On Selection Biases in Book-to-Market Based Tests of Asset Pricing Models

William J. Breen and Robert A. Korajczyk

Northwestern University

Revised: July 1995

Direct correspondence to:

Robert A. Korajczyk
Kellogg Graduate School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL 60208-2006
Telephone: (708) 491-8336
FAX: (708) 491-5719
Internet: r-korajczyk@nwu.edu

We would like to thank Fischer Black, Stephen Brown, Eugene Fama, Kenneth French, Robert Holthausen, Ravi Jagannathan, Josef Lakonishok, Jianping Mei, Richard Sloan, and seminar participants at University of British Columbia, Columbia University, Cornell University, Loyola University, Massachusetts Institute of Technology, Michigan State University, New York University, Northwestern University, University of Pennsylvania, Rice University, University of Wisconsin, the Conference on Financial Economics and Accounting, the CRSP Seminar, and the Western Finance Association, for helpful comments.

On Selection Biases in Book-to-Market Based Tests of Asset Pricing Models

Abstract

Many studies have documented portfolio strategies that provide returns in excess of those expected, given the level of risk of the portfolio. Variables that seem to have predictive power for equity returns include the market capitalization of the firm's equity and the ratio of the firm's book equity to market equity (BE/ME). Firms with low market capitalization and high book-to-market values seem to earn high returns. With respect to the book-to-market anomaly, it has been argued that the apparent superior performance is due to a subtle selection bias in a typical data source used to implement the tests of asset pricing models, the COMPUSTAT data. We use a sample of COMPUSTAT data that is free from this bias to investigate whether the previous evidence on the book-to-market anomaly is an artifact of this selection bias. The postulated selection bias does not seem to be important for samples restricted to NYSE/AMEX firms. There is some difference when NASDAQ firms are included in the standard COMPUSTAT sample. This may be due to a truly stronger BE/ME effect or to a more severe selection bias in that sample. Our data do not allow us to disentangle these two possible explanations.

Standard asset pricing models imply that assets' risk premia are determined by their sensitivity to innovations in investors' marginal utility. In the Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965), Mossin (1966), Treynor (1961) and Black (1972), investors' marginal utility is determined by the rate of return on the market portfolio (the portfolio of all assets, weighted by their relative market values). Thus, assets' risk premia are determined by their sensitivity to unexpected movements in the market portfolio. For asset i, this sensitivity is measured by $\beta_i = \mathrm{cov}(r_i, r_m)/\mathrm{var}(r_m)$, where r_i is the return on asset i, r_m is the return on the market portfolio, $\mathrm{cov}(\cdot,\cdot)$ denotes covariance, and $\mathrm{var}(\cdot)$ denotes variance. The predicted relation between expected returns and β is:

$$E(r_i) = \gamma_0 + \gamma_1\, \beta_i \tag{1}$$

where $E(\cdot)$ denotes the expectation, γ_0 is the zero beta (or riskless) rate of return, and γ_1 is the risk premium for market risk. The CAPM implies that there is a linear relation between expected returns and β and that, after controlling for β, no other variable should be able to explain differences in assets' expected returns. That is, if we estimate a cross-sectional regression of asset returns on β and on a vector of other variables, say Z:

$$r_i = \gamma_0 + \gamma_1\, \beta_i + \delta Z_i + \varepsilon_i \tag{2}$$

then δ should be equal to zero.

Over the past twenty years, the CAPM has been subjected to an enormous amount of empirical scrutiny. A number of studies have documented variables that seem to predict differences in asset returns in excess of those expected given the differences in asset betas. That is, δ is not zero in (2). For example, market capitalization [Banz (1981)], price to earnings (P/E) ratios [Basu (1983)], book equity to market equity (BE/ME) ratios [Stattman (1980)], dividend yield [Keim (1985)], and leverage [Bhandari (1988)] are among the variables that appear to be useful in constructing portfolios that earn high returns even after adjusting for risk. In addition, finding estimates of γ_1 that are statistically significantly different from zero has proven difficult [Tiniç and West (1984)].

Fama and French (1992a) investigate the ability of a number of variables to explain cross-sectional differences in returns earned by stocks. They conclude that the standard measure of non-diversifiable risk, β, has no explanatory power for returns once one controls for differences in size and book-to-market equity (BE/ME). Fama and French (1992a) argue, quite forcefully, that this accumulation of anomalous results calls into question the usefulness of the CAPM, as typically implemented. The arguments of Fama and French (1992a) have elicited much interest, both on the part of academics and of practitioners.

We focus on an issue raised by Kothari, Shanken, and Sloan (1995). They argue that a portion of the apparent ability of BE/ME ratios to predict risk-adjusted returns is due to a selection bias induced by the manner in which data are included on the COMPUSTAT data files. These files include historical accounting data for a wide sample of firms. Many published studies use this source of accounting data when constructing many of the variables that seem to be able to explain cross-sectional differences in asset returns, such as BE/ME, P/E ratios, and leverage ratios. Kothari, Shanken, and Sloan (1995) argue that there are two related potential sources of bias.

Regarding the first source of bias, they argue that when COMPUSTAT adds a firm to its data file, it often "back-fills" data. That is, if a firm is added in 1983, for example, COMPUSTAT might fill in data for the firm back to 1978. Consider the high BE/ME firms that are not on the COMPUSTAT file in 1978. These firms have low market value relative to book value and, therefore, are likely to be performing poorly. Some will become more financially distressed and disappear. These are unlikely candidates for addition to the COMPUSTAT file in 1983. Other firms will overcome their financial distress, i.e., have high returns. These are more likely to be added to the COMPUSTAT file, ex post.

If one uses the current versions of the COMPUSTAT data in testing (2), one would include the high BE/ME firms that subsequently did well in ones regressions for 1978. The high BE/ME firms that were not on COMPUSTAT in 1978 and did poorly in later years (so they were not added later) would be excluded from our tests. This induces a upward bias to the measured relation between returns and book-to-market ratios. A valid testing strategy using the COMPUSTAT data in 1978 would have excluded all of the securities without COMPUSTAT data, not just the ones that later performed poorly.

Kothari, Shanken, and Sloan (1995) argue that there is a second source of bias for firms on the COMPUSTAT database that become financially distressed. They may stop reporting financial results (for example due to bankruptcy). Those that recover from the financial distress may report financial data, retroactively, for the non-reporting period. This imparts a bias toward having data in the database for firms that ultimately recover from distress.

Kothari, Shanken, and Sloan (1995) provide two sets of evidence regarding their hypothesis. The first piece of evidence is regression results, as in (2), using an alternative source for book-to-market ratios. Since they do not have a COMPUSTAT sample purged of the selection bias, they use BE/ME ratios and share prices for approximately 100 industries reported in the *S&P Analyst's Handbook*. While there is likely to be some survivorship bias in the sample, they argue that the bias will be small and would, in fact, be a bias in favor of finding a significant relation between BE/ME and subsequent returns.

Kothari, Shanken, and Sloan (1995) find that there is no statistically significant cross-sectional relation between returns and BE/ME using the S&P industry data. At the same time there is a significant relation between returns and BE/ME using the COMPUSTAT data to form portfolios similar in composition to the S&P industry data. Given that the COMPUSTAT industry results are significant, they argue that the lack of significance of BE/ME using the S&P industry data is unlikely to be due solely to using more highly aggregated industry portfolios.

The second set of evidence offered by Kothari, Shanken, and Sloan (1995) is a comparison of the average returns of firms available on the data files from the Center for Research in Security Prices (CRSP) and on COMPUSTAT to the average returns of firms on CRSP but *not* on COMPUSTAT. The average annual return on the COMPUSTAT sample is higher than the average annual return on the non-COMPUSTAT sample by 1.9% (which is marginally significant). This is true even though the COMPUSTAT sample consists of larger firms, on average, than the non-COMPUSTAT sample. Historically, larger firms have yielded lower returns than smaller firms. Also, the fraction of firms delisted in any given year is much smaller in the COMPUSTAT sample than in the non-COMPUSTAT sample. While these observations are indirect evidence, they are consistent with the selection bias hypothesis.

Fama and French (1993) construct book-to-market based portfolios in a manner designed to minimize the effects of the hypothesized selection bias. They require firms to have two years of COMPUSTAT data before they are eligible for inclusion in the book-to-market portfolios in subsequent years. They argue that COMPUSTAT rarely includes more than two years of historical data when it adds a firm to the database. Even with the two-year data requirement, they find that a book-to-market factor has significant explanatory power for the cross-section of average asset returns.

The potential for selection biases creeping into the results of tests of asset pricing models and portfolio trading strategies is sufficiently important to warrant a further look at the data. In this paper, we investigate book-to-market-based tests of asset pricing models using a COMPUSTAT sample free of selection biases. We have COMPUSTAT data which were collected month by month from January 1974 to the present. For each month, data which are on that month's COMPUSTAT file are saved. Only firms that actually had data on the file, at the portfolio formation date, are eligible for inclusion in our tests. No back-filled data are used in portfolio construction or tests. Therefore, we are able to shed light on the size of the book-to-market effect by providing estimates that are free of selection biases.

A variety of experimental designs have been used in the literature to test asset pricing models. For example, the cross-sections in (2) could be individual assets or portfolios formed by sorting on some asset-specific characteristic. An alternative to cross-sectional regression tests, as in (2), are time-series based tests. The time-series based tests regressions of asset or portfolio excess returns on the excess returns on a market proxy portfolio yield intercepts (Jensen alphas) that can be interpreted as deviations of average returns from those predicted by the asset pricing model. The effects of the postulated selection bias will not be independent of the experimental designs chosen for the tests. For example, cross-sectional regression tests [like (2)] using individual assets or portfolios formed by ranking on the variable inducing the selection bias will be more affected by the bias than tests using portfolios formed by ranking on variables independent of the bias. In particular, since the firms subject to the postulated bias will tend to be smaller firms, tests based on value-weighted portfolios formed by ranking on variables independent of the bias will be less affected by the postulated bias. In our cross-sectional tests we use individual assets. This follows the design used in Fama and French (1992a).

There is a large literature suggesting that many "yield-like" variables (that is, variables with price or market value in the denominator) can explain the cross-section of stock returns [Hawawini and Keim (1992)]. Some of these results, such as book-to-market, earnings-price, and leverage related anomalies, may be subject to selection biases similar to those discussed in Kothari, Shanken, and Sloan (1995). Others are less likely to be subject to the alleged selection biases, such as dividend yields (dividend-price ratios). In fact, Miller and Scholes (1982) find that the inverse of the stock price has significant explanatory power. While selection biases may not explain all of the book-to-market effect, it is important to assess what fraction, if any, seems to be due to biases.

In Section I, we describe our data set and the methods used to estimate and test the asset pricing model. Our empirical results are presented in Section II. We provide a summary and concluding remarks in Section III.

I. Data and Statistical Methods

A. Accounting Data from COMPUSTAT Collected in Real-Time

Our sample covers the period from January 1974 through December 1992. The sample of firms analyzed from January 1974 through March 1992 are those included in the COMPUSTAT Primary/Supplemental/Tertiary file in each of the respective months. Beginning in April 1992, additional firms on the COMPUSTAT OTC file are also included.

These files are updated monthly by COMPUSTAT. On the last business day of each month, COMPUSTAT mails a version of the current data to its subscribers, via overnight delivery. Therefore, a client would generally have a version of the data by the first business day of the month. Selected items were saved from each month's COMPUSTAT file from January 1974 to December 1992. Because of this method of constructing the dataset, firms subsequently added to the tape with back-filled data will not appear on our earlier months' data files. Firms will only appear in the COMPUSTAT file in the month that they were actually added to the database. Therefore, we avoid the selection bias induced by using back-filled data.

Since an investor would not have access to the COMPUSTAT data for month t until after the close of trading on the last day of month t, we assume that the accounting information for month t is used to form a portfolio position at the close on the last trading day of month t + 1. Since we have a month-by-month record of the then publicly available accounting information, we are able to implement portfolio formation rules with current data. For example, say that on February 15 a firm files its accounting statements for the quarter ending December 31. These data will be reported on the February COMPUSTAT tape and will be used in our tests starting at the end of March. By contrast, Fama and French (1992a) use accounting data for a given fiscal year to implement tests starting at the end of the following June. In many instances, there will be several months between the release of the information and the point at which it is included in the tests. If there is any mean-reversion in the book-to-market effects, there should be an advantage, in terms of power, to having more current data.

We use the same COMPUSTAT definition of book value of equity as Fama and French (1992a). As in Fama and French (1992a), we exclude from the BE/ME analysis any firms with negative BE/ME ratios.

B. Accounting Data from the Standard COMPUSTAT File

It is well known that some asset pricing anomalies can vary in magnitude across time periods and across cross-sectional samples. For example, there are periods where the size anomaly seems to be reversed [Brown, Kleidon, and Marsh (1983) and Fuller (1993)]. The time period and cross-sectional sample covered by our Real-Time COMPUSTAT sample and the Fama and French (1992a) sample overlap, but are not identical. We would not want differences in estimates attributed to selection biases if they are due to period specificity of the effects being measured. We compare the results obtained with our Real-Time COMPUSTAT sample to results obtained using data from the regular COMPUSTAT files (Industrial Current, Back Data, and Research files) over the same time period. This will control for any period-specific variation in the measured effects.

An additional difference between our Real-Time COMPUSTAT sample and the Fama and French (1992a) sample is that they have a greater representation of NASDAQ firms. There is no way to completely control for this difference. As a partial remedy, we compare our results for the Real-Time COMPUSTAT data to the results obtained with the standard COMPUSTAT data for NYSE/AMEX firms, as well as to the results obtained with the standard COMPUSTAT data for the full sample of NYSE/AMEX/NASDAQ firms.

C. Stock Return Data from CRSP

Our stock return data are from the NYSE/AMEX and NASDAQ data files from the Center for Research in Security Prices. We use monthly returns in our tests. Our proxy for the market portfolio is the CRSP value-weighted index of NYSE/AMEX/NASDAQ stocks.

D. Portfolio Allocations and Estimation of β

Cross-sectional tests of the asset pricing model as in (1) and (2) require an estimate of the assets' market risk, β. We could estimate beta individually for each asset and use that estimate in the cross-sectional regressions. The main difficulty with this approach is that betas are estimated with a fair amount of error, so that the regressions would have a severe errors-in -variables (EIV) problem. In order to reduce the EIV problem, we follow the approach of Fama and French (1992a) in which firms are allocated to portfolios and the portfolios' betas are used as estimates of the individual assets' betas in the cross-sectional regression (2). The non-beta variables, Z, in (2) are the natural log of market capitalization of the firm and the natural log of the firm's BE/ME ratio. Since these variables are measured with little or no error, the individual assets' size and BE/ME values are used in the regression. To maintain consistency with Fama and French (1992a) and Kothari, Shanken, and Sloan (1995), we use sets of portfolios sorted by market capitalization and β which are described in more detail below.

i. Pre-ranking β (β_{PRE})

Each month, we allocate assets to β portfolios on the basis of a beta estimated over the previous sixty months. For example, for January 1974 we use data from January 1969 to December 1973 to estimate β. We use all of the available data over that period but exclude firms that have less than twenty-four observations. We use the same estimator as Fama and French (1992a), which includes contemporaneous and lagged market returns in the regression:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_{i,0}(r_{m,t} - r_{f,t}) + \beta_{i,-1}(r_{m,t-1} - r_{f,t-1}) + \varepsilon_{i,t}$$

and estimates the beta for asset i as:

$$\beta_{i,PRE} = \beta_{i,0} + \beta_{i,-1}.$$

This approach is meant to adjust for nonsynchronous trading.

ii. Portfolio Allocation for Post-Ranking Beta Estimation

Given our estimate of the pre-ranking beta, and the market capitalization of the firm at the end of month t, we form portfolios at the end of month t + 1. We have two sets of portfolios ranked by size and by the pre-ranking beta.

In one set of portfolios we first allocate firms into size-deciles using the NYSE size-decile cutoff points. Each month, we rank NYSE firms by market capitalization. The bottom 10% are allocated to the first decile, the next 10% are allocated to the second decile, and so on. These allocations determine the market capitalizations for each decile. All of the firms available from CRSP (NYSE/AMEX/NASDAQ) are allocated to size-based portfolios based on the NYSE cutoff capitalizations. Within each size-decile we then rank assets by their pre-ranking β and allocate them to ten beta portfolios. This yields 100 portfolios which we refer to as the size-beta portfolios.

In the other set of portfolios we first allocate firms into beta-deciles, based on the pre-ranking beta. Within each beta-decile we then rank assets by their market capitalization and allocate them to ten size-based portfolios. This yields 100 portfolios which we refer to as the beta-size portfolios. Thus, size-beta refers to portfolios ranked first on size then on beta, while beta-size refers to portfolios ranked first on beta then on size.

The assets are equally-weighted within the portfolios. We have also performed our analysis on value-weighted portfolios. The results were essentially unchanged. Assets are re-ranked and allocated to portfolios each month.

iii. Post-ranking β (β_{POST})

Once assets are allocated to portfolios, we estimate the beta for each portfolio using the entire post-ranking period. Our estimate of the beta of each asset is the beta of the portfolio to which the asset has been assigned. We use several alternative methods to estimate the portfolios' betas. The first method is the standard regression of the portfolio's (say portfolio j) excess return on the contemporaneous return on the market portfolio proxy:

$$r_{j,t} - r_{f,t} = \alpha_j + \beta_j(r_{m,t} - r_{f,t}) + \varepsilon_{j,t} \qquad t = 1, 2, ..., T.$$

The second method is the Fama and French (1992a) approach to accounting for nonsynchronous trading by including contemporaneous and lagged market returns:

$$r_{j,t} - r_{f,t} = \alpha_j + \beta_{j,0}(r_{m,t} - r_{f,t}) + \beta_{j,-1}(r_{m,t-1} - r_{f,t-1}) + \varepsilon_{j,t} \tag{3}$$

with the estimated beta for portfolio j as:

$$\beta_j = \beta_{j,0} + \beta_{j,-1}. \tag{4}$$

The last two methods for estimating portfolio betas are motivated by the results of Handa, Kothari, and Wasley (1989) and Kothari, Shanken, and Sloan (1995). They suggest using longer horizon returns to estimate betas. In particular, they suggest using annual returns for beta estimation. Among other things, this form of beta estimation procedure will alleviate biases in beta due to nonsynchronous trading.

To illustrate the impact of using longer horizon returns on sample betas, we estimate betas of the CRSP NYSE/AMEX decile portfolios for various return horizons. The market portfolio proxy is the CRSP NYSE/AMEX portfolio and the sample period is July 1962 (corresponding to the addition of AMEX firms to the CRSP database) to December 1992. For every monthly horizon, p, between one month and one year we estimate the regression:

$$r^p_{j,t} - r^p_{f,t} = \alpha_j + \beta_j(r^p_{m,t} - r^p_{f,t}) + \varepsilon^p_{j,t} \qquad t = p, 2, ..., T \qquad (5)$$

where $r^p_{j,t}$ is the p-period return on portfolio j from t - p to t; $r^p_{f,t}$ is the p-period return to rolling over 1-month Treasury Bills from t - p to t; and $r^p_{m,t}$ is the p-period return on the market portfolio proxy from t - p to t. While Kothari, Shanken, and Sloan (1995) use non-overlapping annual return periods, we use overlapping returns so that there is an p - 1 month overlap between adjacent p-month observations. We do this in order to improve the efficiency of the estimates [see Hansen and Hodrick (1980)].

The beta estimates for deciles 1, 4, 7, and 10 are plotted in Figure 1 and the estimates for deciles 2, 3, 5, 6, 8, and 9 are given in Figure 2. The cross-sectional dispersion in betas increases until the five-month or six-month horizon, after which the dispersion declines. At the twelve-month horizon advocated by Kothari, Shanken, and Sloan (1995) there is still substantially more dispersion than at the one-month horizon. We use betas estimated using six-month and twelve-month horizons. The former corresponds approximately to the horizon with maximal dispersion while the twelve-month horizon corresponds to the methods of Handa, Kothari, and Wasley (1989) and Kothari, Shanken, and Sloan (1995).

E. Cross-Sectional Regressions

Given our post-ranking estimates of betas and our size and BE/ME instruments, we apply the cross-sectional methods of Fama and MacBeth (1973) to estimate the parameters of (2). For each month, t, from March 1974 to December 1992 we estimate variants of the regression:

$$r_{i,t} = \gamma_{0,t} + \gamma_{1,t}\,\beta_{i,POST} + \delta_t Z_{i,t-2} + \varepsilon_{i,t} \qquad i = 1, 2, ..., n$$

where the variants differ by the definition of $\beta_{i,POST}$ and the composition of Z. From the resulting time-series of coefficient estimates, $\hat{\gamma}_{0,t}$, $\hat{\gamma}_{1,t}$, and $\hat{\delta}_t$, we estimate the parameters by the time-series mean of the month-by-month coefficients:

$$\hat{\gamma}_0 = \frac{1}{T}\sum_{t=1}^{T}\hat{\gamma}_{0,t} \qquad \hat{\gamma}_1 = \frac{1}{T}\sum_{t=1}^{T}\hat{\gamma}_{1,t} \qquad \hat{\delta} = \frac{1}{T}\sum_{t=1}^{T}\hat{\delta}_t$$

while the time-series variance of the parameter estimates allows us to test hypotheses about the parameters of interest.

F. Time-Series Tests

As an alternative to the cross-sectional regressions described in Section I.E we perform time-series tests of the restrictions imposed by the asset pricing model. Let $R_{i,t}$ denote the return on asset i in excess of the riskless interest rate ($R_{i,t} = r_{i,t} - r_{f,t}$) and $R_{m,t}$ be the return on the market portfolio proxy in excess of the riskless interest rate. With n assets in the cross-section, let R_t denote the $n \times 1$ vector of asset returns for period t, $R_t = (R_{1,t}, R_{2,t}, ..., R_{n,t})'$, and let β denote the $n \times 1$ vector of asset betas relative to the market. Consider the time-series multivariate regression:

$$R_t = \alpha + \beta R_{m,t} + \varepsilon_t \qquad (6)$$

where α is an $n \times 1$ vector of intercept coefficients. A testable restriction implied by the unconditional pricing model is that $\alpha = 0$. Testing the joint restriction, $\alpha = 0$, is generally not feasible for the full set of n individual assets since there are many more assets than time-series observations. We group the assets into sets of portfolios formed on the basis of book-to-market, size, and pre-ranking beta. For each of the three ranking variables, we form 20 ventile portfolios and 10 decile portfolios. For the BE/ME and beta portfolios each of the ventile portfolios contains 5% of the assets and each of the decile portfolios contains 10% of the assets. For the size portfolios the ventile and decile cutoff points are defined by the NYSE market capitalizations. Since our sample contains AMEX and NASDAQ securities as well as NYSE securities, the small capitalization ventile and decile portfolios will contain more than 5% or 10% of the sample, respectively, while large capitalization ventile and decile portfolios will contain less than 5% or 10% of the sample, respectively. We form the portfolios with both value weighting and equal weighting.

We use a modified likelihood ratio test [see Rao (1973, pp. 554-556)] to test the hypothesis that $\alpha = 0$. For testing $\alpha = 0$ in (6) the test statistic is:

$$\left[\frac{|\hat{V}_r|}{|\hat{V}_u|} - 1\right] \cdot \frac{T - 1 - n}{n}$$

where n is the number of portfolios in regression (6) ($n = 10$ for the deciles and $n = 20$ for the ventiles), T is the size of the time-series sample (226 monthly observations), $\hat{V}_u$ is the estimated covariance matrix of ε_t in the unrestricted regression (6), and $\hat{V}_r$ is the estimated covariance matrix of ε_t in (6) when α is restricted to be equal to the zero vector. Under the hypothesis that the vector ε_t is independent, identically distributed, and multivariate normal, the test statistic has an exact small-sample distribution which is $F_{n,\, T - 1 - n}$.

II. Empirical Results

A. Summary Statistics

In Table 1 we report summary statistics on decile portfolios formed on the basis on BE/ME, size, and pre-ranking beta. The results for BE/ME-based deciles are in Panel A. The mean returns generally increase with increases in the BE/ME ratio. The post ranking betas, however, are somewhat U-shaped, being high for low and high BE/ME portfolios and being low for intermediate BE/ME portfolios. Fama and French (1992a) also find this U-shaped pattern in betas, although the pattern seems less pronounced in their data. Also, higher BE/ME ratio stocks tend to be those with lower market capitalization. The results for size-based deciles are in Panel B. We see the usual declining mean return as the market capitalization increases. There is a slight decline in β as size increases. In Panel C the mean returns on beta-sorted portfolios do not show any pronounced pattern. Market capitalization tends to decrease as beta increases.

Tables 2 and 3 present summary statistics for portfolios sorted by size and beta. Table 2 contains data for portfolios sorted by beta first and then sorted by size. There is not a very strong correspondence between mean returns and beta. Within a beta decile there seems to be a stronger inverse relation between mean returns and size. Table 3 contains data for portfolios sorted by size first and then sorted by beta. Again, there is a stronger relation between size and mean returns than between beta and mean returns. Thus, the evidence in Tables 1 - 3 is consistent with a weak or no relation between beta and mean returns and a strong relation between size and mean returns, as has been found by previous authors.

B. Cross-Sectional Regressions

The results of our cross-sectional regression tests are presented in Tables 4 and 5. The difference across tables is the manner in which an asset's beta is estimated. In every case the asset is allocated to a companion portfolio each month. The post-ranking beta of the companion portfolio is used as the beta of the asset. The instruments, Z, of the individual assets are used. In Table 4 the companion portfolios are determined by first ranking all assets in the sample by their pre-ranking beta and allocating them to beta deciles. Within each beta decile, the assets are ranked by market capitalization and allocated to decile portfolios. This provides us with 100 companion portfolios which we will refer to as the beta/size portfolios, denoting ranking by betas first and then by size. Firms are reallocated to companion portfolios each month on the basis of capitalization and size data lagged by a month. For example, firms are allocated to companion portfolios for March 1974 (i.e., the portfolio is formed at the end of February 1974) on the basis of size and beta data available at the end of January 1974. We include in the sample only the firms for which we have BE/ME data.

In Table 5 the companion portfolios are determined by first allocating all assets into ten size-based portfolios using the NYSE size decile cutoff points. Within a size decile, assets are ranked by the pre-ranking beta and allocated to beta deciles. This provides us with 100 companion portfolios which we will refer to as the size/beta portfolios, denoting ranking by size first and then by beta. As for the other portfolios, assets are reallocated to companion portfolios each month on the basis of capitalization and size data lagged by a month.

The first column in Tables 4 and 5 reports the average value of $\hat{\gamma}_{0,t}$, $\hat{\gamma}_0$. Under the null hypothesis that there is unrestricted borrowing or lending at the riskless interest rate and that the single beta model correctly prices assets, we should find that $\gamma_{0,t} = r_{f,t}$. The second column reports the mean difference between $\hat{\gamma}_{0,t}$ and $r_{f,t}$, where we use the return on one-month Treasury bills as our proxy for $r_{f,t}$. The third column reports the estimated average premium for beta risk, $\hat{\gamma}_1$. Under the null hypothesis that the single beta model correctly prices assets, we should find that $\gamma_{1,t} = r_{m,t} - r_{f,t}$. The fourth column reports the average difference between $\hat{\gamma}_{1,t}$ and $R_{m,t}$, where $R_{m,t} = r_{m,t} - r_{f,t}$. The fifth column reports the average coefficient, δ, when ln(ME) is included in the cross-sectional regression and the sixth column reports the average coefficient, δ, when ln(BE/ME) is included in the cross-sectional regression.

Tables 4 and 5 use the post-ranking beta estimation procedure of Fama and French (1992a) and described by (3) and (4) above. Panel A in Tables 4 and 5 reports results from our Real-Time COMPUSTAT sample. The first regression in the panel is a regression of returns on a constant and beta. In both Table 4 and Table 5, $\hat{\gamma}_0$ is significantly different from zero (at the 5% level) and significantly different from $\overline{r}_f$ in Table 5. As has been found by Fama and French (1992a) and others, $\hat{\gamma}_1$ is not significantly different from zero, nor is it significantly different from $\overline{R}_m = \overline{r}_m - \overline{r}_f$. Thus, the question of the precision of the estimate $\hat{\gamma}_1$, raised by a number of authors, seems to be an important issue. The point estimates indicate monthly risk premia of 0.29% (Table 4) and 0.19 (Table 5). These estimates correspond to an annual risk premium in excess of 3.5% and 2.3%, respectively.

The second regression in the tables is a regression of returns on a constant and ln(ME). We find the typical negative relation between returns and market capitalization. Our point estimate of -0.11 is very close to the estimate of -0.15 found by Fama and French (1992a, Table III). The relation, however, is statistically insignificant in our sample. This might be due to lower precision of our estimates since the sample size is lower, or due to time-variation in the size anomaly.

The third regression in the tables is a regression of returns on a constant and ln(BE/ME). If our sample corrects for a selection bias hypothesized by Kothari, Shanken, and Sloan (1995), we should find a smaller coefficient on ln(BE/ME). Our coefficient on ln(BE/ME) is a statistically insignificant 0.22, which is less than half of the estimate of 0.50 obtained by Fama and French (1992a, Table III). Thus, the results are in the hypothesized direction. However, the results for $\hat{\gamma}_1$, in which it is insignificantly different from both zero and $\bar{R}_m$, should lead us to be cautious about interpreting the results for ln(BE/ME) as an indication of a selection bias in their results. A reasonable question is whether our estimate of 0.22 is significantly below their estimate of 0.50. We cannot answer this question definitively since it requires knowledge of the correlation between their estimate and ours. There is certainly some correlation since both estimates are obtained from cross-sectional regressions of an overlapping cross-sectional sample of assets over overlapping time periods. However, from the reported estimates we can obtain a range of correlations that would lead to rejection and a range of correlations that would not lead to rejection. A correlation of 0.28 or greater between our estimate and the Fama and French (1992a) estimate will lead us to reject the hypothesis that the estimates are equal (i.e., the t-statistic will be greater than 1.96 for correlations greater than 0.28). As in the case with the measured size effect, the lack of statistical significance of our estimated book-to-market effect might be due to lower precision of our estimates since the sample size is lower, or due to time-variation in the book-to-market anomaly. We address these possibilities below.

The fourth regression in the tables is a regression of returns on a constant, beta, and ln(ME). The estimate of γ_0, $\hat{\gamma}_0$, is significantly greater that $\bar{r}_f$. As is typical, $\hat{\gamma}_1$ is insignificantly different from zero and is significantly below $\bar{R}_m$. The negative relation between returns and capitalization continues to be insignificant after accounting for any relation between returns and betas.

The fifth regression in the tables is a regression of returns on a constant, beta, and ln(BE/ME). Again, the estimate of γ_0, $\hat{\gamma}_0$, is significantly greater that $\bar{r}_f$. The estimate of $\hat{\gamma}_1$ is negative but insignificantly different from both zero and $\bar{R}_m$. The relation between returns and ln(BE/ME) is positive but insignificant after accounting for any relation between returns and betas.

The sixth regression in the tables is a regression of returns on a constant, beta, ln(ME), and ln(BE/ME). The estimate of γ_0, $\hat{\gamma}_0$, is significantly greater that $\bar{r}_f$. The estimate $\hat{\gamma}_1$ is again negative, insignificantly different from zero, and significantly below $\bar{R}_m$. The negative relation between returns and capitalization is marginally significant (at approximately the 10% level). The relation between returns and ln(BE/ME) is small, with $\hat{\delta}$ estimated to be 0.06. It is still insignificantly different from zero.

Our sample and the sample of Fama and French (1992a) overlap but are not identical either cross-sectionally or in the time period studied. The difference observed here might be due, for example, to the fact that we study different time periods. In order to check whether our results are driven by the use of a different sample period, we reestimate the cross-sectional regressions using the standard COMPUSTAT data to define the book-to-market ratio. As in Fama and French (1992a), we use the book value of equity for fiscal year t - 1 to define the book-to-market ratio from the end of June of year t through the end of May of year t + 1. The market value of equity for a given month is defined as the market value one month prior to the portfolio formation date.

The results using the standard COMPUSTAT data over the same time period over which our Real-Time COMPUSTAT data are available are reported in Panel B of Tables 4 and 5. The estimates of γ_1 when only beta is included in the regression are 0.28 (Table 4) and 0.12 (Table 5), neither of which is significantly different from zero or from the excess return on the market portfolio proxy. Fama and French (1992a) have an estimate of 0.15 which is also insignificantly different from zero. Our estimate of the size effect is -0.18 and is significant, whereas the Fama and French (1992a) estimate is -0.15 (also significant). Our estimate of the book-to market effect is 0.33 and is significant, while the Fama and French (1992a) estimate is 0.50 (also significant). As in Fama and French (1992a), when the cross-sectional regressions include size as well as β, the estimated market risk premium becomes negative and significantly below the observed risk premium on the market proxy.

The results reported in Panel B of Tables 4 and 5 indicate that the difference between the results obtained using the Real-Time COMPUSTAT sample and the results reported in the previous literature are not merely an artifact of the different time periods studied. We can address the issue of whether the observed differences across the two COMPUSTAT samples are merely due to lack of precision in the estimate, that is due to sampling error, by testing for differences across the two sets of estimates in panels A and B. Panel C of Tables 4 and 5 report the t-statistics for the tests of the hypotheses that the coefficients in Panel A and Panel B are the same. There is one regression, out of eight, where the estimate of the market risk premium, γ_1, is significantly different across the panels (the regression including β and size in Table 5). The estimated size coefficient is always significantly different across the two panels. The estimated book-to-market effects are significantly different across Panels A and B when either β or size is included in the regression. However the difference is not significant when only BE/ME is included in the regression.

While part of the difference in the estimated book-to-market effects using the Real-Time versus the usual COMPUSTAT samples may be due to selection biases or pure sampling error, another explanation for the difference may be that the book-to-market effect is concentrated in NASDAQ firms which are more well-represented in the standard COMPUSTAT sample than in our Real-Time sample. It is not really feasible to completely disentangle these possibilities. As an imperfect attempt to shed some light on this issue, we estimated the cross-sectional regressions using the standard COMPUSTAT data, but restricting the cross-sectional sample to be NYSE/AMEX firms. The full results are not reported here, but the estimated book-to-market effect using the standard COMPUSTAT data on NYSE/AMEX firms is generally insignificantly smaller than the real-time COMPUSTAT results. This implies that the differences between panels A and B are concentrated in the NASDAQ sample. Some of this difference may be due to a truly stronger book-to-market effect in NASDAQ stocks. Some of the difference may be attributable to the same kind of selection bias since the expansion of the COMPUSTAT files in the 1970's (and, hence, the data back-filling problem) was more concentrated in NASDAQ firms.

The cross-sectional regressions in Tables 4 and 5 were repeated using the long horizon betas [as in (5)] rather than the estimates from (3) and (4). While there seems to be a slight increase in the estimated beta risk premium, $\hat{\gamma}_1$, the basic conclusions of Tables 4 and 5 are unaltered using the long-horizon betas and, therefore, are not reported in detail. This small difference seems to conflict with the findings of Handa, Kothari, and Wasley (1989). However, their comparison is between long-horizon estimates of beta and 1-month horizon ordinary least squares (OLS) estimates of beta. We are comparing long-horizon estimates of beta with the Fama/French beta which incorporates an adjustment for non-synchronous trading. Our results indicate that the Fama/French method of estimating beta is providing almost identical cross-sectional results as we obtain from the long-horizon betas on the Real-Time COMPUSTAT sample. We have also run the same analysis using 1-month horizon OLS estimates of β. In this case we find results consistent with Handa, Kothari, and Wasley (1989). The estimated values of γ_1 tend to be much smaller (e.g., -0.49 versus 0.29 in regression 1 of Table 4 or -0.21 versus 0.19 in regression 1 of Table 5). Thus, the Fama/French method of estimating beta seems to be performing much the same task as the long-horizon regression estimates of beta.

C. Time-Series Tests

Table 6 contains the modified likelihood ratio statistics for testing the hypothesis that $\alpha = 0$ in (6). If we use a 5% critical value, the equal-weighted BE/ME decile portfolios are the only BE/ME portfolios that lead to a rejection of $\alpha = 0$. The equal-weighted size decile and ventile portfolios have test statistics that are statistically significant (at the 5% critical level). All of the beta sorted portfolios, except the value-weighted decile portfolios, have test statistics that are statistically significant (at the 5% critical level).

Figure 3 is a plot of α for the BE/ME decile portfolios. Consistent with the positive BE/ME effects in the cross-sectional regressions [$\hat{\delta} > 0$ in (2)], there is a positive relation between α_i and BE/ME. However, as in the cross-sectional results, the effect is usually statistically insignificant, with the exception of the equal-weighted BE/ME decile portfolios. Figures 4 and 5 have the plots of α for the size and beta portfolios, respectively. The figures show the standard negative relation between α and the market capitalization of the firm as well as the negative relation between α and beta found, for example, by Black, Jensen, and Scholes (1972).

III. Conclusions

In this paper we have investigated hypothesized selection biases in tests of asset pricing models which are due to the use of accounting data from the COMPUSTAT files. Our data provide us with a sample from COMPUSTAT which is free of the hypothesized selection biases.

Our results regarding the explanatory power of unconditional betas and size are basically the same as those of Fama and French (1992a). In isolation, the estimated market risk premium is insignificantly different from zero and from the average risk premium on the market proxy. When size is included, the estimated market risk premium is negative and is significantly below the average risk premium on the market proxy. The estimated book-to-market effect is less than half of their estimated effect. In the cross-sectional regressions, the book-to-market effect in the Real-Time sample is insignificantly different from zero, but significantly below the estimated book-to-market effect using the standard COMPUSTAT data (when β and ME are included in the regressions). This difference seems to be due to the lower representation of NASDAQ firms in our Real-Time sample.

In time-series tests of the single factor model we can reject the model's joint restriction across equations for equal-weighted BE/ME-based decile portfolios. While there appears to be a BE/ME pattern in the estimated mispricing across portfolios, the estimates are rather imprecise.

It is interesting to note that other research [Davis (1994b)] using a survivorship bias-free sample finds an estimated book-to-market effect almost identical to ours (0.26% versus our 0.29%). His estimate is statistically significant while ours is not. The postulated selection bias does not seem to be important for samples restricted to NYSE/AMEX firms. There is some difference when NASDAQ firms are included in the standard COMPUSTAT sample. This may be due to a truly stronger BE/ME effect or to a more severe selection bias in that sample. Our data do not allow us to disentangle these two possible explanations.

Table 1: Summary Statistics for Decile Portfolios Formed on the Basis of BE/ME, Size (ME), and Pre-Ranking Beta: March 1974-December 1992.

The BE/ME decile portfolios are formed each month by ranking firms with positive book equity by BE/ME. The 10% of the firms with the smallest BE/ME are allocated to portfolio 1, etcetera. The size decile portfolios are formed each month by ranking firms by ME. The 10% of the firms with the smallest ME are allocated to portfolio 1, etcetera. The beta decile portfolios are formed each month by ranking firms by beta estimated over the previous 60 months, adjusting for nonsynchronous trading, as in (3) and (4). The 10% of the firms with the smallest beta are allocated to portfolio 1, etcetera. Assets within portfolios are equally-weighted. BE/ME, ME, or beta data at the end of period t-1 are used to form portfolios at the end of period t. Return: is the average monthly return (in percent). β_{POST} is the post ranking beta of the portfolio estimated as in (3) and (4) over the entire period (March 1994-December 1992). ln(ME) is the time-series average of the natural log of the average market capitalization of assets in the portfolio (ME is in millions of dollars). ln(BE/ME) is the time-series average of the natural log of the average BE/ME ratio of assets in the portfolio.

Portfolio	1	2	3	4	5	6	7	8	9	10
Panel A: Stocks Sorted on BE/ME										
Return	1.29	1.37	1.44	1.40	1.48	1.59	1.73	1.73	1.83	1.77
β_{POST}	1.35	1.27	1.25	1.17	1.15	1.07	1.07	1.13	1.22	1.45
ln(ME)	7.26	7.02	6.91	6.84	6.56	6.44	6.25	5.93	5.48	4.96
ln(BE/ME)	-1.50	-0.90	-0.61	-0.40	-0.23	-0.08	0.06	0.23	0.44	1.02
Firms	202	202	201	202	202	202	203	202	201	193
Panel B: Stocks Sorted on Size										
Return	1.62	1.38	1.43	1.45	1.46	1.43	1.42	1.36	1.25	1.08
β_{POST}	1.30	1.26	1.23	1.22	1.17	1.13	1.10	1.08	1.02	0.94
ln(ME)	2.51	3.84	4.34	4.78	5.22	5.67	6.14	6.66	7.27	8.89
Firms	2932	618	424	354	304	244	214	192	175	172
Panel C: Stocks Sorted on Beta										
Return	1.43	1.47	1.51	1.50	1.64	1.59	1.56	1.73	1.60	1.38
β_{POST}	0.68	0.78	0.85	0.96	1.04	1.14	1.19	1.29	1.36	1.46
ln(ME)	6.09	6.65	6.65	6.33	6.12	5.98	5.56	5.22	4.71	4.12
Firms	445	451	452	454	453	453	451	450	447	439

Table 2: Summary Statistics for 100 Portfolios Formed on the Basis of Pre-Ranking Beta and Size (ME): March 1974-December 1992.

The portfolios are formed by first ranking into deciles based on pre-ranking beta then by ranking by size deciles within each beta decile. Return: is the average monthly return (in percent). Beta is the post ranking beta of the portfolio estimated as in (3) and (4) over the entire period (March 1994-December 1992). ln(ME) is the time-series average of the natural log of the average market capitalization of assets in the portfolio (ME is in millions of dollars). Assets within portfolios are equally-weighted. ME and beta data at the end of period t-1 are used to form portfolios at the end of period t.

	ME-1	ME-2	ME-3	ME-4	ME-5	ME-6	ME-7	ME-8	ME-9	ME-10
Panel A: Average Monthly Returns (%)										
Low-β	3.49	1.75	1.40	1.08	1.09	1.03	1.03	1.15	1.08	1.29
β-2	2.11	1.93	1.48	1.21	1.33	1.33	1.42	1.55	1.26	1.19
β-3	2.47	1.58	1.39	1.37	1.45	1.50	1.38	1.53	1.34	1.20
β-4	1.79	1.51	1.46	1.48	1.44	1.43	1.70	1.45	1.58	1.27
β-5	2.68	1.75	1.47	1.54	1.67	1.54	1.51	1.64	1.53	1.15
β-6	2.23	1.65	1.51	1.68	1.74	1.67	1.55	1.44	1.27	1.28
β-7	2.30	1.39	1.39	1.77	1.44	1.72	1.68	1.53	1.39	1.15
β-8	3.45	1.72	1.31	1.54	1.40	1.77	1.66	1.72	1.69	1.23
β-9	3.55	1.49	1.54	1.15	1.63	1.30	1.32	1.46	1.48	1.27
High-β	4.15	1.79	0.81	1.02	0.79	1.12	0.89	1.11	1.01	1.42
Panel B: Post Ranking Beta										
Low-β	0.97	1.00	0.86	0.83	0.87	0.75	0.70	0.63	0.64	0.57
β-2	1.01	0.97	0.95	0.90	0.82	0.85	0.77	0.77	0.73	0.76
β-3	1.04	0.99	0.97	0.91	0.98	0.95	0.96	0.99	0.94	0.87
β-4	1.10	1.15	1.14	1.13	1.03	1.04	1.04	1.07	1.08	0.97
β-5	1.19	1.29	1.18	1.18	1.18	1.13	1.09	1.11	1.10	1.08
β-6	1.27	1.29	1.27	1.31	1.27	1.23	1.26	1.17	1.20	1.15
β-7	1.39	1.33	1.27	1.30	1.31	1.33	1.36	1.25	1.30	1.21
β-8	1.56	1.46	1.44	1.50	1.41	1.45	1.43	1.42	1.33	1.30
β-9	1.51	1.60	1.61	1.55	1.59	1.54	1.57	1.56	1.55	1.45
High-β	1.96	1.66	1.63	1.66	1.66	1.74	1.71	1.72	1.74	1.70

Panel C: Size: ln(ME)										
Low-β	0.42	1.42	2.16	2.73	3.28	3.85	4.46	5.23	6.18	8.31
β-2	0.78	1.85	2.53	3.12	3.68	4.25	4.85	5.58	6.51	9.10
β-3	0.95	2.02	2.71	3.30	3.86	4.39	4.99	5.69	6.59	9.03
β-4	1.06	2.10	2.80	3.40	3.96	4.52	5.13	5.80	6.67	8.59
β-5	1.07	2.09	2.78	3.36	3.90	4.47	5.10	5.80	6.62	8.46
β-6	1.04	2.03	2.66	3.24	3.80	4.37	5.00	5.70	6.56	8.28
β-7	0.98	1.96	2.57	3.11	3.60	4.11	4.65	5.29	6.15	7.82
β-8	0.94	1.89	2.48	2.99	3.46	3.92	4.40	4.98	5.78	7.48
β-9	0.75	1.66	2.22	2.70	3.15	3.60	4.04	4.55	5.21	6.82
High-β	0.55	1.35	1.84	2.29	2.70	3.11	3.55	4.04	4.68	6.18

Table 3: Summary Statistics for 100 Portfolios Formed on the Basis of Size (ME) and Pre-Ranking Beta : March 1974-December 1992.

The portfolios are formed by first ranking into ten size-based portfolios on the basis of market capitalization and then by ranking into beta deciles by on pre-ranking beta within each size decile. Return: is the average monthly return (in percent). Beta is the post ranking beta of the portfolio estimated as in (3) and (4) over the entire period (March 1974-December 1992). ln(ME) is the time-series average of the natural log of the average market capitalization of assets in the portfolio (ME is in millions of dollars). Assets within portfolios are equally-weighted. ME and beta data at the end of period t-1 are used to form portfolios at the end of period t.

	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
Panel A: Average Monthly Returns (%)										
Small-ME	1.69	1.65	1.67	1.73	1.89	1.60	1.76	1.81	1.84	1.48
ME-2	1.05	1.51	1.46	1.53	1.65	1.76	1.62	1.71	1.29	1.12
ME-3	1.24	1.44	1.56	1.41	1.49	1.71	1.68	1.57	1.53	1.32
ME-4	1.26	1.52	1.34	1.55	1.72	1.52	1.70	1.71	1.50	1.40
ME-5	1.43	1.46	1.56	1.64	1.38	1.63	1.55	1.73	1.45	1.20
ME-6	1.38	1.50	1.48	1.51	1.56	1.55	1.50	1.20	1.60	1.37
ME-7	1.19	1.66	1.28	1.57	1.70	1.54	1.39	1.33	1.40	1.44
ME-8	1.30	1.27	1.46	1.50	1.54	1.51	1.24	1.09	1.46	1.18
ME-9	1.19	1.28	1.35	1.16	1.61	1.62	1.22	1.21	1.35	0.83
Large-ME	1.19	1.30	1.31	1.12	1.05	1.00	1.07	0.89	1.03	0.88
Panel B: Post Ranking Beta										
Small-ME	0.91	0.92	1.00	1.16	1.25	1.29	1.43	1.51	1.62	1.73
ME-2	0.67	0.88	0.98	1.07	1.22	1.33	1.34	1.51	1.50	1.76
ME-3	0.67	0.84	0.99	1.08	1.14	1.27	1.39	1.44	1.60	1.71
ME-4	0.68	0.85	0.95	1.04	1.15	1.23	1.30	1.40	1.59	1.77
ME-5	0.65	0.90	1.03	1.01	1.11	1.22	1.20	1.37	1.41	1.63
ME-6	0.57	0.76	1.00	1.08	1.07	1.11	1.24	1.30	1.38	1.65
ME-7	0.52	0.86	1.00	1.01	1.12	1.20	1.22	1.31	1.26	1.55
ME-8	0.50	0.79	0.95	1.06	1.08	1.09	1.19	1.18	1.36	1.50
ME-9	0.57	0.70	0.89	0.98	1.07	1.12	1.11	1.19	1.22	1.35
Large-ME	0.57	0.77	0.75	0.89	0.96	0.98	1.00	1.06	1.16	1.36

Panel C: Size: ln(ME)										
Small-ME	2.38	2.55	2.60	2.63	2.63	2.64	2.65	2.63	2.58	2.48
ME-2	3.93	3.93	3.93	3.94	3.94	3.94	3.94	3.94	3.93	5.65
ME-3	4.44	4.45	4.45	4.45	4.44	4.44	4.45	4.45	4.44	4.44
ME-4	4.89	4.90	4.90	4.90	4.90	4.90	4.90	4.89	4.90	4.89
ME-5	5.35	5.34	5.35	5.36	5.35	5.35	5.34	5.34	5.35	5.33
ME-6	5.81	5.81	5.82	5.81	5.81	5.81	5.81	5.82	5.81	5.80
ME-7	6.29	6.30	6.29	6.29	6.30	6.29	6.29	6.29	6.28	6.29
ME-8	6.84	6.83	6.84	6.81	6.82	6.81	6.82	6.82	6.83	6.81
ME-9	7.48	7.48	7.46	7.46	7.47	7.48	7.47	7.45	7.45	7.44
Large-ME	9.46	10.00	9.41	9.24	9.04	9.07	8.99	8.83	8.64	8.48

Table 4: Cross-Sectional Regressions Using Post Ranking β's
Defined as in Fama and French (1992a)

Cross-sectional regression estimates of (2): $r_i = \gamma_0 + \gamma_1 \beta_i + \delta Z_i + \varepsilon_i$, where β is the post-ranking beta formed by summing the multiple regression slope coefficients from a regression of the companion portfolio return on the contemporaneous and lagged market returns as in equations (3) and (4). The companion portfolios are formed by first ranking into deciles based on pre-ranking beta then by ranking by size deciles within each beta decile. ME is the market capitalization of the equity of the firm (in thousands of dollars). BE/ME is the ratio of book equity to market equity of the firm. Regressions are run for each month over the time period is March 1974 through December 1992. The reported coefficients are the time-series averages of the monthly coefficients. The t-statistics (in parentheses) are given by the time-series average divided by its time-series standard error. P-values are reported in brackets.

Panel A: Real-Time COMPUSTAT

$\hat{\gamma}_0$	$\hat{\gamma}_0-\bar{r}_f$	$\hat{\gamma}_1$	$\hat{\gamma}_1-\bar{R}_m$	δ: ln(ME)	δ: ln(BE/ME)
1.20	0.58	0.29	-0.24		
(3.94)	(1.91)	(0.66)	(-0.64)		
[0.00]	[0.06]	[0.51]	[0.53]		
2.66				-0.11	
(2.60)				(-1.53)	
[0.01]				[0.13]	
1.54					0.22
(4.07)					(1.57)
[0.00]					[0.12]
2.73	2.11	-0.06	-0.59	-0.10	
(3.41)	(2.63)	(-0.17)	(-2.28)	(-1.65)	
[0.00]	[0.01]	[0.87]	[0.02]	[0.10]	
1.24	0.62	0.23	-0.30		0.17
(4.59)	(2.30)	(0.57)	(-0.88)		(1.26)
[0.00]	[0.02]	[0.57]	[0.38]		[0.21]
2.47	1.84	-0.03	-0.56	-0.08	0.06
(3.56)	(2.66)	(-0.08)	(-2.18)	(-1.56)	(0.59)
[0.00]	[0.01]	[0.93]	[0.03]	[0.12]	[0.56]

Panel B: Standard COMPUSTAT					
$\hat{\gamma}_0$	$\hat{\gamma}_0-\bar{r}_f$	$\hat{\gamma}_1$	$\hat{\gamma}_1-\bar{R}_m$	δ: ln(ME)	δ: ln(BE/ME)
1.23	0.61	0.28	-0.24		
(4.76)	(2.37)	(0.72)	(-0.74)		
[0.00]	[0.02]	[0.47]	[0.46]		
3.45				-0.18	
(3.98)				(-3.00)	
[0.00]				[0.00]	
1.68					0.33
(4.49)					(3.33)
[0.00]					[0.00]
3.82	3.19	-0.15	-0.68	-0.19	
(5.30)	(4.43)	(-0.41)	(-2.70)	(-3.31)	
[0.00]	[0.00]	[0.68]	[0.01]	[0.00]	
1.19	0.57	0.39	-0.13		0.30
(4.81)	(2.31)	(1.04)	(-0.42)		(3.19)
[0.00]	[0.02]	[0.30]	[0.67]		[0.00]
3.41	2.78	0.01	-0.51	-0.17	0.17
(5.20)	(4.24)	(0.04)	(-2.07)	(-3.10)	(2.10)
[0.00]	[0.00]	[0.97]	[0.04]	[0.00]	[0.04]

Panel C: Difference Between Standard and Real Time COMPUSTAT			
$\hat{\gamma}_0$	$\hat{\gamma}_1$	δ: ln(ME)	δ: ln(BE/ME)
(0.19)	(-0.04)		
(2.15)		(-2.60)	
(2.45)			(1.55)
(2.65)	(-0.65)	(-3.30)	
(-0.38)	(1.31)		(2.38)
(2.30)	(0.34)	(-3.05)	(2.04)

Table 5: Cross-Sectional Regressions Using Post Ranking β's Defined as in Fama and French (1992a)

Cross-sectional regression estimates of (2): $r_i = \gamma_0 + \gamma_1 \beta_i + \delta Z_i + \varepsilon_i$, where β is the post-ranking beta formed by summing the multiple regression slope coefficients from a regression of the companion portfolio return on the contemporaneous and lagged market returns as in equations (3) and (4). The companion portfolios are formed by first ranking into deciles based on market capitalization then by ranking into pre-ranking beta deciles within each size decile. ME is the market capitalization of the equity of the firm (in thousands of dollars). BE/ME is the ratio of book equity to market equity of the firm. Regressions are run for each month over the time period is March 1974 through December 1992. The reported coefficients are the time-series averages of the monthly coefficients. The t-statistics (in parentheses) are given by the time-series average divided by its time-series standard error. P-values are reported in brackets.

Panel A: Real-Time COMPUSTAT

$\hat{\gamma}_0$	$\hat{\gamma}_0-\bar{r}_f$	$\hat{\gamma}_1$	$\hat{\gamma}_1-\bar{R}_m$	δ: ln(ME)	δ: ln(BE/ME)
1.31	0.69	0.19	-0.33		
(4.73)	(2.50)	(0.45)	(-0.96)		
[0.00]	[0.01]	[0.65]	[0.34]		
2.66				-0.11	
(2.60)				[-1.53]	
[0.01]				[0.13]	
1.54					0.22
(4.07)					[1.57]
[0.00]					[0.12]
2.69	2.07	-0.05	-0.57	-0.10	
(3.48)	(2.67)	(-0.13)	(-2.26)	(-1.64)	
[0.00]	[0.01]	[0.90]	[0.02]	[0.10]	
1.29	0.67	0.18	-0.34		0.17
(5.10)	(2.65)	(0.46)	(-1.07)		(1.31)
[0.00]	[0.01]	[0.65]	[0.29]		[0.19]
2.44	1.82	-0.01	-0.54	-0.08	0.06
(3.61)	(2.68)	(-0.04)	(-2.16)	(-1.54)	(0.58)
[0.00]	[0.01]	[0.97]	[0.03]	[0.12]	[0.56]

Panel B: Standard COMPUSTAT					
$\hat{\gamma}_0$	$\hat{\gamma}_0-\bar{r}_f$	$\hat{\gamma}_1$	$\hat{\gamma}_1-\bar{R}_m$	δ: ln(ME)	δ: ln(BE/ME)
1.43	0.81	0.12	-0.41		
(5.75)	(3.27)	(0.30)	(-1.29)		
[0.00]	[0.00]	[0.77]	[0.20]		
3.45				-0.18	
(3.98)				(-3.00)	
[0.00]				[0.00]	
1.68					0.33
(4.49)					(3.33)
[0.00]					[0.00]
3.99	3.37	-0.29	-0.82	-0.19	
(5.54)	(4.66)	(-0.80)	(-3.32)	(-3.33)	
[0.00]	[0.00]	[0.42]	[0.00]	[0.00]	
1.38	0.76	0.23	-0.29		0.30
(5.79)	(3.20)	(0.62)	(-0.95)		(3.15)
[0.00]	[0.00]	[0.53]	[0.34]		[0.00]
3.59	2.96	-0.13	-0.65	-0.17	0.16
(5.45)	(4.49)	(-0.37)	(-2.70)	(-3.12)	(2.03)
[0.00]	[0.00]	[0.71]	[0.01]	[0.00]	[0.04]

Panel C: Difference Between Standard and Real Time COMPUSTAT			
$\hat{\gamma}_0$	$\hat{\gamma}_1$	δ: ln(ME)	δ: ln(BE/ME)
(0.86)	(-0.60)		
(2.15)		(-2.60)	
(2.45)			(1.55)
(3.24)	(-2.00)	(-3.33)	
(0.75)	(0.42)		(2.12)
(2.86)	(-0.96)	(-3.07)	(1.96)

Table 6: Time-Series Tests of the Unconditional Single Factor Model

The table reports the modified likelihood ratio tests of α = 0 in (6) for decile and ventile portfolios formed on the basis of BE/ME, market capitalization and beta. Panel A reports results for portfolios formed on the basis of BE/ME. Panel B reports results for portfolios formed on the basis of market capitalization. Panel C reports results for portfolios formed on the basis of beta. The sample period is March 1974 through December 1992.

Weighting of LHS Assets	Number of Portfolios (n)	$F_{n,226-1-n}$	P-Value
Panel A: Portfolios Formed by BE/ME			
Value weighted	20	1.17	0.28
Value weighted	10	1.37	0.20
Equal weighted	20	1.16	0.29
Equal weighted	10	2.11	0.02
Panel B: Portfolios Formed by Market Capitalization			
Value weighted	20	1.03	0.42
Value weighted	10	0.92	0.52
Equal weighted	20	1.89	0.01
Equal weighted	10	2.07	0.03
Panel C: Portfolios Formed by Beta			
Value weighted	20	1.63	0.05
Value weighted	10	0.81	0.62
Equal weighted	20	1.97	0.01
Equal weighted	10	3.11	0.00

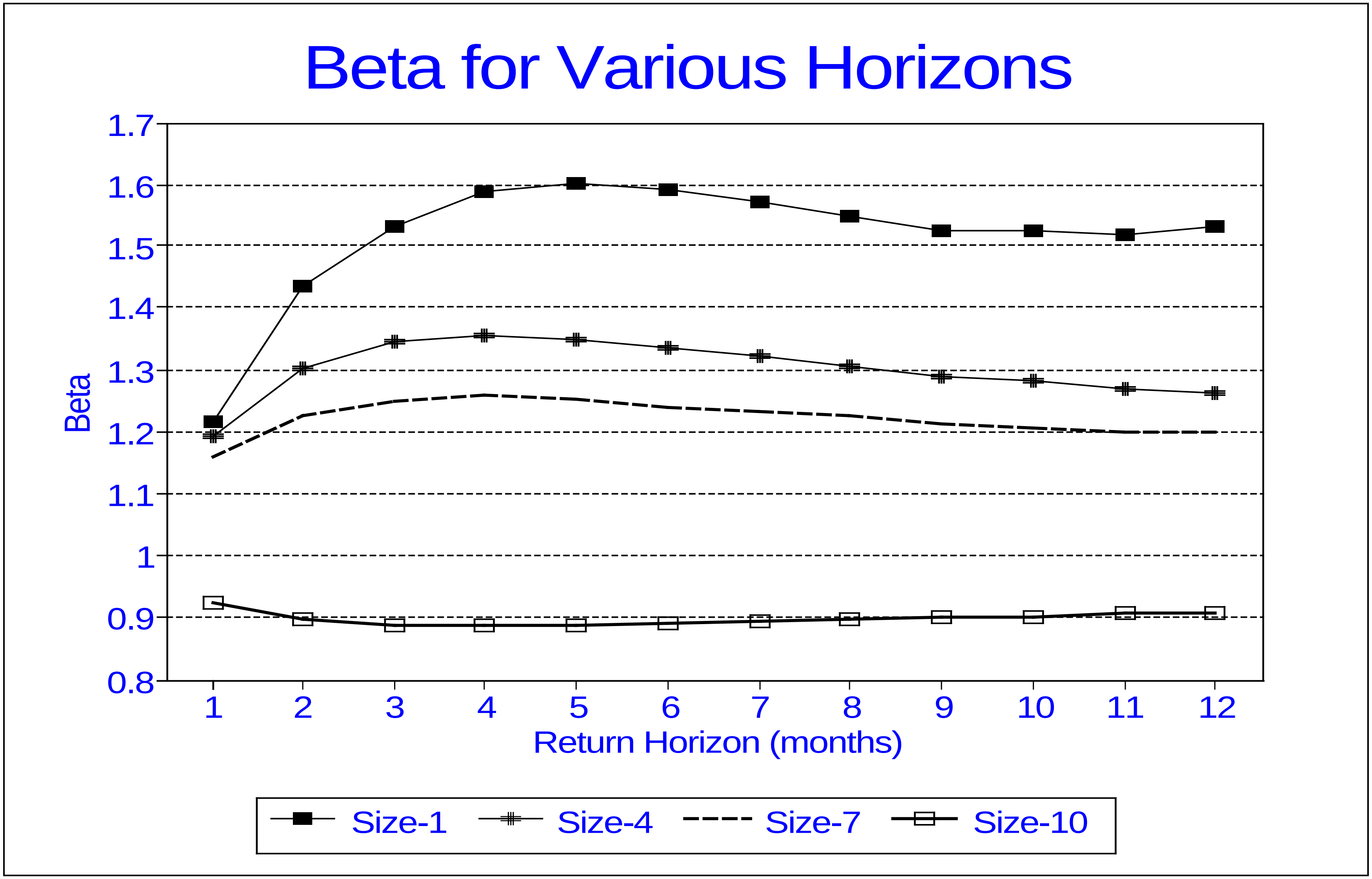
Beta for Various Horizons
Beta
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
1
2
3
4
5
6
7
8
9
10
11
12
Return Horizon (months)
Size-1
Size-4
Size-7
Size-10

Figure 1

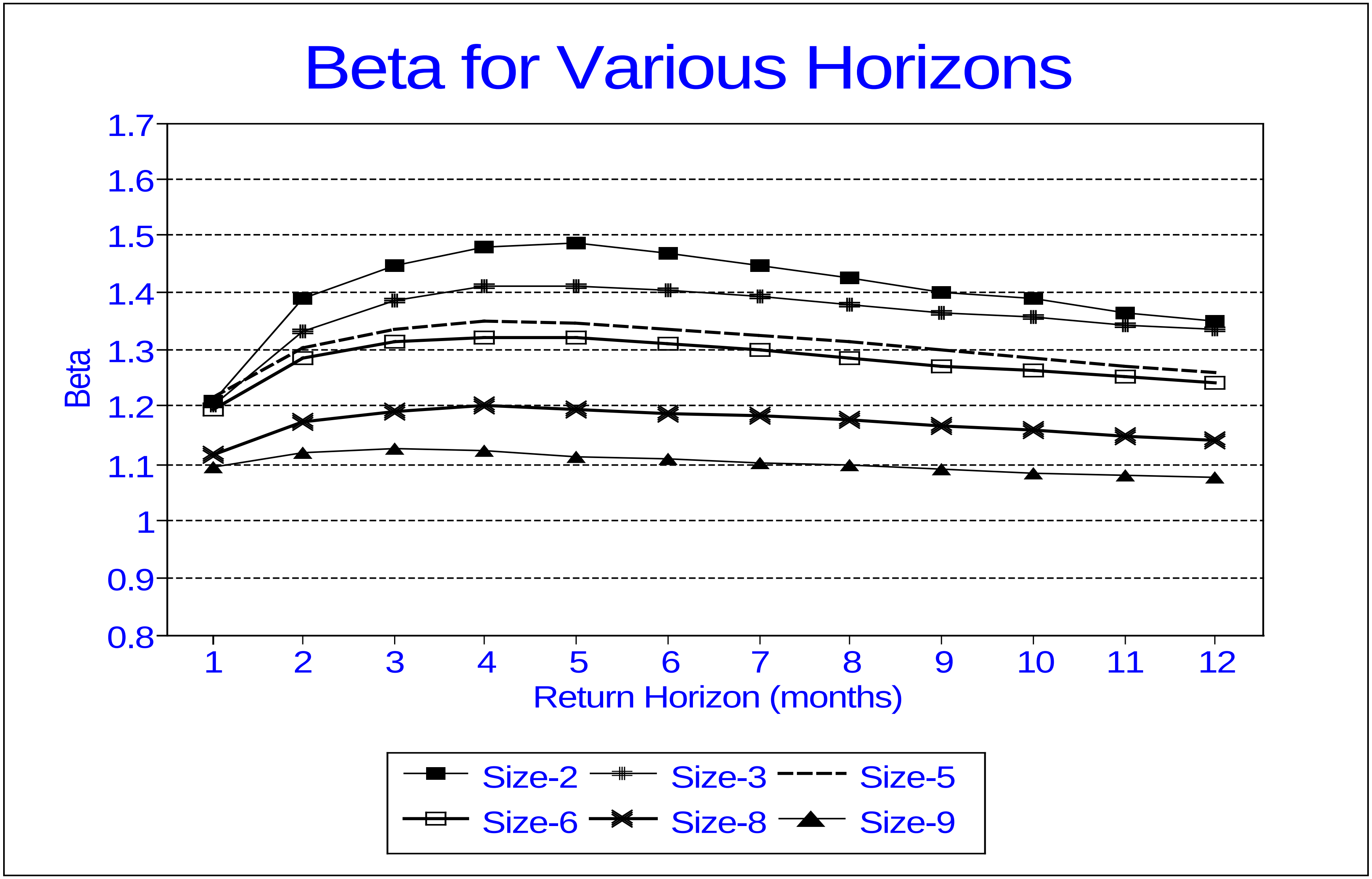
Beta for Various Horizons
Beta
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
1
2
3
4
5
6
7
8
9
10
11
12
Return Horizon (months)
Size-2
Size-3
Size-5
Size-6
Size-8
Size-9

Figure 2

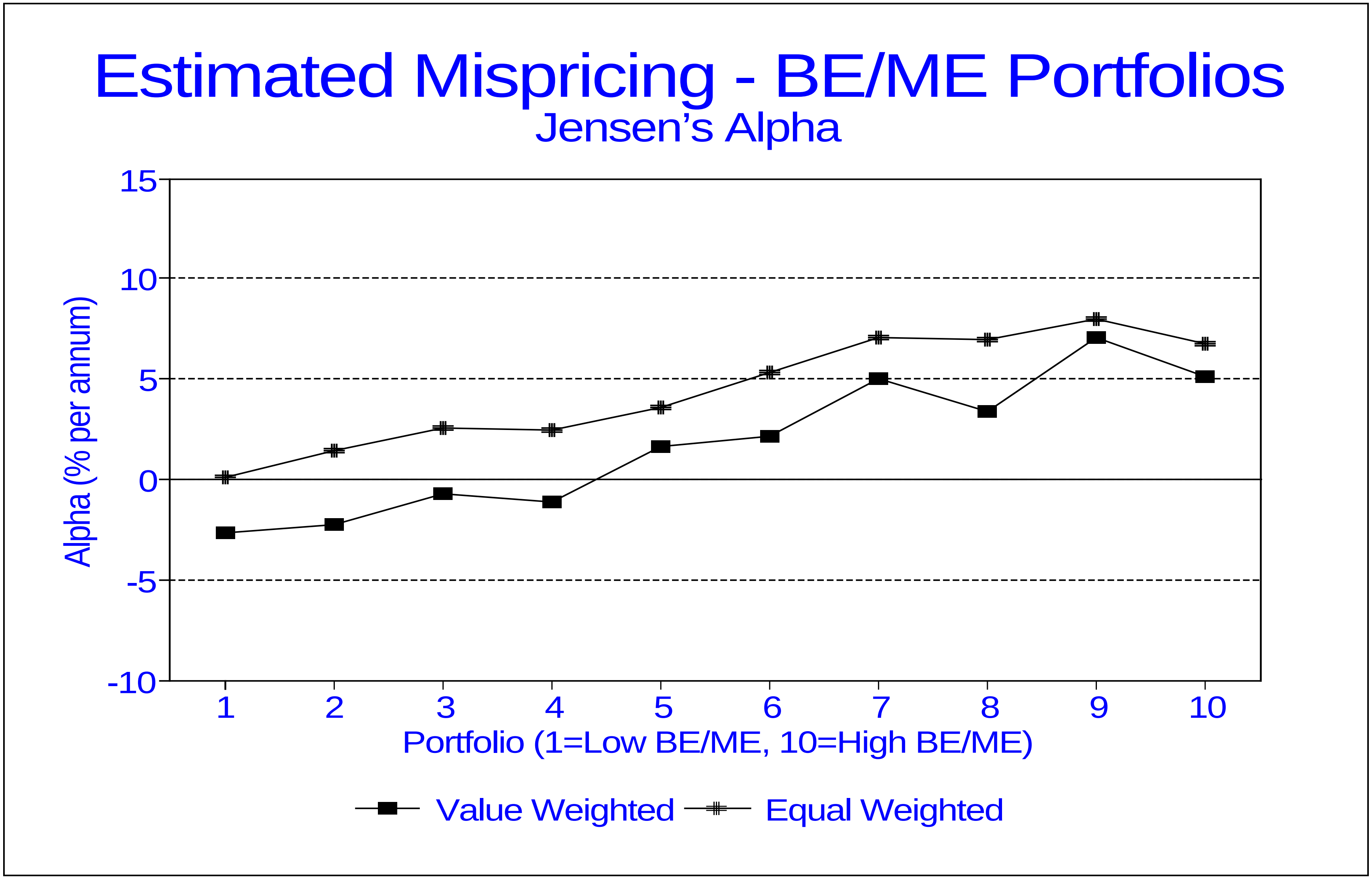
Estimated Mispricing - BE/ME Portfolios
Jensen's Alpha
Alpha (% per annum)
15
10
5
0
-5
-10
1
2
3
4
5
6
7
8
9
10
Portfolio (1=Low BE/ME, 10=High BE/ME)
Value Weighted
Equal Weighted

Figure 3

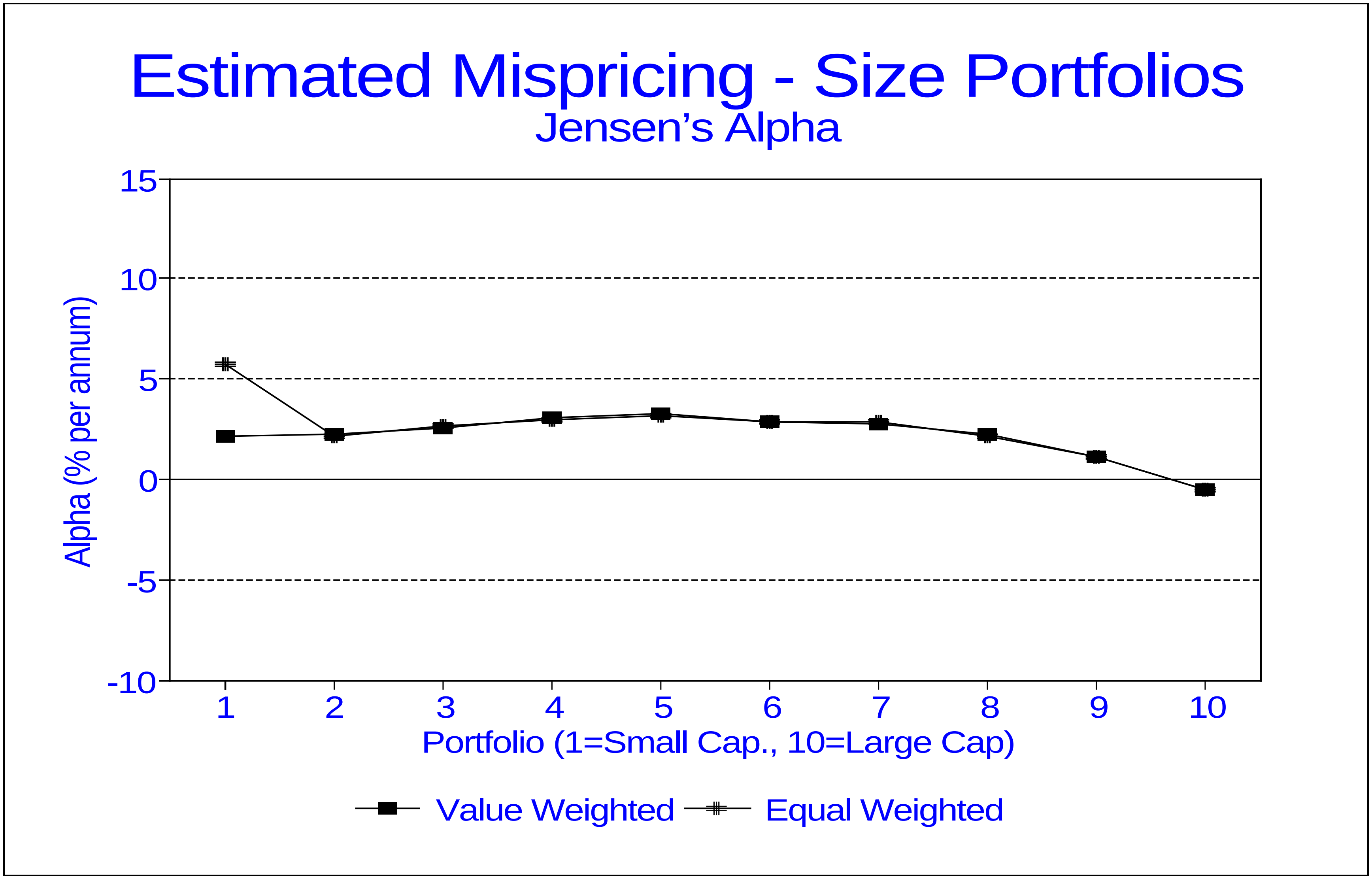
Estimated Mispricing - Size Portfolios
Jensen's Alpha
Alpha (% per annum)
15
10
5
0
-5
-10
1
2
3
4
5
6
7
8
9
10
Portfolio (1=Small Cap., 10=Large Cap)
Value Weighted
Equal Weighted

Figure 4

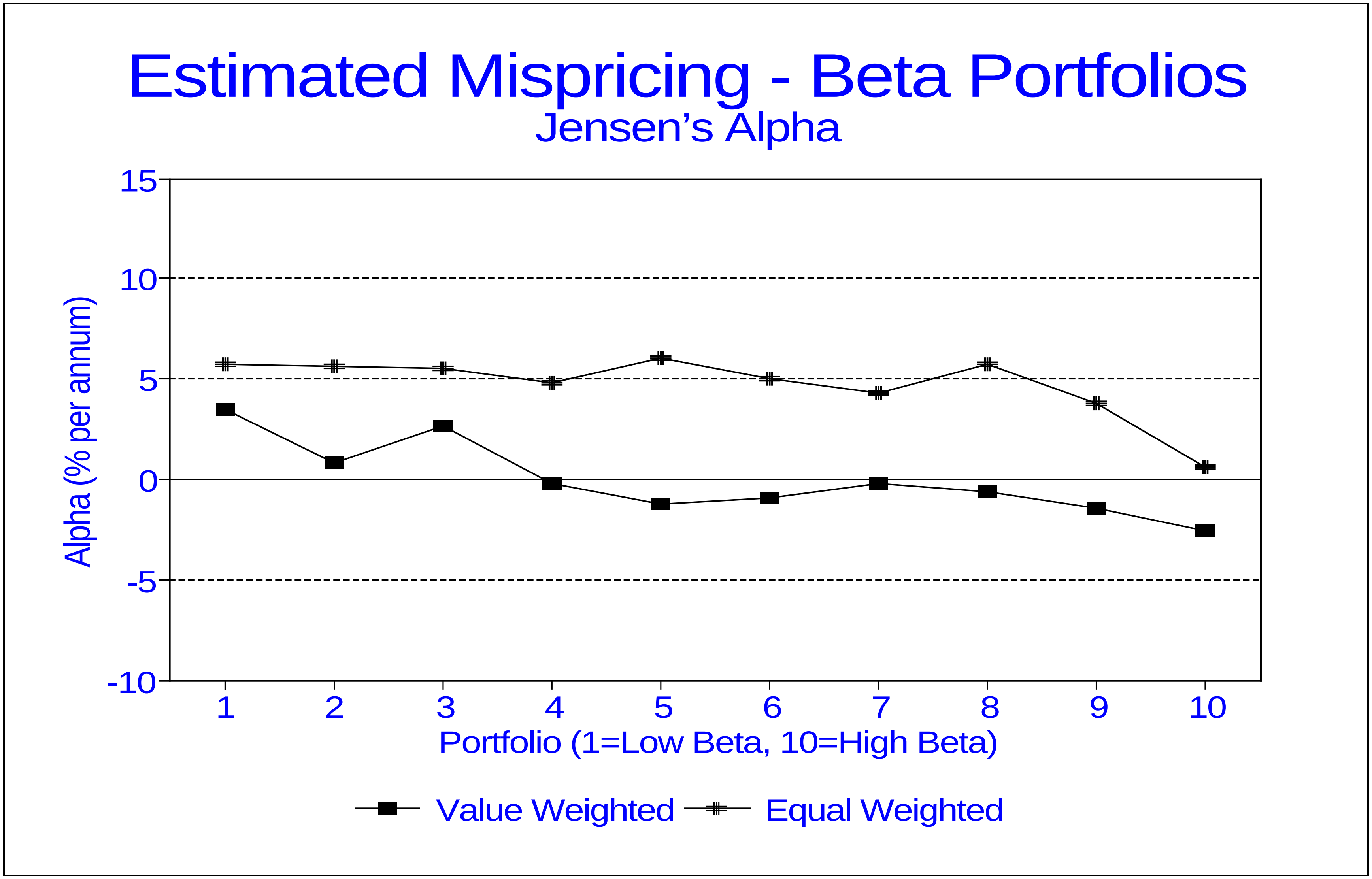
Estimated Mispricing - Beta Portfolios
Jensen's Alpha
Alpha (% per annum)
15
10
5
0
-5
-10
1
2
3
4
5
6
7
8
9
10
Portfolio (1=Low Beta, 10=High Beta)
Value Weighted
Equal Weighted

Figure 5

References

Alford, Andrew W., Jennifer J. Jones, and Mark E. Zmijewski, "Extensions and Violations of the Statutory SEC Form 10-K Filing Requirements." *Journal of Accounting and Economics* **17** (January 1994): 229-254.

Amihud, Yakov, Bent Christensen, and Haim Mendelson, "Further Evidence on the Risk-Return Relationship." Working paper, New York University, 1992.

Banz, Rolf W., "The Relationship Between Return and Market Value of Common Stock." *Journal of Financial Economics* **9** (March 1981): 3-18.

Banz, Rolf W. and William J. Breen, "Sample Dependent Results Using Accounting and Market Data: Some Evidence." *Journal Finance* **41** (September 1986): 779-793.

Basu, Sanjoy, "The Relationship Between Earnings' Yield, Market Value and Return for NYSE Common Stocks." *Journal of Financial Economics* **12** (1983): 129-156.

Bhandari, Laxmi Chand, "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence." *Journal of Finance* **43** (June 1988): 501-528.

Berk, Jonathan B., "A Critique of Size-Related Anomalies." *Review of Financial Studies* **8** (Summer 1995): 275-286.

Berk, Jonathan B. and Nobuya Takezawa, "A Re-examination of the Size Anomaly." Working paper 93-BJ-001, University of British Columbia, March 1993.

Black, Fischer, "Capital Market Equilibrium with Restricted Borrowing." *Journal of Business* **45** (July 1972): 444-455.

Black, Fischer, "Beta and Return." *Journal of Portfolio Management* **20** (Fall 1993): 8-18.

Black, Fischer, Michael C. Jensen, and Myron Scholes, "The Capital Asset Pricing Model: Some Empirical Tests." In Michael C. Jensen (ed.) *Studies in the Theory of Capital Markets*. New York: Praeger, 1972.

Brown, Philip, Allan W. Kleidon, and Terry Marsh, "New Evidence on the Nature of Size-Related Anomalies in Stock Prices." *Journal of Financial Economics* **12** (June 1983): 33-56.

Chan, K. C., "Market Value Changes and Time Varying Risk." Working paper, University of Chicago, August 1985.

Chan, Louis K. C. and Josef Lakonishok, "Are the Reports of Beta's Death Premature?" *Journal of Portfolio Management* **19** (Summer 1993): 51-62.

Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok, "Evaluating the Performance of Value versus Glamour Stocks: The Impact of Selection Bias." *Journal of Financial*

Economics **38** (July 1995): 269-296.

Chen, Nai-fu, Richard Roll, and Stephen A. Ross, "Economic Forces and the Stock Market." *Journal of Business* **59** (July 1986): 383-403.

Davis, James L., "The Cross-Section of Stock Returns and Survivorship Bias: Evidence from Delisted Stocks." Working paper, Kansas State University, August 1994a.

Davis, James L., "The Cross-Section of Realized Stock Returns: The Pre-COMPUSTAT Evidence." *Journal of Finance* **49** (December 1994b): 1579-1593.

Fama, Eugene F., *Foundations of Finance*. New York: Basic Books, 1976.

Fama, Eugene F. and Kenneth R. French, "The Cross-Section of Expected Stock Returns." *Journal of Finance* **47** (June 1992a): 427-465.

Fama, Eugene F. and Kenneth R. French, "The Economic Fundamentals of Size and Book-to-Market Equity." Working paper # 361, Center for Research in Security Prices, University of Chicago, September 1992b.

Fama, Eugene F. and Kenneth R. French, "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* **33** (February 1993): 5-56.

Fama, Eugene F. and James D. MacBeth, "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy* **71** (May/June 1973): 607-636.

Fluck, Zsuzsanna, Burton G. Malkiel, and Richard E. Quant, "The Predictability of Stock Returns and the Efficient Market Hypothesis." Memorandum #139, Financial Research Center, Princeton University, May 1993.

French, Kenneth R., "Expected Stock Returns and Dimensions of Risk: Beta, Size and Book-to-Market." In *Rethinking Beta and Expected Stock Returns*. American Society of Appraisers, April 1993.

Fuller, Russell J., "The Small-Firm Anomaly, Liquidity Premiums, and Time-Varying Returns to Size." In Diana R. Harrington and Robert A. Korajczyk (eds.) *The CAPM Controversy: Policy and Strategy Implications for Investment Management*. Charlottesville: Association for Investment Management and Research, 1993.

Gibbons, Michael R., "Multivariate Tests of Financial Models: A New Approach." *Journal of Financial Economics* **10** (1982): 3-27.

Grinold, Richard C. "Is Beta Dead Again?" *Financial Analysts Journal* **49** (July/August 1993): 28-34.

Handa, Puneet, S. P. Kothari, and Charles Wasley, "The Relation Between the Return Interval and Betas: Implications for the Size Effect." *Journal of Financial Economics* **23** (June 1989): 79-100.

Hansen, Lars Peter and Robert J. Hodrick, "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis." *Journal of Political Economy* **88** (October 1980): 829-853.

Hawawini, Gabriel and Donald B. Keim, "On the Predictability of Common Stock Returns: World-Wide Evidence." Working paper 23-92, University of Pennsylvania, 1992. Forthcoming in *Finance Handbook*, edited by R. Jarrow, V. Maksimovic, and W. Ziemba. Amsterdam: North Holland.

Jacobs, Bruce I. and Kenneth N. Levy, "Disentangling Equity Return Regularities: New Insights and Investment Opportunities." *Financial Analysts Journal* **44** (May/June 1988): 18-43.

Jaffe, Jeffrey, Donald B. Keim, and Randolph Westerfield, "Earnings Yields, Market Values, and Stock Returns." *Journal of Finance* **44** (March 1989): 135-148.

Jagannathan, Ravi, and Zhenyu Wang, "The Cross-Section of Expected Stock Returns: Do Size and Book to Market Equity Measure Systematic Risk Better Than Beta?" Working paper, University of Minnesota, August 1992.

Jagannathan, Ravi, and Zhenyu Wang, "The CAPM is Alive and Well." Research Department Staff Report #165, Federal Reserve Bank of Minneapolis, November 1993.

Kandel, Shmuel and Robert F. Stambaugh, "Portfolio Inefficiency and the Cross-Section of Mean Returns." *Journal of Finance* **50** (March 1995): 157-184.

Keim, Donald B., "Dividend Yields and Stock Returns: Implications of Abnormal January Returns." *Journal of Financial Economics* **14** (September 1985): 473-489.

Kothari, S. P., Jay Shanken, and Richard G. Sloan, "Another Look at the Cross-section of Expected Stock Returns." *Journal of Finance* **50** (March 1995): 185-224.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, "Contrarian Investment, Extrapolation, and Risk." *Journal of Finance* **49** (December 1994): 1541-1578.

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

Lev, Baruch, and Theodore Sougiannis, "Penetrating the Book-to-Market Black Box." Working paper, University of California, Berkeley, CA, August 1993.

Lintner, John, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics* (February 1965): 13-

37.

Lys, Thomas and Jowell S. Sabino, "Research Design Issues in Grouping-Based Tests." *Journal of Financial Economics* **32** (December 1992): 355-387.

MacKinlay, A. Craig, "Multifactor Models do not Explain Deviations from the CAPM." Forthcoming, *Journal of Financial Economics*.

Miller, Merton H., and Myron S. Scholes "Dividends and Taxes: Some Empirical Evidence." *Journal of Political Economy* **90** (December 1982): 1118-1141.

Mossin, Jan, "Equilibrium in a Capital Asset Market," *Econometrica* (October 1966): 768-783.

Rao, C. Radhakrishna, *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley, 1973.

Roll, Richard, and Stephen A. Ross, "On the Cross-Sectional Relation Between Expected Returns and Betas." *Journal of Finance* **49** (March 1994): 101-121.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, "Persuasive Evidence of Market Inefficiency." *Journal of Portfolio Management* **11** (Spring 1985): 9-16.

Ross, Stephen A., "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* **13** (December 1976): 341-360.

Scholes, Myron and Joseph Williams, "Estimating Betas from Nonsynchronous Data." *Journal of Financial Economics* **5** (December 1977): 309-327.

Shanken, Jay, "On the Estimation of Beta-Pricing Models." *Review of Financial Studies* **5** (1992): 1-33.

Sharpe, William F., "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance* **19** (September 1964): 425-442.

Stambaugh, Robert F., "On the Exclusion of Assets from Tests of the Two-Parameter Model: A Sensitivity Analysis." *Journal of Financial Economics* **10** (1982): 237-268.

Stattman, Dennis, "Book Values and Stock Returns." *The Chicago MBA: A Journal of Selected Papers* **4** (Spring 1980): 25-45.

Tiniç, Seha M. and Richard R. West, "Risk and Return: January vs. Rest of the Year." *Journal of Financial Economics* **13** (December 1984): 561-574.

Treynor, Jack L. "Toward a Theory of Market Value of Risky Assets," unpublished manuscript, 1961.