A STATISTICAL COMPARISON OF THE CAPM TO THE FAMA-FRENCH THREE FACTOR MODEL AND THE CAHART'S MODEL

Zakri Y. Bello[1]

belloz@ccsu.edu

Central Connecticut State University, USA.

ABSTRACT

The goal of this study is to compare the CAPM to the Fama-French (FF) Three Factor Model and to Carhart's extension of the FF Model with regard to (1) statistical goodness of fit, and (2) the quality of prediction. My sample consists of actively managed domestic equity mutual funds and the sample period is April 1986 to March 2006. My results indicate that each of the three regression lines explains about 71% of equity fund returns. Thus, with respect to the statistical goodness of fit, the difference between the three models is not significant. However, with respect to the quality of prediction, the FF Three Factor Model is a remarkable improvement over the CAPM, and the Carhart Model is a significant improvement over the FF Model. I do not find any evidence of harmful collinearity in my analyses.

Key words: Asset Pricing Models, Statistical Goodness of Fit, Model Specification, Multicollinearity

JEL Classification: G11, G12, C12, C13, C25.

[1] Contact Details: 1615 Stanley Street, New Britain, CT 06050, Tel: (860) 832-3262

I. INTRODUCTION

> "The goal of statistical model specification is to assist in understanding the system from which the data is extracted, to learn which regression terms are important, and to learn which model is the best in terms of prediction" and that "an adroit job of model selection does not require that one actually locate the correct model, and that the correct model may indeed never be found." (Myers, 1990).

The purpose of this study is to compare the specification of three related asset pricing models in terms of (1) the statistical goodness of fit, and (2) the qualities of prediction. I use several traditional methods of model selection and nontraditional model selection criteria which are prediction oriented. The asset pricing models that I focus on are those that specify risk in microeconomic terms, i.e. those that use the characteristics of the underlying sample of securities. Connor (1995) compares the explanatory power of the three types of multifactor models of asset returns, including a macroeconomic factor model, a fundamental factor model, and a statistical factor model. He finds that the statistical factor model and the fundamental factor model substantially outperform the macroeconomic factor model, and that the fundamental factor model outperforms the statistical factor model. I concentrate on the three most popular fundamental factor models, including the single factor CAPM, the Fama-French Three Factor Model (Fama and French, 1993 and 1996), and Carhart's extension of the Fama-French Three Factor Model (Carhart, 1997).[2]

The CAPM posits that the variation in security returns is the only relevant source of a security's systematic risk, and that a properly selected proxy of the market portfolio can be used to estimate this systematic risk. Therefore the risk premium on an individual security or on a portfolio of securities is a function of systematic risk as measured by the beta on the relevant benchmark index. Fama and French (1993) extended the CAPM into a three factor model whereby the risk premium on a security is a function of the systematic risk as measured by the betas on three factors including the CAPM's market portfolio, a portfolio that represents the difference in returns of small versus large firms (SMB) and a portfolio that represents the difference in returns of firms with high versus low book-to-market value ratios (HML).[3]

[2] Daniel and Titman (1997) found that the return premia on small capitalization and high book-to-market ratio stocks do not arise because of the comovements of the stocks with the factors. That it is the characteristics, rather than the covariance structure of returns, that explain the cross-sectional variation in stock returns.

[3] Fama and French (1993) and Lakonishok, Shleifer and Vishney (1994) discussed the ability of the Fama and French Model to explain security returns and offered alternative interpretations of the two factors, SMB and HML, as to whether the related patterns of returns are consistent with market efficiency.

The Fama and French risk-return framework is further extended by Carhart (1997) who introduced a price momentum factor as the fourth systematic risk factor. The price momentum factor represents the tendency of firms with negative past returns to earn negative future returns, and for firms with positive past returns to earn positive future returns. The Fama and French Model is estimated using statistical regression as follows:

$$r_{jt} - r_{ft} = \alpha_j + \beta_{j1}(r_{mt} - r_{ft}) + \beta_{j2}(SMB_t) + \beta_{j3}(HML_t) + \varepsilon_{jt} \qquad (1)$$

Where,

r_{jt} = the realized return on security j during time period t;

r_{mt} = the realized return on the market during period t. I obtained the series of realized excess returns on the market, (r_{mt} – r_{ft}), from Ken French's Website[4] where it is described as the value weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate.

r_{ft} = the nominal risk-free rate during time period t, represented here by the monthly yields on three-month Treasury bills.

α_j = the intercept , predicted by the Arbitrage Pricing Model to be equal to zero;

β_{j1} to β_{j3} = factor betas on the three risk factors including the excess return on the market, SMB, and HML;

ε_{jt} = the residual excess return on portfolio j during time period t;

SMB_t = the difference in returns on small firms versus large firms during time period t; and

HML_t = the difference in returns of firms with high book-to-market value (B/M) ratios versus the returns of firms with low B/M ratios.

Carhart's (1997) extension of the Fama and French Three Factor Model is as follows:

$$r_{jt} - r_{ft} = \alpha_j + \beta_{j1}(r_{mt} - r_{ft}) + \beta_{j2}(SMB_t) + \beta_{j3}(HML_t) + \beta_{j4}MOM_t + \varepsilon_{jt} \qquad (2)$$

Where the price momentum factor (MOM) is the average return on two-high-prior-return portfolios minus the average return on two-low-prior-return portfolios. That is,

---

[4] Http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/.

the average return on securities with the best return performance during the past year less the average return on securities that had the worst return performance.

II. STATISTICAL METHODOLOGY

> "There is some consensus that models are metaphors, or windows, through which researchers view the observable world, and that their adoption depends not upon whether they can be deemed to be 'true' but rather upon whether they can be said to (a) correspond to the facts, and (b) be useful" (Kennedy, 2003, p. 81).

> "The analyst must accept that linear models are merely empirical approximations and that several models can be fit that would be nearly equal in effectiveness." (Myers, 1990, p. 164).

The standard criteria for comparing models include the coefficient of determination, R^2, and an estimate of the error variance, S^2. However, R^2 is the measure of the model's capability to fit the data and is not prediction oriented (Myers, p. 100). Never-the-less,I use the adjusted R-squared together with an alternative measure, the Amemiya's criterion, to compare the goodness of fit of the three asset pricing models, including the CAPM, the Fama and French (FF) Three Factor Model, and Carhart's extension of the FF Model.

The estimate of the error variance, S^2 (i.e. the residual mean square), can provide valuable information for selecting the best model for prediction and is used in this study both for assessing the goodness of fit and prediction. S^2 is also used in the calculation of the standard errors of coefficients for hypothesis testing. I use both the R^2 and S^2 to compare the three asset pricing models. The model with smallest S^2 or largest R^2 is of course preferable. Additionally, we use less traditional criteria for comparing models, including the PRESS (i.e. the prediction sum of squares) statistic and Mallow's C_p statistic, both of which are prediction oriented. The PRESS statistic is calculated as follows:

$$PRESS = \sum_{i=1}^{n} (Y_i - \hat{Y}_{i,-i})^2 \quad (3)$$

Where, Y_i is the response and $\hat{Y}_{i,-i}$ (i = 1, 2, …, n) is the prediction, calculated by removing the first observation , then the second, then the third, et cetera, and each time fitting the model using the remaining observations, and then estimating the first observation (i.e. $\hat{Y}_{i,-i}$), then the second observation, et cetera. The PRESS residuals are then calculated as $(Y_i - \hat{Y}_{i,-i})$. Using the set of PRESS residuals, I calculated the PRESS

Statistic as shown in equation (3). The model with the smallest PRESS statistic is of course preferable.

When comparing models that differ with regard to the number of regressors, a compromise must be made between a biased model and one that has inflated variance. It is known that an underfitted (or short) model will have poor fit (i.e. will be biased) but will have low variance of prediction and the S^2 produced will overestimate the true population error variance, leading to problems in hypothesis testing (Myers, page 88). In other words, an underfitted model will produce bias in the prediction, $(\hat{Y})$, in the regression coefficients, and in the estimate of the variance, S^2 (i.e. the residual sum of squares that reflects the bias in prediction). On the other hand, an overfitted (or long) model is well fitted, but will have high variance of prediction, high multicollinearity, and regression coefficients that are too large. In other words, a model with fewer regressors produces biased coefficients and biased prediction, whereas a model with more regressors produces large variances in the coefficients and in prediction. Moreover, it is known that the additional variance produced by the addition of variables depends largely on multicollinearity introduced by the added regressors.

The choice of a suitable set of regressors is therefore one that strikes a balance between bias and variance. In this respect, the Mallow's C_p is considered an appropriate compromise. The Mallow's C_p is calculated as follows:

$$C_p = p + \frac{(S^2 - \hat{\sigma}^2)(n-p)}{\hat{\sigma}^2} \quad (4)$$

Where,

P=the number of estimated parameters, including the intercept;

$\hat{\sigma}^2$ =the residual mean square for the model with the greatest number of terms (i.e. the most complete

model); and

S^2= estimate of the error variance. This is calculated as: $S^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-p}$, where n is the number of data points, p is number of parameters including the intercept, and $Y_i - \hat{Y}_i$ (i= 1,2, …, n) are the ordinary residuals, i.e. the observed errors of fit.

The C_p statistic is a measure of total error of prediction, which includes (1) bias, equivalent to the difference between C_p and P, and (2) the variance of prediction. Equivalently, the C_p statistic is measured as follows:

$$C_p = \sum_{i=1}^{n} \frac{\{[Variance \hat{Y}_i] + [Bias \hat{Y}_i]^2\}}{\sigma^2} \quad (5)$$

In this study, I use the Mallow's C_p, the PRESS statistic, and the S^2 to select the model that performs best from the prediction standpoint. I assess statistical fit using the traditional measure of fit, the adjusted R^2, which is often used as an alternative to the Akaike Criterion, Amemiya's Criterion and the Schwartz Criterion (Kennedy, 2003, p. 117). The Amemiya's Criterion is used in this study to supplement the adjusted R^2.

III. THE DATA

The sample consists of actively managed domestic equity mutual funds from the following investment objective categories: Aggressive Growth, Growth, Growth and Income, Equity Income, and Small Company. Index funds, funds of funds, master feeder funds, and money market funds are not included. Since most of the domestic-equity- mutual funds tend to have substantial holdings of foreign stocks and bonds, I selected only those funds that have at least 50% of their portfolios values invested in domestic stocks and no more than 15% invested in foreign stocks. I also selected funds that have no more than 15% of their portfolios invested in bonds. The sample period is April 1986 to March 2006.

Monthly mutual fund returns and monthly yields on three-month Treasury bills were obtained from the Morningstar Principia database. Monthly excess return on the market, monthly Fama-French Factors (SML and HMB), and the monthly Momentum Factor (MOM), were obtained from Ken French's Web site.[5]

A profile of the sample is shown in Table 1. The size of the average equity mutual fund in the sample, as measured by net assets, is about $1.9 billion. The standard deviation of net assets is relatively large at about $3.9 billion, suggesting that the sample consists of mutual funds of diverse sizes. Domestic stocks make up about 92% of the average fund's portfolio value, and foreign stocks and bonds make up about 4% and 0.07% of the average fund's portfolio, respectively. The sample truly consists of actively managed funds as indicated by the portfolio turnover of about 70%, which suggests that the average fund in this sample buys and replaces its assets in about 17 months . Moreover, the average fund invests about 27% of its portfolio in the top-ten companies it holds, as judged by the average "Top-Ten" shown in the Table.

[5] Http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/.

Table 1
Sample Profile of Domestic Equity Mutual Funds
(April 1986 – March 2006)

Variable	N	Mean	Std. Dev.
Net Assets ($m)	489	1856.970	**3921.510**
Domestic Stock (%)	628	92.297	**5.686**
Foreign Stock (%)	628	4.431	**3.824**
Bonds ($m)	628	0.069	**0.594**
Top-Ten (%)	628	27.337	**9.962**
Turnover (%)	**628**	**69.608**	**48.257**

Variable	Observations	Mean	Std. Dev.
Treasury Bill	126524	0.368	**0.161**
MKT	126524	0.614	**4.262**
SMB	126524	-0.054	**3.684**
HML	126524	0.494	**3.335**
MOM	**126524**	**0.854**	**4.737**

Note: N is the number of mutual funds with non-missing data, and "Std. Dev." is short for standard deviation. Top-Ten is the percentage of the mutual-fund portfolio invested in the top-ten companies held by the average mutual fund. The variables MKT, SMB, HML, and MOM are as defined in Equation (1) and Equation (2).

IV. RESULTS

Assessing Multicollinearity

In the presence of multicollinearity, the variance of the estimated parameters is quite large, parameter estimates are unreliable and hypothesis testing has little power. To detect if the regressors of the FF and Carhart models are collinear, I computed variance inflation factors as well as condition numbers. The variance inflation factor (VIF) is calculated as follows:

$$VIF = (1 - R_i^2)^{-1},$$

where R_i^2 is the coefficient of multiple determination that is obtained when a particular independent variable is regressed against the other independent variables in the equation. As the R_i^2 approaches unity, the VIF gets quite large, indicating that accuracy in estimating the coefficient is decreasing. A VIF greater than 10 indicates harmful collinearity (Kennedy 2003, page 213).

The condition number of the correlation matrix of the independent variables is an alternative indicator of collinearity. Instability in the independent variables is evidenced by a large condition number (CN). When a CN exceeds 1000, one should be concerned with the problems caused by multicollinearity (Myers 1990, page 370).

The variance inflation factors shown in Table 2 are all substantially less than 20, and the condition numbers are quite small. This suggests that harmful collinearity is not present in both the FF and Carhart equations.

FIT OF THE REGRESSION LINES

The coefficient of determination (the R squared), measures the proportion of the variation in the dependent variable that is explained by the regression equation. As the R squared approaches unity, ordinary residuals all approach zero and the fit of the regression equation to the data approaches the ideal. Statisticians agree that the R squared is a dangerous criterion for comparison of models (Myers 1990, page 38, and Kennedy 2003, pages 90-106) because any additional regressor will cause the R squared to increase, and because R squared does not imply superior prediction equation. And although the adjusted R-squared is supposed to correct the bias in the R-squared as an estimate of the population R-squared, it is known that this is in fact not true. The adjusted R-squared is not known to be a better estimator of the population R-squared than the traditional unadjusted R-squared. Nevertheless, the adjusted R-squared is used in this study instead of the unadjusted R-squared to supplement the other statistical measures of model fit.

The results shown in Table 2 indicate a minimal change in adjusted R squared when the two FF factors (SMB and HML) are added to the CAPM and when the momentum the factor

(MOM) is added to the FF model. The increase in R squared is less than 1% in each case.[6] These results are not greatly altered by the Amemiya's Criterion, the alternative measure of statistical goodness fit. The results suggest that the set of regressors in each of the three equations explain about 71% of the variability of equity-mutual-fund returns during the 1986 to 2006 sample period. Moreover, all of the estimated parameters are statistically significant at the 1% significance level. In comparison, Davis, Fama and French (2000) found that the FF model explains greater than 91% of the variability of stock returns, and Carhart (1997) finds that the inclusion of a momentum factor into the FF model increases R squared by about 15%.

Table 2
Statistical Properties of Asset Pricing Models

Variable	CAPM	FF Model		Cahart's Model	
	Coefficient	Coefficient	VIF	Coefficient	VIF
Intercept (α_0)	0.4107	0.3484	0.0000	0.3355	0.0000
	(53.56)*	(44.44)*		(42.00)*	
r_m (β_1)	0.9718	0.9962	1.3128	0.9982	1.3319
	(556.60)*	(502.91)*		(500.41)*	
SMB (β_2)		0.1043	1.2498	0.1055	1.2549
		(45.73)*		(46.18)*	
HML (β_3)		0.1064	1.5873	0.1069	1.5880
		(37.47)*		(37.66)*	
MOM (β_4)				0.0134	1.0229
				(8.33)*	
Measures of goodness of fit and of prediction.					
R_a^2	0.7100	0.7156	---	0.7158	---
Amemiya's Crt.	0.2900	0.2844	---	0.2842	---
C_p	2568.0620	72.3278	---	5.0000	---
PRESS	7.2875	7.1473	---	7.1438	---
S^2	2.6995	2.6733	---	2.6725	---
Bounds on CN	1, 1	1.03 , 4.12	---	1.59, 20.79	---

Note: "Variable" definitions are as given in equation (1) and equation (2). FF Model refers to the Fama-French Three Factor Model, and VIF denotes the variance inflation factor. The t values are in parentheses. R_a^2 is the adjusted R-squared, Amemiya's Crt is the Amemiya's Criterion and C_p is the Mallow's C_p statistic. PRESS is calculated as defined in equation (3) divided by the number of data points (n). S^2 is the estimate of the error variance, and "bounds on CN" are the bounds on condition number. *Statistically significant at the 1% significance level.

PREDICTION ORIENTED CRITERIA

[6] Kleinbaum, Kupper, muller, and Nizam (1998, page 120) point out that R^2 always increases as more and more variables are added to a model , but a very small increase in R^2 may be neither practically nor statistically important.

The best model for prediction can be selected using the Mallow's C_p, the PRESS statistic, or the estimate of the error variance, S^2. Table 2 shows that the Mallows C_p statistic declines remarkably, from 2568 to 72, when the FF factors are added to the CAPM equation, and declines substantially when the momentum factor is added to the FF equation. These results suggest that both the FF equation and the Carhart's equation are superior to the CAPM equation in predicting stock-mutual-fund returns.

Further, it is well known that the smaller the PRESS statistic, the better the equation in terms of prediction. Similarly, the smaller the error variance, S^2, the more predictive the equation is. The PRESS statistic and the error variance, shown in Table 2, both decline as regressors are added to the CAPM equation, in support of the evidence provided by the C_p statistic. According to the C_p statistic, Carhart's equation has zero bias of prediction as measured by the difference between the calculated C_p statistic and the number of estimated parameters.

The implications of these results are that the managers of domestic mutual funds, other institutional investors, and individual investors ought to consider using either the Fama and French Three Factor Model or the Cahart's Model , when measuring the risk adjusted performance of mutual funds, instead of measures that are based on the CAPM Model. This is because the Fama and French Three Factor Model and the Cahart's Model are superior in terms of statistical goodness of fit and better in predicting mutual funds returns.

V. SUMMARY AND CONCLUSIONS

The goal of this study is to compare the specification of three related asset pricing models with regard to (1) statistical goodness of fit, and (2) the quality of prediction. I compare the CAPM to the Fama and French (FF) Three Factor Model and to Carhart's extension of the FF Model. My sample consists of actively managed domestic equity mutual funds and the sample period is April 1986 to March 2006.

My results indicate that the set of regressors in each of the three equations explains about 71% of equity fund returns. Thus, with respect to the statistical goodness of fit, the difference between the three models is not significant. However, with respect to the quality of prediction, the FF Three Factor Model is a remarkable improvement over the CAPM, and the Carhart Model is a significant improvement over the FF Model.

Because multicollinearity among the regressors results in unreliable parameter estimates and hypothesis testing that has little power, I tested for evidence of harmful collinearity in my analyses using both variance inflation factors and condition numbers. Both of these indicators suggest that harmful collinearity is not present in my analyses.

REFERENCES

Carhart, Mark M., 1997, On The Persistence in Mutual Fund Performance, The Journal of Finance 52, No. 1, 57 – 82.

Conner, Gregory, 1995, The Three Types of Factor Models: A Comparison of Their Explanatory Power, Financial Analysts Journal, 42 – 46.

Davis, L. D., E. F. Fama, and K. R. French , 2000, Characteristics, Covariances, and Average Returns: 1929 to 1997, The Journal of Finance 55, No. 1, 389 – 406.

Daniel, K., and S. Titman, 1997, Evidence on the Characteristics of Cross Sectional Variation in Stock Returns, The Journal of Finance 52, No. 1, 1 – 33.

Fama E. F., and K. R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, Journal of Financial Economics 33, 3 – 56.

Fama, E. F., and K. R. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, The Journal of Finance 51, No.1, 55 – 84.

Kennedy, Peter, 2003. A Guide to Econometrics, 5th Edition, The MIT Press, Cambridge, Massachusetts.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., and Nizam, A., 1998, Applied Regression Analysis and Other multivariate Methods, 3rd Edition, Duxbury Press (Pacific Grove, CA).

Lakonishok, J., A. Shleifer, and R. W. Vishny, 1994, Contrarian Investment, Extrapolation, and Risk, The Journal of Finance 49, No. 5, 1541 – 1578.

Myers, Raymond H., 1990, Classical and Modern Regression with Applications, PWS-Kent Publishing Company, 2nd Edition (Boston, Massachusetts).