Journal of Banking & Finance 25 (2001) 1187–1208
Journal of BANKING & FINANCE
www.elsevier.com/locate/econbase

Forecasting correlation among equity mutual funds

Parvez Ahmed *

School of Business Administration, Pennsylvania State University – Harrisburg, Middletown, PA 17057-4898, USA

Received 8 February 1999; accepted 13 April 2000

Abstract

Investors often buy multiple funds that are actively managed, specializing within narrowly defined market segments. To successfully implement a strategy of diversification investors must obtain accurate estimates of correlation among mutual fund returns. This paper forecasts mutual fund correlations using eight models that are broadly classified into historical, mean and index. Results indicate that estimate of future correlations from the Multi-Style Index, Dynamic and Fama–French 3-Factor models have the lowest prediction errors. Moreover, the relative ranks of Multi-Style Index and Fama–French 3-Factor models have lower dispersion across different forecasting time periods and in sub-samples of funds belonging to similar or different 'style' categories. © 2001 Elsevier Science B.V. All rights reserved.

JEL classification: G11; G14

Keywords: Correlations; Mutual funds; Style investing

1. Introduction

Correlation among securities is one of the most important determinants of portfolio risk. An investor's ability to produce accurate ex-ante estimates of

* Tel.: +1-717-948-6162; fax: +1-717-948-6456.
E-mail address: pxa13@psu.edu (P. Ahmed).

PII: S 0 3 7 8 - 4 2 6 6 (0 0) 0 0 1 2 9 - 1

correlations will determine the optimal choice of securities in their portfolio. This paper evaluates various models to determine their efficacy in forecasting correlations among equity mutual funds. Estimating the correlation structure of stock returns has been the subject of many prior researches. However, no paper has demonstrated the use of forecasting models in estimating the correlation structure of mutual funds. This study distills estimates of future correlations using information contained in past fund returns. The results of this paper have important implication in managing a portfolio of mutual funds.

Since Fama and French's (1992) celebrated paper many studies have shown that stocks with some common characteristics like their market size and book-to-market ratios tend to perform similarly over several economic and market cycles. Thus, fund managers tend to invest in stocks with similar firm-specific characteristics as their price-to-earnings ratios, price-to-book ratios, dividend yields, market values, etc. With this method of asset allocation, commonly referred to as 'style investing', becoming popular it makes sense for investors to pursue a multiple fund strategy to portfolio management. Although a typical equity mutual fund holds more than 60 stocks in its portfolio, recent research by O'Neal (1997) shows that a multi-fund portfolio is far less risky than its single-fund counterpart. [1] Moreover, the popular press also advises investors to pursue a multi-fund strategy. [2]

Estimates of future correlation among mutual funds come from three different classes of models – *historical, mean and index*. One of the easiest ways to estimating future correlation is imputing them from pairwise correlation over some historical time period. Elton and Gruber (1973), Elton et al., (1978) and Eun and Resnick (1984) find such historical correlations to be poor estimators of future correlation among stocks. They attribute this to the historical model's inability to filter out noise. Averages of the historical correlations can eliminate unnecessary noise in the forecasts. For stocks, Eun and Resnick (1992) show such an average to be the best estimator of future correlation. Index models provide yet another alternative for calculating correlation. Using an assumption of how securities move together, index models allow implicit derivation of the correlation matrix. The derived correlation matrix may do well in explaining historical returns but perform poorly in forecasting future correlations. Thus, it is a matter of empirical investigation to determine which models predict mutual fund correlations with the lowest errors. Results show the

[1] As of March 1999, the median number of stocks held by mutual funds was 78 for Morningstar classified large cap funds, 139 for small cap funds, 78 for growth funds and 87 for value funds.

[2] McGough (1998) writes, 'Funds may be invested in types of stocks that have been popular recently (why else would you be looking at them?) but that may not be so popular a year from now (the market is inscrutable). Since none of us can predict which funds, or even types of funds, will excel in the future, it makes sense to hold several different types of funds.'

arbitrage pricing theory (APT) inspired Multi-Style Index and Fama–French 3-Factor models to produce forecasts that consistently have low prediction errors.

Sharpe (1992) shows that appropriate 'style' classification will enable an investor to effectively diversify. Holdings-based method and returns-based method are most commonly used to classify funds. Services like *Morningstar* use the holdings-based method which categorizes funds on the basis of average market capitalization and average price-to-earnings of the fund portfolio. The holdings-based method may be a better way for classifying funds as historical correlations are poor predictors of future correlations (Christopherson, 1995). However, this method requires establishment of boundaries on some differentiating characteristic which are vague. [3] Trzcinka (1995), Brown and Goetzman (1997), Gallo and Lockwood (1997) and diBartolomeo and Witkowski (1997) demonstrate the successful use of the returns-based classification method. Its success may be due primarily to the scheme being parsimonious with the data, simple to model, and cost effective in its use. The returns-based method also reduces the management incentive to 'game' the styles in order to improve ex-post rankings and they provide sufficient discrimination between funds which results in significant diversification benefits. Thus, this paper uses a returns-based classification method similar to Gallo and Lockwood (1997).

The rest of the paper is organized as follows: Section 2 describes the data and a description of the returns-based method used for classifying the mutual funds. Section 3 illustrates the use of alternative forecasting models in estimating future correlations. Section 4 describes methods for evaluating forecasting models. Section 5 presents the results of the study and Section 6 offers a summary of the paper.

2. Data and fund classification

The sample for this study consists of 202 equity mutual funds from *Morningstar's Mutual Fund OnDisc* database. To get a long series of fund returns, which will then allow tests of inter-temporal stability in forecasts, only funds that have all 240 monthly returns available over the period April 1979–March 1999 are selected. [4] It is a common practice among researchers to use a long

[3] *Morningstar* classifies all funds with average price-to earnings and price-to-book ratios 12.5% below those of the S&P 500 Index as 'value' funds.

[4] By limiting the sample to surviving funds, the study may induce some survivorship bias into the results. While survivorship bias may affect tests of performance evaluation (Brown et al., 1992; Grinblatt and Titman, 1989) its effect on models of forecasting is not clear.

series of data, usually 7–10 years when using monthly observations, in estimating model parameters. Many papers show the betas or sensitivities of security returns to various risk factors change over different economic conditions (Ahmed and Lockwood, 1998). Thus, it may be useful for a forecasting model to span at least one of the different economic/business cycles. This allows parameter estimates to be averaged across different business and economic conditions and not be specific to any economic regime. Historical market condition data show different economic conditions to sometimes span as many as 80 months. [5]

This study uses 10 years of monthly observations to estimate correlations using various forecasting models. Out-of-sample data with two different forecasting (investing) horizons of 3 and 5 years test the forecasts. [6] To allow inferences on the inter-temporal stability of the models, the process of creating matched model and test periods are repeated by moving 1 year forward while dropping the first year. This process creates eight, M1–M8, data sets for estimating correlations.

Following Gallo and Lockwood (1997), this study classifies a mutual fund into one of four style categories on the basis of the sensitivity of a fund's standardized returns to those of the four *Wilshire* indices – large capitalization growth (LCG), small capitalization growth (SCG), large capitalization value (LCV) and small capitalization value (SCV). [7] Eq. (1) shows the time-series regression of each fund's monthly standardized returns against the standardized monthly return on the four *Wilshire* indexes:

$$R_{it} = b_{i0} + b_{i1}SSCV_t + b_{i2}SSCG_t + b_{i3}SLCV_t + b_{i4}SLCG_t + \varepsilon_{it}, \tag{1}$$

Standardized returns are obtained by subtracting the mean monthly return from the actual return in month t and dividing the difference by standard deviation of the returns. Regression coefficients (betas), b_i's, are the sensitivities

[5] Weisenberger Investment Companies, Business Conditions Digest and National Bureau of Economic Research.

[6] Investing horizons of 3–5 years seems to be of most practical relevance. The existence of contingent deferred sales charges imposes significant costs for early exit from a mutual fund. These charges often start at 5% for money withdrawn within a year and declines thereafter until they disappear in 3–4 years. Additionally, the study also conducts out-of-sample tests for 1- and 10-year investing horizons. These results are not presented partially out of concern for the length of the paper. However, they are available upon request.

[7] The average correlation among the four *Wilshire* indices is 0.84 over the April 1979–March 1999 period. In contrast, the average correlation between the *Morningstar* categories of equity income, growth and income, growth, aggressive growth is 0.96. The lower correlations among the *Wilshire* indexes suggest greater potential for diversification by choosing funds from these mutually exclusive categories.

of the standardized returns of a mutual fund i on the standardized returns of the *Wilshire* indices. Funds are classified into the category of its highest beta weight.

3. Forecasting models

This study compares eight models for their ability to forecast correlations among mutual funds. The eight models can be grouped into three major categories – *historical*, *mean*, *and index models*. Funds in each model period are stratified by its style class from Eq. (1). It is possible that a model may do well in predicting correlations between funds belonging to a certain style (intra-style) but fail to do so for funds in different style categories (inter-style). Thus, this study also compares both the forecasting ability of each model in intra- and inter-style sub-samples. The difference (error) between the forecasts and the actual correlations in the holdout sample will determine the efficacy of each model. The best model will have the lowest mean of the squared errors or the lowest mean of the absolute errors.

3.1. Historical model

The most direct way of estimating future correlation is to compute pairwise correlation over some historical period and use them as estimates of future correlations. The Historical model computes pairwise correlations from the historical time-series returns of each fund in the model period. This model assumes that these correlations are the best estimates of future pairwise correlations among the mutual funds. Since this model directly estimates each pairwise correlation, it is the most disaggregate of all models.

3.2. Mean models

The mean model assumes that the past correlation matrix contains information about the average correlation between funds in the future but cannot account for their individual differences from the average. The model computes the average or mean of the pairwise correlations from the historical correlation matrix and treats this mean as a forecast of the future pairwise correlation between all pairs of funds. Elton and Gruber (1995, p. 168) call this model the Overall Mean model and suggest that more elaborate models need to be judged against this naive model.

Many studies have shown firm size, price-to-earnings ratios, and book-to-market ratios explain the cross-sectional variation in stock returns. [8] Consistent with these findings, others have shown that market capitalizations and value/growth differentials explain monthly returns of mutual funds. [9] A more disaggregate averaging assumes that there are common mean correlations in subgroups of mutual funds stratified by their returns-based style class. In such a model, labeled the Style Mean Model, the historical mean correlation of all funds in one category is assumed to be a predictor of correlations among the funds in that category. For funds in different style categories the averages of all historical pairwise correlations between funds in two different styles predicts the pairwise correlation coefficient between the two funds from those two style categories.

3.3. Index models

A third class of models, the index models, assumes securities move together because of their response to a set of common factors. Depending on the assumption of how many factors (one or many) influence security returns the category of index models can be further subdivided into Single-Index and Multi-Index type. In its simplest form, the market model, shows changes in stock returns being correlated with changes in a broad market index. Following Ross's (1976) arbitrage pricing theory, several studies have shown stock returns to depend on factors beyond the market index or on factors other than the market index.

3.3.1. Single Index model

The Single Index model assumes that mutual fund returns vary in response to changes in the equity markets. A broad equity market index such as the *Wilshire 5000* proxies the changes in an overall market portfolio. Eq. (2) describes a Single Index model:

$$R_i = \alpha_i + \beta_i R_{\mathrm{M}} + \varepsilon_i, \quad i = 1, \ldots, n. \tag{2}$$

R_i and R_{M} are the monthly rates of return on a mutual fund i and the *Wilshire 5000* index, respectively. This model assumes that errors, ε, from the regression of two funds will be unrelated [$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$]. Also, by construction, the mean error for a fund is zero [$E(\varepsilon_i) = 0$]. Eq. (3) then calculates the correlation coefficient between two funds i and j as [10]

[8] See Carlton and Lakonishok (1986), Banz (1981), Reinganum (1981), Rosenberg et al., (1985), Keim (1983), Lamoureux and Sanger (1989), Basu (1983), Chan et al., (1983), and Fama and French (1992).

[9] See Sharpe (1992), Gallo and Lockwood (1997), Tierney and Winston (1991).

[10] For a derivation of this equation refer to Elton and Gruber (1995, pp. 132–133).

$$\rho_{ij} = \frac{\beta_i \beta_j \sigma_{\mathrm{M}}^2}{\sigma_i \sigma_j}, \quad i \neq j. \tag{3}$$

3.3.2. Multi-index model

The multi-index models search for a set of factors that can capture the common movement in mutual fund returns. Elton and Gruber (1995) caution that, 'while it is easy to find a set of indices that is associated with nonmarket effects over any period of time, ...it is quite another matter to find a set that is successful in predicting covariances that are not market related.' Eq. (4) provides a general representation of the multi-index model:

$$R_i = \alpha_i + \sum_{k=1}^{K} \gamma_{ik} R_k + \eta_i, \quad i = 1, \ldots, n. \tag{4}$$

In this equation, R_i and R_k are the monthly rates of return on a mutual fund i and K sets of indexes. γ_{ik} measure the response of a fund i's return to changes in index k. Its meaning is analogous to b_i in the Single Index model. For the correlations to have desirable mathematical properties (e.g. bounded by -1 and $+1$), it is necessary to have the K indices uncorrelated or orthogonal to each other. In its orthogonalized form, the sensitivity γ_{ik} can be thought of as the response of a fund to changes in index k holding all other index levels constant. [11] By construction, the expected value of the error term, h, is zero, $E(\eta_i) = 0$. Moreover, the covariance between indices $k = 1$ and $k = 2$ is zero, $\mathrm{Cov}(R_1, R_2) = 0$ and the covariance between the residual for fund i and index k is zero, $\mathrm{Cov}(\eta_i, R_k) = 0$. Finally, the multi-index model assumes that the only reason the funds move together is because of their common response to these set of indices, thus, $\mathrm{Cov}(\eta_i, \eta_j) = 0$. Eq. (5) shows the calculation of pairwise correlation coefficients from any multi-index model:

$$\rho_{ij} = \frac{\sum_{k=1}^{K} \gamma_{ik} \gamma_{jk} \sigma_K^2}{\sigma_i \sigma_j}, \quad i \neq j. \tag{5}$$

This paper studies five multi-index models. Sections 3.3.2.1–3.3.2.4 describe the factors assumed to affect fund returns. The calculations of pairwise correlations between funds will follow Eq. (5).

[11] The process of orthogonalization poses no theoretical problems for it is possible to take a set of correlated indices and transform them into uncorrelated indices. In all index models each index or factor is orthogonalized by regressing it on the 'preceding' indices and the resulting residuals are used in the model. Elton and Gruber (1995, pp. 174–175) provide detailed illustration on the process of orthogonalization.

3.3.2.1. Style Index model. A special case of the multi-index model is the Style Index model. This model attempts to capture the influence of a fund's commonality with a certain investment style. In this model two factors, market index M and style index s are assumed to affect fund returns. [12] R_{M} represents the market return and R_{s} the return of the style index to which the fund belongs. While the correlation between funds in the same style category is given as in Eq. (5), for funds in different style categories, Eq. (6) calculates the pairwise correlations as follows:

$$\rho_{ij} = \frac{\gamma_{i\mathrm{M}}\gamma_{j\mathrm{M}}\sigma_{\mathrm{M}}^2}{\sigma_i\sigma_j}, \quad i \neq j. \tag{6}$$

3.3.2.2. Multi-Style Index model. Eun and Resnick (1992) use a multi-index model in which stock returns are dependent on three size based indexes representing large, mid and small capitalization firms. However, several academic studies indicate that both market size and value/growth differential explain the average cross-sectional variation in stock returns. [13] These findings motivate the inclusion of factors that combine size and value/growth characteristics. Eq. (1) represents this model with actual monthly returns as the model variables instead of standardized returns. The Multi-Style Index model uses the four Wilshire style indices (k = LCV, LCG, SCV and SCG) as the four factors that explain mutual fund returns. Each index represents a mutually exclusive style class. Gallo and Lockwood (1997) show that this model is effective in explaining the historical correlation structure of mutual fund returns.

3.3.2.3. Fama–French 3-Factor model. Kothari and Warner (1997) study various performance benchmarks to evaluate which one, if any, provides a realistic evaluation of mutual fund performance. Their results indicate that a procedure based on Fama and French (1993) 3-Factor model outperforms traditional CAPM based measures. They view their results as indicating the usefulness of style analysis in benchmarking fund returns. The three factors in this model are $R_m - R_f$, HML and SMB, where $R_m - R_f$ is the excess market return, HML is the book-to-market factor which is computed as high minus low book-to-market portfolio return and SMB is the size factor return computed as small minus big firm size portfolio return. [14]

[12] This model may suffer from significant benchmark (tracking) error if the style index beta does not contain the majority of a fund's total style exposure (total beta). This concern maybe mitigated by the average difference in beta between the index with the highest beta and the other indices being as large as 0.4976. Further, on average, the beta of the classified index was 59% of the total beta. Finally, 71% of the funds had its highest beta exceed 50% of the total beta of the fund.

[13] See Banz (1981), Reinganum (1981, 1990), Basu (1977), and Fama and French (1992).

[14] Eugene Fama is gratefully acknowledged for providing the three factors.

3.3.2.4. Dynamic model. Most of the common measures of fund performance are unconditional models that use historical average returns to estimate expected performance. Ferson and Warther (1996) assert that the shortcoming of the unconditional measures is that they do not reflect the time-variation of risk and expected returns. They suggest a conditional or Dynamic model in which the three factors are RM_t, $RM_t \times (D/P)_{t-1}$ and $RM_t \times (TB)_{t-1}$ where RM_t, is the return in month t for the S&P 500 index, $(D/P)_{t-1}$ is the lagged value of the market dividend yield and $(TB)_{t-1}$ is lagged value of the short-term Treasury yield. The product terms, $RM_t \times (D/P)_{t-1}$ and $RM_t \times (TB)_{t-1}$, essentially pick up the movement through time of the conditional betas as they relate to market indicators. These additional factors account for the dynamic strategies followed by many fund managers.

4. Measures of forecasting performance

The mean squared error (MSE) evaluates and ranks the forecasting power of each model. Given n pairs of matching forecasts (F_i) and actual values (A_i), Eq. (7) computes the MSE as follows:

$$\mathrm{MSE} = \sum_{i=1}^{n} \frac{(F_i - A_i)^2}{n}. \tag{7}$$

To rank the various competing models, this study calculates the differences, D_i, in the squared forecast errors between matched pairs of the forecasting models:

$$D_i = (F_{i1} - A_i)^2 - (F_{i2} - A_i)^2. \tag{8}$$

F_{i1} and F_{i2} refer to the forecast values of Model 1 and Model 2 for the ith entry of the correlation matrix while A_i refers to the actual correlation from the holdout test sample. If the mean of these paired differences, D_i, is negative and statistically different from 0 at the 5% level, then Model 1 produces superior forecasts of correlation than Model 2. Two-tailed t-test is used to evaluate the statistical significance of the differences in mean.

Theil (1971) shows the MSE can be decomposed into three terms each of which represent a particular type of forecast error:

$$\mathrm{MSE} = \sum_{i=1}^{n} \frac{(F_i - A_i)^2}{n} = (\overline{F} - \overline{A})^2 + (\sigma_{\mathrm{F}} - \sigma_{\mathrm{A}})^2 + 2(1-\rho)\sigma_{\mathrm{F}}\sigma_{\mathrm{A}}. \tag{9}$$

where $(\overline{F}, \overline{A})$, and $(\sigma_{\mathrm{F}}, \sigma_{\mathrm{A}})$ are the means and standard deviations of the forecasts and the actual correlations and the ρ is the correlation between the forecasts and the actual values. The first term measures the error due to biased forecasts; the second term measures the error due to unequal variation; and the third term measures the error due to imperfect correlation.

An alternative method to evaluate the forecasting ability of a model is to compute the mean absolute errors (MAE) in a manner analogous to the computations for MSE. Eq. (10) shows the formula for calculating MAE:

$$\mathrm{MAE} = \sum_{i=1}^{n} \frac{|F_i - A_i|}{n}. \tag{10}$$

5. Empirical results

5.1. Style classification and sample characteristics of mutual funds

Table 1 shows the cross-sectional averages of the betas for funds stratified into one of four style classes of large-cap growth (LCG), large-cap value (LCV), small-cap growth (SCG) or small-cap value (SCV). Betas measure the sensitivity of mutual fund returns against the four *Wilshire* style indexes. A mutual fund is assigned to one of four style classes based on its highest beta weight. All statistics in Table 1 are averages across the eight model periods, M1–M8. N indicates the number of funds in each style class. The return-based classification scheme in Eq. (1) on average categorizes 120 funds as large-cap growth, 22 funds as large-cap value, 52 funds as small-cap growth and 9 funds as small-cap value. [15] The minimum, maximum and standard deviation in N indicates a 'style drift' with large number of funds changing a style category in different time periods. Table 1 also shows that for each style class the mean beta weight within matched or classified style is the highest in each row and is significantly different from out-of-style beta weights. The results indicate that the returns-based style classification method allows funds to be highly correlated with their allocated style index and have very low correlation with the other style indexes.

Table 2 reports the risk and return statistics for the 202 mutual funds stratified by matched model and test periods. Results show that for most funds, average fund return in the model period is a poor estimator of fund performance in the test period. The t-statistics show the significance of difference in mean returns and variances between the cross-section of funds in their model and test periods and the column NUM shows the number of funds that report these significant differences. In only two matched pairs (T1–3 and T2–3 for the 3-year forecasting horizon) fewer than 50% of the funds report significantly different means and standard deviations between their model and test periods.

[15] Classifying funds over the 1986–1993 period, Gallo and Lockwood (1997) report the following proportions for the style categories: LCG (50%), LCV (5%), SCG (35%) and SCV (9%).

Table 1
Summary statistics on classification of funds into 'style' categories – sensitivity of funds to Wilshire style indexes

Variables	Style class	N	Beta LCG	Beta LCV	Beta SCG	Beta SCV
Mean		120	**0.5421**	0.2371	0.1837	0.0317
t-Statistics			**6.3280***	1.5072	2.9143*	0.3444
Diff in means (*t*-statistics)	LCG			10.35*	10.07*	7.43*
S.D.		12	**0.0273**	0.0449	0.0160	0.0442
Min		103	**0.5260**	0.0748	0.1532	−0.0367
Max		134	**0.6009**	0.1848	0.1992	0.0872
Mean		22	0.2372	**0.4606**	0.1909	0.0591
t-Statistics			2.9692*	**5.6688***	3.2491*	0.6920
Diff in means (*t*-statistics)	LCV		10.35*		13.27*	5.58*
S.D.		9	0.0293	**0.0146**	0.0233	0.0705
Min		12	0.1803	**0.4129**	0.0827	0.0104
Max		34	0.2628	**0.4567**	0.1507	0.2058
Mean		52	0.2362	0.1909	**0.5966**	0.1573
t-Statistics			2.2333	0.1920	**7.6246***	0.2291
Diff in means (*t*-statistics)	SCG		10.07*	13.27*		6.90*
S.D.		6	0.0252	0.0235	**0.0204**	0.0216
Min		46	0.2102	−0.0224	**0.5867**	−0.0301
Max		66	0.2828	0.0404	**0.6501**	0.0462
Mean		9	0.1693	0.0591	0.1573	**0.3602**
t-Statistics			2.0995	1.2166	2.6345	**4.0610***
Diff in means (*t*-statistics)	SCV		7.43*	5.58*	6.90*	
S.D.		5	0.0672	0.0688	0.0757	**0.0566**
Min		3	0.0590	−0.0558	0.1109	**0.3020**
Max		18	0.2351	0.1558	0.3049	**0.4404**

* Significant at 1%.

These results indicate that investors are better off pursuing a multi-fund strategy.

5.2. Forecasting performance

MSE and MAE from Eqs. (7) and (10) measure the overall performance of the eight competing models. For each model, the first step computes the cross-sectional averages of the squared and absolute errors for the 202 funds. The next step averages MSE and MAE across all matched model-test periods. [16] Table 3 reports MSE and MAE for both forecasting horizons of 3 and 5 years.

[16] The period-by-period MSE and MAE statistics are available upon request.

Table 2
Summary statistics on fund's risk-return characteristic stratified by matched model and test periods

Model and matched test periods	All funds			LCG	LCV	SCG	SCV
	Mean return (S.D.)	*t*-Test diff in mean return (S.D.)	NUM	Mean return (S.D.)	Mean return (S.D.)	Mean return (S.D.)	Mean return (S.D.)
Model M1: April	16.39			15.84	16.54	16.96	16.93
1979–March 1989	(17.93)			(17.08)	(14.49)	(20.00)	(14.95)
Test T1-3: April	15.04	3.25*	48, 33	14.55*	11.23*	16.49	13.09*
1989–March 1992	(16.04)	(1.25)		(15.65)	(14.58)	(18.42)	(14.77)
Test T1-5: April	12.65	11.21*	79, 104	12.33*	10.88*	13.54*	11.33*
1989–March 1994	(14.32)	(3.86*)		(13.41*)	(12.33)	(16.05*)	(12.56)
Model M2: April	16.62			16.31	17.02	16.82	17.26
1980–March 1990	(17.25)			(16.75)	(14.75)	(19.36)	(14.09)
Test T2-3: April	14.13,	6.82[a],	52, 45	13.46*	14.73	15.22*	12.61*
1990–March 1993	15.65	1.53		(14.63)	(14.08)	(17.63)	(13.16)
Test T2-5: April	11.41	16.13*	114, 110	10.75*	12.06*	12.39*	10.36*
1990–March 1995	(13.74)	(3.96*)		(12.92*)	(12.22)	(15.39*)	(11.57*)
Model M3: April	13.93			14.06	14.21	13.38	15.14
1981–March 1991	(17.49)			(19.87)	(15.02)	(20.18)	(12.59)
Test T3-3: April	11.11	7.94*	87, 145	10.51*	9.12*,	11.94*	12.04*
1991–March 1994	(12.35)	(6.19*)		(11.64*)	(12.03*)	(13.34*)	(12.17*)
Test T3-5: April	13.67	0.82	30, 190	13.26*	12.57*	14.11	14.75
1991–March 1996	(11.20)	(8.05*)		(10.58*)	(10.87)	(12.02*)	(11.34)
Model M4: April	16.50			16.60	15.96	16.42	16.07
1982–March 1992	(17.46)			(17.05)	(15.17)	(19.55)	(13.61)
Test T4-3: April	9.03	22.59*	166, 198	8.76*	7.77*	9.42*	9.57*
1992–March 1997	(9.99)	(9.97*)		(9.61*)	(9.17*)	(10.24*)	(11.28*)
Test T4-5: April	13.71	10.61*	62, 193	13.12*	12.91*	13.30*	12.95*
1992–March 1994	(10.63)	(8.74*)		(10.23*)	(9.59*)	(10.95*)	(12.62)
Model M5: April	16.35			13.87	13.81	12.12	14.54
1983–March 1993	(16.73)			(16.35)	(14.21)	(19.29)	(11.80)
Test T5-3: April	13.57	−0.20	35, 198	13.26	15.85*,	13.96	14.11
1993–March 1998	(9.81)	(9.83*)		(9.57*)	(8.97*)	(9.90*)	(11.40*)
Test T5-5: April	17.74	−12.33*	96, 166	17.75*	17.92*	17.98*	16.29
1993–March 1994	(11.96)	(6.18)		(11.61*)	(10.96)	(12.12[a])	(14.09)
Model M6: April	13.59			14.16	12.50	13.69	13.34
1984–March 1994	(16.47)			(16.21)	(14.21)	(18.82)	(11.13)
Test T6-3: April	15.90	−5.62*	67, 180	16.27*	17.54*	15.43*	14.86
1994–March 1997	(10.97)	(7.25*)		(10.67*)	(10.00*)	(11.17*)	(12.55*)
Test T6-5: April	18.22	−9.99*	110, 43	19.51*	20.63*	16.86*	15.21*
1994–March 1999	(15.07)	(1.57)		(14.75)	(14.11)	(15.43)	(17.10)
Model M7: April	13.20			13.25	13.21	13.10	12.77
1985–March 1995	(16.00)			(15.60)	(13.85)	(18.52)	(13.05)

Table 2 (Continued)

Model and matched test periods	All funds			LCG	LCV	SCG	SCV
	Mean return (S.D.)	t-Test diff in mean return (S.D.)	NUM	Mean return (S.D.)	Mean return (S.D.)	Mean return (S.D.)	Mean return (S.D.)
Test T7-3: April 1995–March 1998	24.67 (12.59)	−29.79* (4.27*)	189, 134	25.20* (11.94*)	24.68* (12.25*)	24.33* (13.39*)	22.92* (13.00)
Model M8: April 1986–March 1996	12.71 (18.70)			12.59 (15.26)	12.37 (13.18)	13.60 (18.35)	9.72 (14.77)
Test T8-3: April 1996–March 1999	18.70 (18.07)	−10.79* (−2.01)	121, 45	20.65* (17.06)	23.40* (18.43*)	17.26* (18.72)	10.35 (20.62*)

* Indicate statistically significant differences in mean or standard deviation between model and test periods. Statistical significance is measured at the 1% level.

For 3-year forecasts, the MSE criteria show the Multi-Style (0.012901) and the Dynamic Index (0.012709) models to have the lowest prediction errors. Moreover, the mean squared errors of these two models are significantly different from those of other models but not different from each other. Under the MAE evaluation criteria, the Multi-Style model is the best performer with its mean absolute error of 0.081617 being significantly different from those of all other models. The FF 3-Factor models have the second lowest mean absolute error of 0.083950 followed by that of the Dynamic model. Their mean absolute errors are insignificantly different from each other but significantly different from those of all other models. For 5-year forecasts, the FF 3-Factor model has the lowest MSE and MAE of 0.007413 and 0.068421 respectively. Both error statistics are significantly different from those of all other models. The Multi-Style and the Dynamic Index Models rank next with their mean MSE and MAE being statistically indifferent from each other but different from those of other models.

Among the mean models, for both 3- and 5-year forecasts, the Overall Mean model which is the most aggregate of all models has lower MSE and MAE than the style mean model. However, the Historical model outperforms both Mean models. This indicates that the process of averaging not only eliminates noise but also some useful information contained in the historical correlation matrix. The performance of the Style Index model which is very similar to that of the Single Index model indicates that adding a style index to the market model does not add any more information to the estimates of future correlations. The decomposition of the mean squared error statistics into their relative bias, variance and covariance proportions follows Eq. (9) and is also reported in Table 3. Bias in forecasts which account for approximately 20–40% of the MSE is insignificantly different between various model pairs.

Table 3
Comparison of errors in forecasting correlation of mutual funds by different models[a]

Forecasting models	3-year forecasts					5-year forecasts				
	Mean square error (MSE)	Theil's decomposition			Mean absolute Error (MAE)	Mean square error (MSE)	Theil's decomposition			Mean absolute error (MAE)
		Bias, U_m	Variance, U_v	Co-variance, U_c			Bias, U_m	Variance, U_v	Co-variance, U_c	
A. Historical model										
Full	0.014026	0.319	0.013	0.667	0.085084	0.011838	0.265	0.143	0.593	0.077652
B. Mean models										
Overall	0.017595	0.261	0.230	0.509	0.091904	0.014954	0.207	0.058	0.735	0.083538
Style	0.018056	0.255	0.214	0.531	0.092316	0.015456	0.200	0.172	0.628	0.084415
C. Index models										
Single	0.013853	0.325	0.013	0.660	0.085337	0.011335	0.283	0.156	0.561	0.165794
Style	0.013963	0.320	0.013	0.666	0.085265	0.011476	0.277	0.151	0.572	0.165632
Multi-Style	0.012901	0.349	0.013	0.639	0.081617	0.010712	0.299	0.147	0.554	0.072973
FF 3-Factor	0.013660	0.327	0.013	0.659	0.083950	0.007414	0.278	0.143	0.579	0.068422
Dynamic	0.012709	0.385	0.015	0.598	0.084302	0.010115	0.352	0.207	0.441	0.073631

	Mean (overall)	Mean (style)	Single Index	Style Index	Multi Index	FF 3-Factor	Dynamic
	MSE of 3-year forecasts, MSE of 5-year forecasts						
Historical	1, 1	1, 1	0, 1	0, 0	1, 1	1, 1	1, 1
Mean (overall)		0, 0	1, 1	1, 1	1, 1	1, 1	1, 1
Mean (style)			1, 1	1, 1	1, 1	1, 1	1, 1
Single Index				0, 0	1, 1	0, 1	1, 1
Style Index					1, 1	0, 1	1, 1
Multi Index						1, 1	0, 0
FF 3-Factor							1, 1
	MAE of 3-year forecasts, MAE of 5-year forecasts						
Historical	1, 1	1, 1	0, 1	0, 1	1, 1	1, 1	1, 1
Mean (overall)		0, 0	1, 1	1, 1	1, 1	1, 1	1, 1
Mean (style)			1, 1	1, 1	1, 1	1, 1	1, 1
Single Index				0, 0	1, 1	1, 1	0, 1
Style Index					1, 1	0, 1	0, 1
Multi Index						1, 1	1, 0
FF 3-Factor							0, 1

[a] Statistical significance of MSE and MAE in comparison. (1 indicates significance at the 1% level and 0 otherwise.)

In summary, these results show the Multi-Style Index, FF 3-Factor and Dynamic models as the top contenders for the best forecasting model. The successes of these models reflect the fact that they account for the way mutual fund managers construct their portfolios. Style investing allows fund managers to have their portfolios exposed to specific risk factors. The independent variables in both the Multi-Style Index and the FF 3-Factor models take into account the risk factors of size and value/growth differentials. These results are consistent with Gallo and Lockwood (1997) and Kothari and Warner (1997) who find style analysis to be useful in benchmarking fund returns. The success of the Dynamic model supports Ferson and Warther (1996) with the conditioning of information being important both in statistical and practical terms.

5.3. Inter-temporal consistency of forecasts

Another issue in deciding the efficacy of a forecasting model is its consistency in relative rank across different time periods. Table 4 reports the mean and standard deviation of a model's rank across the various sub-periods over which each model was evaluated. The mean rank across forecasting periods is different from the overall rank based on average MSE and MAE across all matched model-test periods. This along with the dispersion of the ranks around the mean indicates that the relative rank of a model change in each period M1–M8.

The mean rank of the Multi-Style, FF 3-Factor and the Dynamic models is the lowest of all eight models and this is consistent with overall ranks of these models. However, the period-by-period rank of the Dynamic model's rank suffers from high variability. A closer examination of its period by period rank shows that the Dynamic model performs poorly in sub-periods T1–3, T1–5, T2–3, T2–5 and T8–3. Excluding these five sub-periods, the Dynamic model ranks first under the MSE criteria. Thus, the Dynamic model is the best performing model most of the time although its reliability is certainly questionable. On the other hand the Multi-Index model ranks no worse than 3 in most periods and only twice ranks 5. In contrast, the FF 3-Factor model shows the most stability in its inter-temporal rank. For 3-year forecasts, it always ranks either 3 or 4 and for 5-year forecasts it ranks 4 most of the time but occasionally grabs the top finish.

Measuring the overall reliability of the models using the dual characteristic of mean rank and standard deviation of the rank, the Multi-Style-Index model is the best model for both 3- and 5-year forecasts. Its mean rank is the lowest among all models and the standard deviation of its ranks is among the lowest. The FF 3-Factor model is the next best model with its mean rank close to that of the Dynamic model. However, for the FF 3-Factor model the dispersion in rank across the various forecasting time-periods is much lower than those of

Table 4
Mean and standard deviation of ranks for the various forecasting models across different time periods

Forecasting models	Rank based on MSE				Rank based on MAE			
	3 Years		5 Years		3 Years		5 Years	
	Mean (S.D.)	Overall rank ALL	Mean (S.D.)	Overall rank ALL	Mean (S.D.)	Overall rank ALL	Mean (S.D.)	Overall rank ALL
Full Historical	4.50 (2.07)	6	4.50 (2.35)	6	4.50 (2.59)	6	4.00 (2.58)	4
Mean Overall	6.13 (1.96)	7	7.00 (0.00)	7	6.13 (1.31)	7	6.67 (0.00)	5
Mean Style	6.13 (3.18)	8	7.33 (1.03)	8	6.13 (2.71)	8	6.67 (1.55)	6
Single Index	4.63 (1.69)	4	3.67 (1.03)	4	4.63 (1.77)	4	5.00 (1.47)	8
Style Index	5.00 (0.76)	5	4.50 (0.55)	5	5.00 (0.83)	5	5.00 (0.41)	7
Multi-Style	2.63 (1.51)	2	2.33 (0.52)	3	2.63 (1.30)	1	2.17 (0.75)	2
FF-3-Factor	3.63 (0.52)	3	3.33 (1.51)	1	3.38 (1.06)	2	3.33 (1.51)	1
Dynamic	3.38 (3.34)	1	3.33 (3.61)	2	3.75 (3.33)	3	3.50 (3.51)	3

the Dynamic model. [17] These results are in sharp contrast with those from correlation among common stocks in Eun and Resnick (1992). For common stocks, a model analogous to the Mean Style model was the overall winner. The APT inspired, Multi-Style Index model which ranked poorly for stocks (seventh out of eight) was the big winner for correlations among mutual funds. [18]

Historical and mean models which in general do not perform well show tremendous improvement in their relative rank in sub-periods T1–3 and T2–3. The Mean Style, Mean Overall and the Historical models rank 1, 2 and 3, respectively in both sub-periods. Thus, in periods of relative stability between model and test period return distributions the Historical and the mean models, which are parsimonious and very simple to use, generally provide good forecasts of future correlation among mutual funds.

5.4. Inter- and intra-style forecasts

Comparison of forecasting performance for inter- and intra-style funds is in Table 5. In addition to reporting the MSE and MAE averaged across all matched model-test periods the table also shows the mean rank of each model in parentheses. For 3-year forecasts, the Multi-Style model does extremely well in predicting the correlation of funds in both different and similar style categories by finishing first in its mean rank (3.13 (MSE) and 2.88 (MAE) for inter-style; 2.88 (MSE) and 3.13 (MAE) for intra-style). [19] In contrast, the Dynamic model has the lowest MSE and MAE of 0.0162 and 0.0985, respectively in the inter-style sub-sample with its mean rank close to that of the Multi-Style index model's. However, intra-style fund correlation forecasts from the Dynamic model ranks third (MSE) and fourth (MAE) in mean rank trailing in both cases the Single Index model. In contrast, the Single Index model has the opposite performance doing well (second in relative rank)in predicting within style but does poorly (sixth in relative rank)for funds in different style classes. All other models do not exhibit any significant differences in their relative ranks when forecasting performance in either the inter- and intra-style sub-samples.

[17] Similar results were obtained using forecasting horizons of 1 and 10 years. The top two models in those time periods essentially remain the same. The rank of the Multi-Index model is 1 for 1-year forecasts and 2 for 10-year forecasts. The FF 3-Factor model in contrast ranks 3 for 1-year forecasts and 1 for 10-year forecasts.

[18] For stocks the Eun and Resnick (1992) uses only size based indexes in their model analogous to the Multi-Style Index model. In contrast this paper uses indices that are a combination of size and value/growth.

[19] The number of inter-style correlations account for approximately 60% of the total observations.

Table 5
Comparison of errors in forecasting correlation of mutual funds by different models in samples stratified by style

Forecasting models	Inter-style performance				Intra-style performance			
	3-Year forecast		5-Year forecast		3-Year forecast		5-Year forecast	
	MSE (rank)	MAE (rank)	MSE (rank)	MAE (rank)	MSE (rank)	MAE (rank)	MSE (rank)	MAE (rank)
A. Historical model								
Full	0.0188	0.1030	0.0157	0.0928	0.0115	0.0774	0.0094	0.0677
	(4.13)	(4.63)	(3.50)	(4.00)	(4.63)	(4.00)	(4.33)	(4.83)
B. Mean models								
Overall	0.0283	0.1215	0.0230	0.1099	0.0160	0.0870	0.0118	0.0734
	(6.25)	(6.25)	(7.50)	(7.67)	(7.13)	(6.75)	(7.17)	(6.67)
Style	0.0261	0.1160	0.0215	0.1052	0.0142	0.0818	0.0109	0.0695
	(5.75)	(5.75)	(6.83)	(6.67)	(6.50)	(6.50)	(6.17)	(5.33)
C. Index models								
Single	0.0191	0.1038	0.0161	0.0936	0.0101	0.0751	0.0079	0.0641
	(4.63)	(4.38)	(4.67)	(4.17)	(3.13)	(3.63)	(3.67)	(3.67)
Style	0.0191	0.1036	0.0160	0.0934	0.0109	0.0766	0.0088	0.0655
	(4.88)	(4.63)	(5.33)	(4.83)	(3.88)	(4.13)	(4.33)	(4.50)
Multi-Style	0.0176	0.0998	0.0148	0.0897	0.0099	0.0717	0.0080	0.0609
	(3.13)	(2.88)	(2.17)	(2.17)	(2.88)	(3.13)	(2.50)	(2.50)
FF-3-Factor	0.0188	0.1032	0.0155	0.0920	0.0111	0.0761	0.0088	0.0652
	(4.13)	(4.13)	(2.50)	(3.00)	(4.25)	(4.13)	(3.83)	(4.17)
Dynamic	0.0162	0.0985	0.0130	0.0863	0.0107	0.0791	0.0084	0.0691
	(3.13)	(3.38)	(3.50)	(3.50)	(3.63)	(4.00)	(4.00)	(4.33)

In summary, these results indicate that the Multi-Index and the FF 3-Factor models are not only the best performing models but also have the most consistency in their relative ranks when comparing across different forecasting periods and over different forecasting lengths. Moreover, their successes are indifferent in sub-samples of funds in the same or different style categories.

6. Conclusions

Estimates of future correlation coming from the Multi-Style Index and the FF 3-Factor model are very consistent in sub-samples stratified by style and time. The implication of these results is important for managing a portfolio of mutual funds. In contrast, Eun and Resnick (1992) find the aggregate mean model analogous to the Style Mean model in this paper to be the best performer for stocks. The success of the Multi-Style Index model and the FF 3-Factor model agrees with Fama and French (1992) which shows firm size and value/growth differentials as significant priced factors in the US equity markets. These findings are also consistent with Gallo and Lockwood (1997) findings that show trading strategies based on the Multi-Style Index model perform very well. Kothari and Warner (1997) show the FF 3-Factor model to be the best evaluator of mutual fund performance. The success of this model in this study indicates that this model is not only good at explaining the historical correlation structure among mutual funds but also in predicting future correlations. Since most mutual funds actively engage in 'style' investing along size and value/growth differentials, the success of the multi-style index and the FF 3-Factor model perhaps comes as no surprise.

The success of any diversified fund portfolio is also partly dependent on appropriate style classification. This study tracks each fund across the eight model periods. Of the 202 funds only 96 funds remain faithful to their style class. Moreover, Indro et al. (1998) reports that Fidelity Magellan lost two 401(k) plan bids because managers exhibited style drifts. The extent of the problem has led many 401(k) plans to apply pressure on mutual fund complexes to define each portfolio's investing style strictly and maintain them. In response, fund complexes are now hiring 'style cops' to police their manager's adherence to their defined style (Edgerton, 1998). All this implies investors must periodically update their fund's style class.

Future research can continue to experiment with other APT inspired index models (conditional or unconditional) as researchers find success with their use in explaining historical fund returns. Solnik (1974) illustrates the benefit of adding international securities to a portfolio. Several mutual funds have either exclusive or heavy emphasis in foreign investing. Global funds for example may allocate a large part of their portfolio to foreign stocks while regional or country funds invest exclusively in foreign stocks. Future empirical research will have to adapt the various index models to allow their use in forecasting correlation among various 'domestic', 'country' and 'global' funds.

Acknowledgements

Comments from and conversations with Larry Lockwood, Oranee Tawatnuntachai and participants at the FMA Annual Conference in Orlando, Florida are gratefully acknowledged. The author also thanks the Editor Giorgio P. Szegö and two anonymous referees for their suggestions in improving the paper.

References

Ahmed, P., Lockwood, L., 1998. Changes in factor betas and risk premiums over varying market conditions. Financial Review 33 (3), 149–168.
Banz, R., 1981. The relationship between returns and market values of stocks. Journal of Financial Economics 9, 3–18.
Basu, S., 1977. Investment performance of common stocks in relation to their price–earnings ratio: A test of the efficient market hypothesis. Journal of Finance 32, 663–682.
Basu, S., 1983. The relationship between earnings yield, market value, and return for NYSE common stocks. Journal of Financial Economics 12, 129–156.
Brown, S.J., Goetzman, W.N., 1997. Mutual fund styles. Journal of Financial Economics 43, 373–399.
Brown, S.J., Goetzman, W., Ibbotson, R.G., Ross, S.A., 1992. Survivorship bias in performance studies. Review of Financial Studies 5, 553–580.
Carlton, W.T., Lakonishok, J., 1986. The size anomaly: Does industry group matter? Journal of Portfolio Management 12, 36–40.
Chan, L.K., Hamao, Y., Lakonishok, J., 1983. Fundamentals and stock returns in Japan. Journal of Finance 46, 1739–1789.
Christopherson, J.A., 1995. Equity style classifications. Journal of Portfolio Management 21, 32–43.
diBartolomeo, D., Witkowski, E., 1997. Mutual fund misclassification: Evidence based on style analysis. Financial Analysts Journal 53 (5), 32–43.
Edgerton, J, 1998. Who's that peering over your fund manager's shoulder? It's the fund style cop. Money 27 (4), 49.
Elton, E., Gruber, M., Urich, T., 1978. Are betas best? Journal of Finance 39, 1375–1384.
Elton, E., Gruber, M., 1973. Estimating the dependence structure of share prices – implications for portfolio selection. Journal of Finance 28, 1203–1232.
Elton, E., Gruber, M., 1995. Modern Portfolio Theory and Investment Analysis, fifth ed. Wiley, New York.
Eun, C.S., Resnick, B.G., 1984. Forecasting the correlation structure of international share prices. Journal of Finance 39, 1311–1324.
Eun, C.S., Resnick, B.G., 1992. Forecasting the correlation structure of share prices: A test of new models. Journal of Banking and Finance 16, 643–656.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.

Fama, E.F., French, K.R., 1993. Common stock risk factors in returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Ferson, W.E., Warther, V.A., 1996. Evaluating fund performance in a dynamic market. Financial Analysts Journal 52 (6), 20–28.

Gallo, J.G., Lockwood, L., 1997. Benefits of proper style classification of equity portfolio managers. Journal of Portfolio Management 23 (3), 47–55.

Grinblatt, M., Titman, S., 1989. Mutual fund performance: An analysis of quarterly portfolio holdings. Journal of Business 62, 393–416.

Indro, D.C., Jiang, C.X., Hu, M.Y., Lee, W.Y., 1998. Mutual fund performance: A question of style. Journal of Investing 7 (2), 46–53.

Keim, D.B., 1983. Size-related anomalies and stock return seasonality. Journal of Financial Economics 12, 13–32.

Kothari, S.P., Warner, J., 1997. Evaluating mutual fund performance. Working Paper, University of Rochester.

Lamoureux, C.G., Sanger, G.C., 1989. Firm size and turn-of-the-year effects in the OTC/NASDAQ market. Journal of Finance 44, 1219–1245.

McGough, R., 1998. When hunting for performance, don't forget about diversifying. Wall Street Journal, 6 July, R17.

O'Neal, E.S., 1997. How many mutual funds constitute a diversified mutual fund portfolio? Financial Analysts Journal 53 (2), 37–46.

Reinganum, M., 1981. Misspecification of capital asset pricing: Empirical anomalies based on earnings yields and market values. Journal of Financial Economics 9 (1), 19–46.

Reinganum, M., 1990. Market microstructure and asset pricing: An empirical investigation of NYSE and NASDAQ securities. Journal of Financial Economics 28, 127–148.

Rosenberg, B., Reid, K., Lanstein, R., 1985. Persuasive evidence of market inefficiency. Journal of Portfolio Management 11 (3), 9–17.

Ross, S., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.

Sharpe, W.F., 1992. Asset allocation: Management style and performance measurement. Journal of Portfolio Management 18 (2), 7–19.

Solnik, B., 1974. Why not diversify internationally rather than domestically. Financial Analysts Journal, July–August, 48–54.

Theil, H., 1971. Applied Economic Forecasting. North-Holland, Amsterdam.

Trzcinka, C.A., 1995. Equity style classifications: A comment. Journal of Portfolio Management 21, 44–46.

Tierney, D.E., Winston, K., 1991. Using generic benchmarks to present manager styles. Journal of Portfolio Management 17, 33–36.