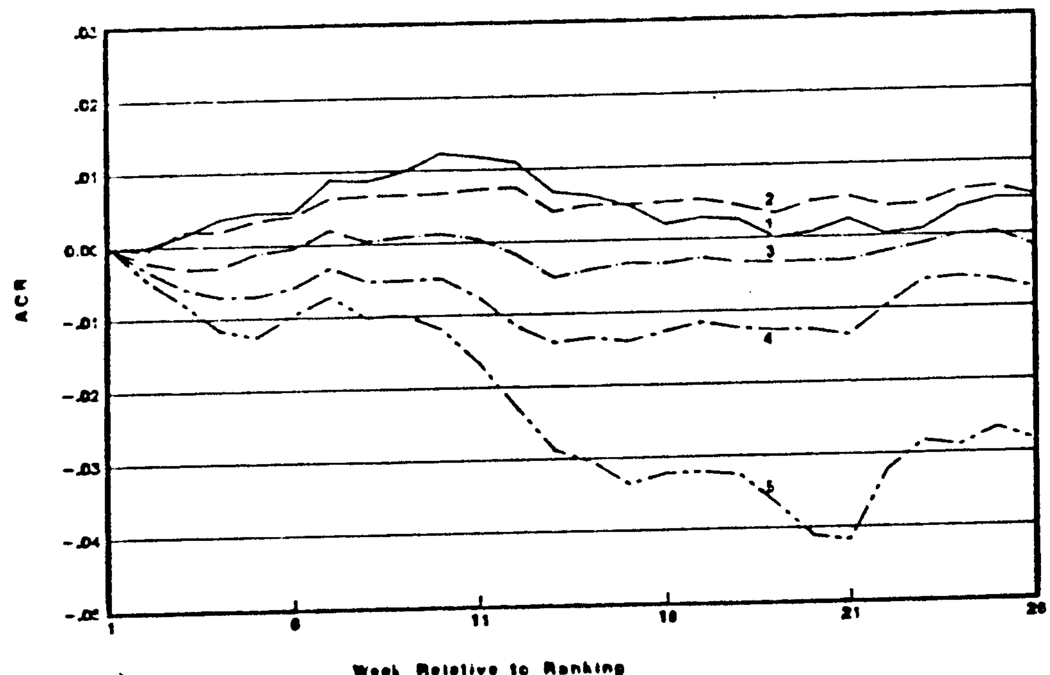
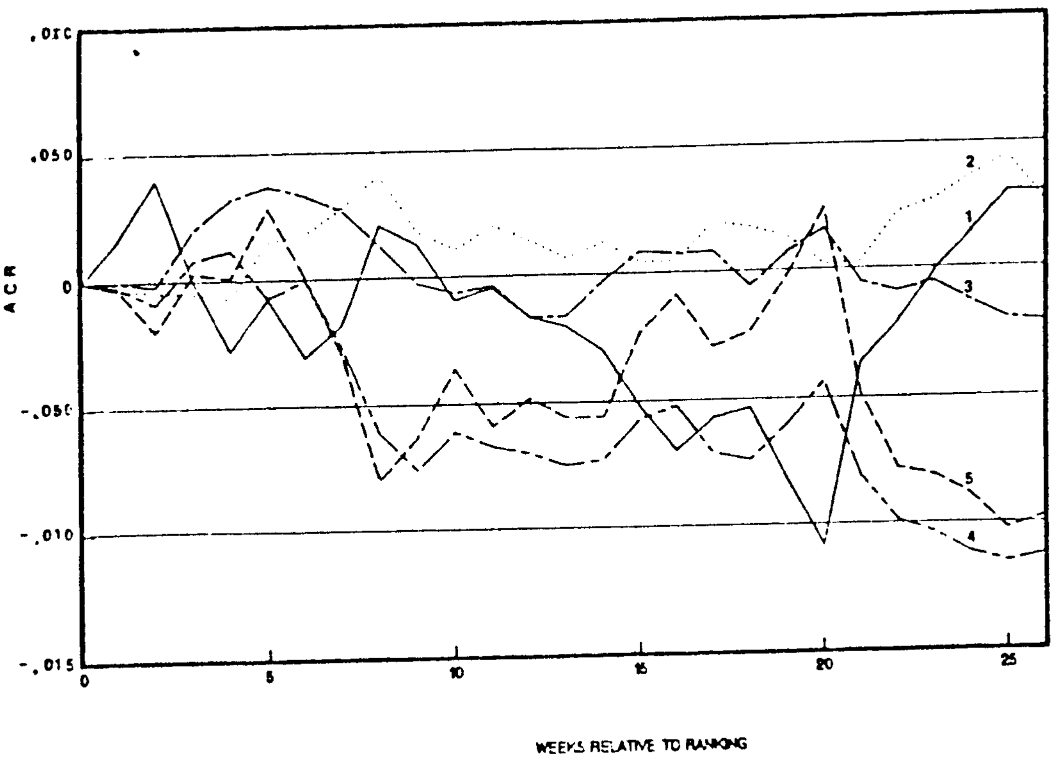
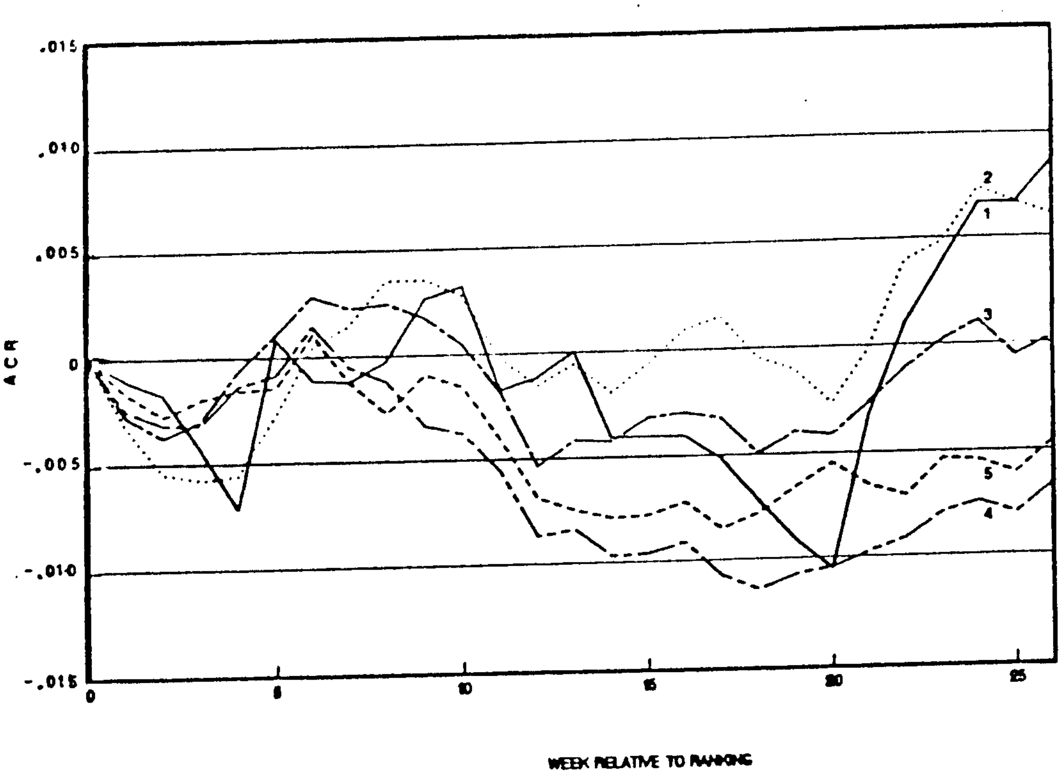


Figure 1 ~~Excess~~ Abnormal rate of return performance plots for five Value Line rankings portfolios (Panels A and B) and for five size-ranked portfolios (Panels C and D) for 24 ranking dates from November 26, 1965, to February 3, 1978. The ACR are the average (across 24 periods) of the weekly ~~excess~~ abnormal rates of return cumulated to the designated week relative to the ranking date in the test period (week zero). The numbers 1 to 5 correspond to the Value Line rankings of the securities in the portfolios, in Panels A and B, and to firm sizes (with 1 being the smallest quintile) in Panels C and D.

Panel C: Size-Ranked Portfolio ACR's, Single Factor Market Model



Panel D: Size-Ranked ACR's, Five Factor Market Model



using the multifactor model or 1.29% using the single factor model.

Figure 1 goes here.

Note, however, that the ACR plots of the size-ranked portfolios for the single factor model reveal negative cross-sectional correlation. Portfolios 1 and 5, and portfolios 2 and 4, seem to be reflections of each other. This is confirmed by large negative correlations between excess returns of portfolios 1 and 5(−.610) and 2 and 4(−.662). However, these large negative correlations are reduced to .281 and −.007 respectively, when the single-factor model is replaced with the five-factor model. This suggests that there is a left out variable in the single factor model which is uncorrelated with the single factor market index.

3.3 Mean Square Error Comparisons –

We observed above that the ACR's of the single and multifactor models are similar and not significantly different from zero. These results are to be expected if the market is efficient and the model parameters are stationary across the benchmark and the test periods. However, the single and multifactor model may still differ in their forecasting power. The forecast error variances, under the stationarity assumption, can be drived by comparing equation (4) and equation (6) with the realized returns.

The prediction error for the single factor model is

$$\begin{aligned} P1_w = & \; R_{pw} - E(R_{pw} | \hat{a}_p, \hat{b}_p, R_{mw}) \\ = & \; a_p + b_p R_{mw} + \epsilon_{pw} - [\hat{a}_p + \hat{b}_p R_{mw}] \\ = & \; [a_p - \hat{a}_p] + [b_p - \hat{b}_p] R_{mw} + \epsilon_{pw} \quad , \end{aligned}$$

and its variance under the usual assumptions is given by (see Theil (1971), p. 106):

$$\text{Var}(P1_w) = [\frac{1}{N} + \frac{(R_{mw} - \overline{R}_m)^2}{\sum_{t=27}^{52} (R_{mt} - \overline{R}_m)^2} + 1]\sigma_\epsilon^2 \tag{10}$$

where σ_ϵ^2 is the variance of ϵ_p, R_{mt} is the market return during the future benchmark period and $\overline{R}_m$ is the average of R_{mt}.

The prediction error for the multifactor model is

$$\begin{aligned} P5_w &= R_{pw} - E(R_{pw}|\hat{\alpha}_p, \hat{b}_{1p}, ..., \hat{b}_{5p}, R_{1w}, ..., R_{5w}) \\ &= [\alpha_p - \hat{\alpha}_p] + \sum_{k=1}^{5}(b_{kp} - \hat{b}_{kp})R_{kw} + \eta_{pw} \ , \end{aligned}$$

and the variance under the usual regression assumptions and the assumption that the five underlying factors are constructed to be orthogonal can be derived as:[16]

$$\text{Var}(P5_w) = [\frac{1}{N} + \sum_{k=1}^{5} \frac{(R_{kw} - \overline{R}_k)^2}{\sum_{t=27}^{52}(R_{kt} - \overline{R}_k)^2} + 1]\sigma_\eta^2 \tag{11}$$

where σ_η^2 is the variance of η_p, R_{kt} is the k^{th} factor return during the future benchmark period and $\overline{R}_k$ is the average $\overline{R}_{kt}$.

As we compare expression (10) with (11), it becomes immediately clear that the condition $\sigma_\eta^2 < \sigma_\epsilon^2$ is <u>not</u> sufficient to imply that the forecast error variance is lower for the multifactor model. If indeed $\sigma_\eta^2 < \sigma_\epsilon^2$, we still have to tradeoff the lower residual variance with the uncertainty arising from parameter estimation,

$$(R_{mw} - \overline{R}_m)^2 / \sum_{t=27}^{52}(R_{mt} - \overline{R}_m)^2 \ \text{ vs. } \ \sum_{k=1}^{5}[(R_{kw} - \overline{R}_k)^2 / \sum_{t=27}^{52}(R_{kt} - \overline{R}_m)^2].$$

This is true even in the special case where factor one, R_1, is constructed to be the market, R_m (*i.e.*, $R_2, ..., R_5$ are obtained by factor analyzing the market residuals). On the other hand, if factor one is a noisy approximation of the market (i.e., $R_1 = R_m + \varsigma$ where ς is noise uncorrelated with R_m), and the right hand side is greater than the left hand side in the above expression, then $\sigma_\eta^2 > \sigma_\epsilon^2$ is sufficient to imply that the forecast error variance of the single factor is lower.

By comparing (10) with (11), we can also infer under what conditions the forecast error variance of the multifactor model is likely to be lower. If we hold all other things (parameters, i.e. $\hat{a}_p, \hat{b}_p, R_m, \hat{\alpha}_p, \hat{b}_{1p}, ..., \hat{b}_{5p}, R_1, ..., R_5$, $\sigma_\epsilon^2 - \sigma_\eta^2$) the same, then since σ_ϵ^2 and σ_η^2 enter (10) and (11) multiplicatively, the lower σ_η^2 is, the more

[16]The orthogonality condition is not crucial for the discussion that follows. It does simplify the expression in Eq. (11).

likely (10) is greater than (11). In other words, the forecast error variance for the multifactor model is likely to be lower if the portfolio is well-diversified.

The above anlysis is derived under certain stationarity assumptions and offers us some guidelines of what to expect. A more informative comparison of the models still must come from the comparison of the realized forecast errors. In this section, we examine the mean square error criterion, which is defined as

$$MSE \equiv \frac{1}{24}\sum_{p=1}^{24}(CR_p)^2. \tag{12}$$

The results of these calculations are reported in Table 3. The multifactor model appears to do better. It has smaller MSE for Value Line portfolios 2, 3, and 4, and for size-ranked portfolios 1, 4, and 5. However, we need a significance test.

Table 3 – Mean Square Error Terms for the Single-Factor Model (column 2), the Five-Factor Model (col. 3), and Their Ratio (2÷3 = column 4)

Panel A: Value Line Portfolios

(1) portfolio number	(2) single factor MSE	(3) five factor MSE	(4) ratio = 2÷3
1	.00387	.00508	0.762
2	.00196	.00176	1.114
3	.00217	.00109	1.991
4	.00298	.00182	1.673
5	.00567	.00581	0.967

Panel B: Size-ranked Portfolios

(1) portfolio number	(2) single factor MSE	(3) five factor MSE	(4) ratio = 2÷3
1	.10387	.10315	1.007
2	.01687	.04518	0.373
3	.01724	.03075	0.561
4	.05118	.02594	1.973
5	.11598	.02604	4.454

If the test-period cumulative forecast errors of the two models were independent, the ratio of mean square errors could be interpreted as a doubly non-central F-test. However, as the cumulative forecast errors of the two models are correlated, a different significance test must be employed. A procedure introduced by Ashley, Granger, and Schmalensee (1980) is used to test the null hypothesis that the difference in the mean square error terms for the two models is zero. To explain their procedure, let ϵ_{1p} and ϵ_{5p} be the test period CR's for the single and multifactor models respectively. The difference between the mean square errors is

$$MSE(\epsilon_1) - MSE(\epsilon_5) = [s^2(\epsilon_1) - s^2(\epsilon_5)] + [m(\epsilon_1)^2 - m(\epsilon_5)^2] \quad (13)$$

where MSE is the sample mean square error, s^2 is the sample variance, and m is the sample mean. Now defining $\triangle_p$ as the difference in error terms and θ_p as their sum,

$$\triangle_p = \epsilon_{1p} - \epsilon_{5p} \; and \; \theta_p = \epsilon_{1p} + \epsilon_{5p}, \quad (14)$$

equation (13) can be written as

$$MSE(\epsilon_1) - MSE(\epsilon_5) = [COV(\triangle, \theta)] + [m(\epsilon_1)^2 - m(\epsilon_5)^2], \quad (15)$$

where COV is the sample covariance. From equation (15), a significant difference between the MSE's for the two models is indicated if we can reject the joint null hypothesis that $COV(\triangle, \theta) = 0$ and $m(\triangle) = 0$. This joint hypothesis can be tested by running the OLS regression

$$\triangle_p = A_1 + A_2[\theta_p - m(\theta_p)] + \mu_p \, , \quad (16)$$

where μ_p is a mean zero error term and is independent of θ_p. Equation (16) also assumes no significant autocorrelation. The null hypothesis of no difference in the MSE is equivalent to the joint hypothesis that $A_1 = A_2 = 0$. The two simple null hypotheses of no difference in prediction bias and of equivalent prediction variances are the same as $A_1 = 0$ and $A_2 = 0$, respectively.

Table 4 shows the results of regression (16) for the five Value Line portfolios and for the size-ranked portfolios. The intercept term, which is an estimate of the difference in excess returns between the single and multifactor models, is insignificant in all regressions. The slope term is significantly positive for Value Line portfolio 3 and size-ranked portfolio 5, indicating superiority of the multifactor model. However, the slope is significantly negative for size-ranked portfolios 2 and 3, indicating superiority for the single-factor model. These results are consistent with the ratios of mean-squared errors which were reported in Table 3.

Table 4 – Results of the OLS regression

$$\triangle_p = A_1 + A_2[\theta_p - m(\theta_p)] + \mu_p$$

for each of the five portfolios.

Panel A: Value Line Portfolios

Value Line Portfolio	Parameter Estimate	t-test	r^2/DW	F-test
1	A_1-.0020	-0.2386	0.0541	1.26
	A_2-.0747	-1.1214	1.9809	
2	A_1-.0023	-0.2994	0.0068	0.15
	$A_2$0.3780	0.3874	2.2257	
3	A_1-.0052	-0.7804	0.1766	4.72*
	$A_2$0.1989	2.1721*	2.1433	
4	A_1-.0041	-0.5535	0.1082	2.67
	$A_2$0.1365	1.6336	2.3342	
5	A_1-.0016	-0.1573	0.0014	0.03
	A_2-.0140	-0.1752	2.1340	

Panel B: Size-ranked Portfolios

Size-ranked Portfolio	Parameter Estimate	t-test	r^2/DW	F-test
1	A_1-.0057	-0.7250	0.1263	3.18
	A_2-.1002	-1.7831	2.7603	
2	A_1-.0034	-0.5673	0.4490	17.93*
	A_2-.3658	-4.2338*	2.3643	
3	A_1-.0024	-0.4106	0.3289	10.78*
	A_2-.3003	-3.2832*	2.3138	
4	A_1-.0047	-0.6668	0.0791	1.89
	$A_2$0.1322	1.3750	2.3993	
5	A_1-.0054	-0.7407	0.3750	13.20*
	$A_2$0.2444	3.6334*	2.6903	

An asterisk (*) indicates that the variable is statistically significant at the 5% confidence level or better.
DW is the Durbin Watson statistic.
F-test degrees of freedom are (2,22) with $F > 3.44$ being significant at the 5% confidence level.

In Table 5, for both the Value Line and Size-Ranked portfolios, we report the average standard deviations of the raw returns (Panel A), as well as the average residual standard deviations, (σ_ϵ), of the single-factor model (Panel B), and (σ_η) of the five-factor model (Panel C) from the benchmark period. For all the Value Line portfolios, their single-factor residual variances (σ_ϵ^2) are larger than the corresponding five-factor residual variances (σ_η^2). The middle portfolio has the smallest five-factor residual variance. From Table 4, it is also the portfolio where the five-factor model does significantly better in terms of forecast MSE. This is consistent with our analysis that the five-factor model is more likely to do better for well-diversified portfolios.

On the other hand, as a consequence of the choice of market index (the equally weighted index), the middle size-ranked portfolio has relatively low single-factor residual variance. The larger firm portfolios, which are usually regarded as more diversified, actually have higher residual variance. In fact, the equally-weighted market index captures so much of the return variations of the smaller firms that the single-factor model has lower residual variances than those of the multifactor model for the lowest three size-ranked portfolios. Even though the first factor from the multifactor model usually is also a proxy for the equally-weighted market, our factor is constructed with only 180 stocks rather than with the thousands of stocks that enter the market index and therefore is relatively noisy. These appear to be sufficient to push the forecast error variances the other way, and, as we observe in Table 4, the simple-factor model is significantly better for the second and third size-ranked portfolios.

Nevertheless, for the larger firm portfolios, their average multifactor model residual variance is lower than that of the smaller firm portfolios. Both the fourth and the fifth size-ranked portfolios have lower forecast MSE in the multifactor model, and significantly so for the largest firm portfolio. This pattern is again consistent with the expectation that the multifactor model performs better than the single-factor model when the portfolio is well-diversified.

Table 5 - Average Standard Deviations for the Simple-Factor and Five-Factor Model		
Panel A: Benchmark Period Raw Return Standard Deviations		
Portfolio	Value Line	Size-Ranked
1	.03332	.03392
2	.03224	.03321
3	.03089	.03226
4	.03014	.03152
5	.03125	.03057
Panel B: Benchmark Period Residuals: Single-factor (σ_ϵ)		
Portfolio	Value Line	Size-ranked
1	.02590	.02523
2	.02451	.02248
3	.02382	.02291
4	.02473	.02411
5	.02595	.02574
Panel C: Residuals from benchmark period: five-factor (σ_η)		
Portfolio	Value Line	Size-ranked
1	.02563	.02529
2	.02345	.02360
3	.02237	.02309
4	.02320	.02268
5	.02481	.02334
Average standard deviations are the average across the 24 holding periods.		

4 Summary and Conclusions

We have compared single and multifactor residual analysis evaluation techniques in a specific application to determine if a multifactor model improves on the more familiar single factor model. The ACR measurements were unbiased and similar whether using a single factor or a multifactor model. Upon reflection it should not be too surprising that the abnormal performance estimates were not different. The future benchmark methodology is essentially a market model technique, whether one uses a single or a multifactor index. Consequently, the intercept term accounts for the average return. Our results show no significant differences in the average cumulative returns. However, when portfolios were based on a CAPM anomoly, namely firm size, the patterns of ACR's were considerably different. The single factor ACR's exhibited significant negative cross-sectional correlation while the multifactor residuals did not.

Even though the single and multiple factor methodologies give the same point estimates of abnormal returns it is still possible that their mean square errors are different. A priori, one would expect the multifactor model to be more powerful when the model parameters are known with certainty. However, since we have to estimate the model parameters empirically, we have to trade off reductions in residual variance against parameter estimation errors. Our results show that the single and multifactor model mean square forecast errors are similar. However, the single factor model tends to have relatively more power for forecasting conditional returns for poorly diversified portfolios while the multifactor model tends to have relatively more power for forecasting conditional returns for well diversified portfolios.

Appendix A: How to Form the Mimicking Portfolios

Assume that returns for assets $(i = 1, ...I)$ are generated by a k-factor $(k = 1, ..., K)$ linear model such as

$$\tilde{r}_i = E_i + b_{i1}\,\tilde{\delta}_1 + ... + b_{iK}\,\tilde{\delta}_K + \tilde{\varepsilon}_i \qquad (A.1)$$

where E_i is the expected return during the next time interval; $\tilde{\delta}_k$ are the mean zero factors common to all assets; b_{ik} is the sensitivity of return on asset i to the fluctuations in factor k; and $\tilde{\varepsilon}_i$ is the idiosyncratic risk for the i^{th} asset with $E(\tilde{\varepsilon}_i, \tilde{\delta}_j) = 0$ for all k.

Chen (1983) has shown that there exists a unique linear- transformation that can generate the factor sensitivities, i.e., the b_{ik} for all assets corresponding to a fixed set of common factors. Chen's theorem enables us to form well diversified "mimicking" portfolios, one for each factor, which have high sensitivity to the k^{th} factor and zero sensitivity to all other factors. The subperiod (e.g., weekly) returns on these mimicking portfolios may be used as estimates of the factor returns (analogous to a K-factor market index), and then employed in a multi- index market model to estimate subperiod abnormal portfolio performance.

The procedure we used for estimating the weekly rates of return on the mimicking portfolios is:

1. Compute the variance covariance matrix for the weekly rates of return of the first 180 securities on the CRSP tape which had weekly returns every week.[17] The data was separated into two halves. The first half began with Friday November 19, 1965 and ended with Friday May 4, 1973, thereby spanning the first 12 holding periods. The second half began on Friday August 4, 1972 and ended with Friday October 26, 1979. It spanned the second 12 holding periods.[18]

2. Compute ten factor loadings for the 180 stocks by using the *Jöreskog* asympototic maximum likelihood method.[19]

---

[17] Of the 4063 companies on the CRSP tape, 999 had weekly returns for each of the 861 weeks between the beginning of July 1962 and the end of December 1978. Of these, we used only the first 180 stocks (chosen alphabetically) because we were constrained by the processing capacity of the IBM 3033 in use.

[18] The mean weekly rate of return for the i^{th} security in the t^{th} year was used to compute the mean deviations that year. This was done yearly in order to emphasize the covariability among securities.

[19] The software package EFAP II, written by *Jöreskog* was utilized.

3. Based on the factor loadings, form 5 mimicking portfolios.

Note that although ten factor sensitivities for each of the original 180 securities were produced by the *Jöreskog* routine, we created only five mimicking portfolios. There are two reasons. First, no one knows the exact number of true underlying factors, but Roll and Ross (1980) suggest there are at least three and probably four "priced" factors. Hence, our choice of five factors is arbitrary, but is probably reasonable. The second reason has to do with a possible misspecification problem. So long as the eleventh and higher order factors are immaterial (and there is good reason to believe that they are) we can be sure that the first five factor loadings are not contaminated by misspecification by setting the loadings for the sixth through tenth factors to zero.

A linear programming model (using 180 securities) was used to determine sets of weights for 5 mimicking portfolios. [20] The objective was to minimize departures from a well diversfied, equally weighted portfolio, subject to constraints that 1) the mimicking portfolios should be uncorrelated with each other and 2) that the weights in the first portfolio should sum to one while the weights in the remaining four portfolios sum to zero. This procedure was repeated twice to produce first half and second half matrices of mimicking portfolio returns (where there were 390 weeks in the first half and 378 weeks in the second). Table A1 shows summary statistics for the mimicking portfolios: the matrix of sensitivities, and a correlation matrix for the first half data.[21] Panel A shows that the p^{th} portfolio has high sensitivity to the k^{th} factor when p = k but virtually zero sensitivity to all other factors where $p \neq k$. Panel B, the correlation matrix, provides a way of checking the quality of the mimicking portfolios. The non-diagonal elements should ideally be zero. The largest two correlations are 6.202% and -3.070%.

---

[20]This is the GUB routine within the elastic programming in the XS mathematical programming system developed by Glenn Graves, Professor of Mathematical Methods, Graduate School of Management, UCLA.

[21]In order to save space we have not printed either the entire period or the second half tables. They are essentially the same except that the maximum correlations which resulted from fitting over the entire period were as high as 16.8% and 11.6%. This is further evidence that non-stationarity was a problem.

Table A1

Summary statistics for 5 mimicking portfolios chosen from 180 securities. Factors estimated using the first half of the returns data (April 65 - January 72)[22]

Panel A:	Factor sensitivities, $\sum w_{ip}\, b_{ik}$. For example, row 1 - column 1 is the sensitivity of the first mimicking portfolio of the first factor.					
				Factor Number		
		1	2	3	4	5
Mimicking	1	-.4154E+00	.3169-06	.4556E-06	.49990E-07	-.4398E-08
Portfolio	2	.6694E-06	.2000E+00	.1146E-06	.2002E-06	-.4398E-08
Number	3	-.1544E-06	.15464E-06	.2000E+00	.5945E-07	-.7503E-09
	4	.2792E-06	.1824E-06	.9242E-07	.1500E+00	-.5650E-07
	5	-.4028E-07	.1214E-07	-.4060E-08	-.4736E-07	.1682E+00

Panel B: Correlations Among the 5 mimicking portfolios						
				Factor Number		
		1	2	3	4	5
Mimicking	1	1.00000	-.00164	-.00023	-.00398	.00766
Portfolio	2		1.00000	-.03076	-.06202	.00450
Number	3			1.00000	-.02809	.01286
	4				1.00000	-.00993
	5					1.00000

[22]The second half factor sensitivities and correlations among the mimicking portfolios were very similar.

References

Ashley, R., C.W.J. Granger, and R. Schmalensee, "Advertising and Aggregate Consumption: An Analysis of Casuality", *Econometrica*, July 1980, 1152-1159.

Brown, S. and J. Warner, "Measuring Security Performance", *Journal of Financial Economics*, September 1980, 205-258.

Brown, S. and M. Weinstein, "The Use of Derived Factors in Event Studies" (working paper, University of Southern California, November 1983).

Chen, N., "Some Empirical Tests of the Theory of Arbitrage Pricing", *Journal of Finance*, December 1983, 1393-1414.

Copeland, T.E. and D. Mayers, "The Value Line Enigma (1965-1978): A Case Study of Performance Evaluation Issues", *The Journal of Financial Economics*, November 1982, 298-321.

Huberman, G., "Arbitrage Pricing Theory: A Simple Approach", *Journal of Economic Theory*, October 1982, 183-191.

King, B., "Market and Industry Factors in Stock Price Behavior", *Journal of Business*, January 1966, pp. 134-190.

Reinganum, M., "The Arbitrage Pricing Theory: Some Empirical Results", *Journal of Finance*, 1981.

Reinganum, M., "Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings Yields and Market Value", *Journal of Financial Economics*, March 198l, 19-46.

Roll, R. and S. Ross, "An Empirical Investigation of the Arbitrage Pricing Theory", *Journal of Finance*, December 1980, 1073-1103.

Roll, R. and S. Ross, "Regulation, the Capital Asset Pricing Model, and the Arbitrage Pricing Theory", *Public Utilities Fortnightly*, May 26, 1983, 22-28.

Ross, S.A., "The Arbitrage Theory of Capital Asset Pricing", *The Journal of Economic Theory*, December 1976, 343-362.

Stickel, S.E., "The Effect of Value Line Investment Survey Rank Changes on Common Stock Prizes", *Journal of Financial Economics*, March 1985, pp. 121-143.

Theil, H., *Principles of Econometrics*, John Wiley and Sons, New York, 1971.