Journal of Financial Economics 9 (1981) 3-18. North-Holland Publishing Company

THE RELATIONSHIP BETWEEN RETURN AND MARKET VALUE OF COMMON STOCKS*

Rolf W. BANZ

Northwestern University, Evanston, IL 60201, USA

Received June 1979, final version received September 1980

This study examines the empirical relationship between the return and the total market value of NYSE common stocks. It is found that smaller firms have had higher risk adjusted returns, on average, than larger firms. This 'size effect' has been in existence for at least forty years and is evidence that the capital asset pricing model is misspecified. The size effect is not linear in the market value; the main effect occurs for very small firms while there is little difference in return between average sized and large firms. It is not known whether size *per se* is responsible for the effect or whether size is just a proxy for one or more true unknown factors correlated with size.

1. Introduction

The single-period capital asset pricing model (henceforth CAPM) postulates a simple linear relationship between the expected return and the market risk of a security. While the results of direct tests have been inconclusive, recent evidence suggests the existence of additional factors which are relevant for asset pricing. Litzenberger and Ramaswamy (1979) show a significant positive relationship between dividend yield and return of common stocks for the 1936-1977 period. Basu (1977) finds that price-earnings ratios and risk adjusted returns are related. He chooses to interpret his findings as evidence of market inefficiency but as Ball (1978) points out, market efficiency tests are often joint tests of the efficient market hypothesis and a particular equilibrium relationship. Thus, some of the anomalies that have been attributed to a lack of market efficiency might well be the result of a misspecification of the pricing model.

This study contributes another piece to the emerging puzzle. It examines the relationship between the total market value of the common stock of a firm and its return. The results show that, in the 1936-1975 period, the common stock of small firms had, on average, higher risk-adjusted returns

*This study is based on part of my dissertation and was completed while I was at the University of Chicago. I am grateful to my committee, Myron Scholes (chairman), John Gould, Roger Ibbotson, Jonathan Ingersoll, and especially Eugene Fama and and Merton Miller, for their advice and comments. I wish to acknowledge the valuable comments of Bill Schwert on earlier drafts of this paper.

110138-OPC-POD-60-1

than the common stock of large firms. This result will henceforth be referred to as the 'size effect'. Since the results of the study are not based on a particular theoretical equilibrium model, it is not possible to determine conclusively whether market value *per se* matters or whether it is only a proxy for unknown true additional factors correlated with market value. The last section of this paper will address this question in greater detail.

The various methods currently available for the type of empirical research presented in this study are discussed in section 2. Since there is a considerable amount of confusion about their relative merit, more than one technique is used. Section 3 discusses the data. The empirical results are presented in section 4. A discussion of the relationship between the size effect and other factors, as well as some speculative comments on possible explanations of the results, constitute section 5.

2. Methodologies

The empirical tests are based on a generalized asset pricing model which allows the expected return of a common stock to be a function of risk β and an additional factor ϕ, the market value of the equity.[1] A simple linear relationship of the form

$$E(R_i)=\gamma_0+\gamma_1\beta_i+\gamma_2[(\phi_i-\phi_m)/\phi_m], \tag{1}$$

is assumed, where

$E(R_i)$ = expected return on security i,
γ_0 = expected return on a zero-beta portfolio,
γ_1 = expected market risk premium,
ϕ_i = market value of security i,
ϕ_m = average market value, and
γ_2 = constant measuring the contribution of ϕ_i to the expected return of a security.

If there is no relationship between ϕ_i and the expected return, i.e., $\gamma_2=0$, (1) reduces to the Black (1972) version of the CAPM.

Since expectations are not observable, the parameters in (1) must be estimated from historical data. Several methods are available for this purpose. They all involve the use of pooled cross-sectional and time series regressions to estimate γ_0, γ_1, and γ_2. They differ primarily in (a) the assumption concerning the residual variance of the stock returns (homoscedastic or heteroscedastic in the cross-sectional), and (b) the treatment of the

[1] In the empirical tests, Φ_i and Φ_m are defined as the market proportion of security i and average market proportion, respectively. The two specifications are, of course, equivalent.

110138-OPC-POD-60-2

errors-in-variables problem introduced by the use of estimated betas in (1). All methods use a constrained optimization procedure, described in Fama (1976, ch. 9), to generate minimum variance (m.v.) portfolios with mean returns γ_i, $i=0,\dots,2$. This imposes certain constraints on the portfolio weights, since from (1)

$$E(R_p) \equiv \gamma_i = \gamma_0 \sum_j w_j + \gamma_1 \sum_j w_j \beta_j + \gamma_2 \left[\left(\sum_j w_j \phi_j - \phi_m \sum_j w_j \right) \Big/ \phi_m \right], \qquad i=0,\dots,2, \tag{2}$$

where the w_j are the portfolio proportions of each asset j, $j=1,\dots,N$. An examination of (2) shows that $\hat{\gamma}_0$ is the mean return of a standard m.v. portfolio ($\sum_j w_j = 1$) with zero beta and $\phi_p \equiv \sum_j w_j \phi_j = \phi_m$ [to make the second and third terms of the right-hand side of (2) vanish]. Similarly, $\hat{\gamma}_1$ is the mean return on a zero-investment m.v. portfolio with beta of one and $\phi_p = 0$, and $\hat{\gamma}_2$ is the mean return on a m.v. zero-investment, zero-beta portfolio with $\phi_p = \phi_m$. As shown by Fama (1976, ch. 9), this constrained optimization can be performed by running a cross-sectional regression of the form

$$R_{it} = \gamma_{0t} + \gamma_{1t} \beta_{it} + \gamma_{2t} [(\phi_{it} - \phi_{mt})/\phi_{mt}] + \varepsilon_{it}, \qquad i=1,\dots,N, \tag{3}$$

on a period-by-period basis, using estimated betas $\hat{\beta}_{it}$ and allowing for either homoscedastic or heteroscedastic error terms. Invoking the usual stationarity arguments the final estimates of the gammas are calculated as the averages of the T estimates.

One basic approach involves grouping individual securities into portfolios on the basis of market value and security beta, reestimating the relevant parameters (beta, residual variance) of the portfolios in a subsequent period, and finally performing either an ordinary least squares (OLS) regression [Fama and MacBeth (1973)] which assumes homoscedastic errors, or a generalized least squares (GLS) regression [Black and Scholes (1974)] which allows for heteroscedastic errors, on the portfolios in each time period.[2] Grouping reduces the errors-in-variables problem, but is not very efficient because it does not make use of all information. The errors-in-variables problem should not be a factor as long as the portfolios contain a reasonable number of securities.[3]

Litzenberger and Ramaswamy (1979) have suggested an alternative method which avoids grouping. They allow for heteroscedastic errors in the cross-section and use the estimates of the standard errors of the security

[2]Black and Scholes (1974) do not take account of heteroscedasticity, even though their method was designed to do so.

[3]Black, Jensen and Scholes (1972, p. 116).

110138-OPC-POD-60-3

betas as estimates of the measurement errors. As Theil (1971, p. 610) has pointed out, this method leads to unbiased maximum likelihood estimators for the gammas as long as the error in the standard error of beta is small and the standard assumptions of the simple errors-in-variables model are met. Thus, it is very important that the diagonal model is the correct specification of the return-generating process, since the residual variance assumes a critical position in this procedure. The Litzenberger–Ramaswamy method is superior from a theoretical viewpoint; however, preliminary work has shown that it leads to serious problems when applied to the model of this study and is not pursued any further.[4]

Instead of estimating equation (3) with data for all securities, it is also possible to construct arbitrage portfolios containing stocks of very large and very small firms, by combining long positions in small firms with short positions in large firms. A simple time series regression is run to determine the difference in risk-adjusted returns between small and large firms. This approach, long familiar in the efficient markets and option pricing literature, has the advantage that no assumptions about the exact functional relationships between market value and expected return need to be made, and it will therefore be used in this study.

3. Data

The sample includes all common stocks quoted on the NYSE for at least five years between 1926 and 1975. Monthly price and return data and the number of shares outstanding at the end of each month are available in the monthly returns file of the Center for Research in Security Prices (CRSP) of the University of Chicago. Three different market indices are used; this is in response to Roll's (1977) critique of empirical tests of the CAPM. Two of the three are pure common stock indices — the CRSP equally- and value-weighted indices. The third is more comprehensive: a value-weighted combination of the CRSP value-weighted index and return data on corporate and government bonds from Ibbotson and Sinquefield (1977) (henceforth 'market index').[5] The weights of the components of this index are derived from information on the total market value of corporate and government bonds in various issues of the *Survey of Current Business* (updated annually) and from the market value of common stocks in the CRSP monthly index file. The stock indices, made up of riskier assets, have both higher returns

[4] If the diagonal model (or market model) is an incomplete specification of the return generating process, the estimate of the standard error of beta is likely to have an upward bias, since the residual variance estimate is too large. The error in the residual variance estimate appears to be related to the second factor. Therefore, the resulting gamma estimates are biased.

[5] No pretense is made that this index is complete; thus, the use of quotation marks. It ignores real estate, foreign assets, etc.; it should be considered a first step toward a comprehensive index. See Ibbotson and Fall (1979).

and higher risk than the bond indices and the 'market index'.[6] A time series of commercial paper returns is used as the risk-free rate.[7] While not actually constant through time, its variation is very small when compared to that of the other series, and it is not significantly correlated with any of the three indices used as market proxies.

4. Empirical results

4.1. *Results for methods based on grouped data*

The portfolio selection procedure used in this study is identical to the one described at length in Black and Scholes (1974). The securities are assigned to one of twenty-five portfolios containing similar numbers of securities, first to one of five on the basis of the market value of the stock, then the securities in each of those five are in turn assigned to one of five portfolios on the basis of their beta. Five years of data are used for the estimation of the security beta; the next five years' data are used for the reestimation of the portfolio betas. Stock price and number of shares outstanding at the end of the five year periods are used for the calculation of the market proportions. The portfolios are updated every year. The cross-sectional regression (3) is then performed in each month and the means of the resulting time series of the gammas could be (and have been in the past) interpreted as the final estimators. However, having used estimated parameters, it is not certain that the series have the theoretical properties, in particular, the hypothesized beta. Black and Scholes (1974, p. 17) suggest that the time series of the gammas be regressed once more on the excess return of the market index. This correction involves running the time series regression (for $\hat{\gamma}_2$)

$$\hat{\gamma}_{2t} - R_{Ft} = \hat{\alpha}_2 + \hat{\beta}_2 (R_{mt} - R_{Ft}) + \hat{\varepsilon}_{2t}. \qquad (4)$$

It has been shows earlier that the theoretical β_2 is zero. (4) removes the effects of a non-zero β_2 on the return estimate $\hat{\gamma}_2$ and $\hat{\alpha}_2$ is used as the final estimator for $\hat{\gamma}_2 - R_F$. Similar corrections are performed for γ_0 and γ_1. The

[6] Mean monthly returns and standard deviations for the 1926–1975 period are:

	Mean return	Standard deviation
'Market index'	0.0046	0.0178
CRSP value-weighted index	0.0085	0.0588
CRSP equally-weighted index	0.0120	0.0830
Government bond index	0.0027	0.0157
Corporate bond index	0.0032	0.0142

[7] I am grateful to Myron Scholes for making this series available. The mean monthly return for the 1926–1975 period is 0.0026 and the standard deviation is 0.0021.

derivations of the $\hat{\beta}_i$, $i=0,\ldots,2$, in (4) from their theoretical values also allow us to check whether the grouping procedure is an effective means to eliminate the errors-in-beta problem.

The results are essentially identical for both OLS and GLS and for all three indices. Thus, only one set of results, those for the 'market index' with GLS, is presented in table 1. For each of the gammas, three numbers are reported: the mean of that time series of returns which is relevant for the test of the hypothesis of interest (i.e., whether or not $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are different from the risk-free rate and the risk premium, respectively), the associated t-statistic, and finally, the estimated beta of the time series of the gamma from (4). Note that the means are corrected for the deviation from the theoretical beta as discussed above.

The table shows a significantly negative estimate for γ_2 for the overall time period. Thus, shares of firms with large market values have had smaller returns, on average, than similar small firms. The CAPM appears to be misspecified. The table also shows that γ_0 is different from the risk-free rate. As both Fama (1976, ch. 9) and Roll (1977) have pointed out, if a test does not use the true market portfolio, the Sharpe–Lintner model might be wrongly rejected. The estimates for γ_0 are of the same magnitude as those reported by Fama and MacBeth (1973) and others. The choice of a market index and the econometric method does not affect the results. Thus, at least within the context of this study, the choice of a proxy for the market portfolio does not seem to affect the results and allowing for heteroscedastic disturbances does not lead to significantly more efficient estimators.

Before looking at the results in more detail, some comments on econometric problems are in order. The results in table 1 are based on the 'market index' which is likely to be superior to pure stock indices from a theoretical viewpoint since it includes more assets [Roll (1977)]. This superiority has its price. The actual betas of the time series of the gammas are reported in table 1 in the columns labeled $\hat{\beta}_i$. Recall that the theoretical values of β_0 and β_1 are zero and one, respectively. The standard zero-beta portfolio with return $\hat{\gamma}_0$ contains high beta stocks in short positions and low beta stocks in long positions, while the opposite is the case for the zero-investment portfolio with return $\hat{\gamma}_1$. The actual betas are all significantly different from the theoretical values. This suggests a regression effect, i.e., the past betas of high beta securities are overestimated and the betas of low beta securities are underestimated.[8] Past beta is not completely uncorrelated with the error of the current beta and the instrumental variable approach to the error-in-variables problem is not entirely successful.[9]

[8] There is no such effect for β_2 because that portfolio has both zero beta and zero investment; i.e., net holdings of both high and low beta securities are, on average, zero.

[9] This result is first documented in Brenner (1976) who examines the original Fama–McBeth (1973) time series of $\hat{\gamma}_0$.

Table 1

Portfolio estimators for γ_0, γ_1 and γ_2 based on the 'market index' with generalized least squares estimation.[a]

$$R_{it} = \hat{\gamma}_{0t} + \hat{\gamma}_{1t}\hat{\beta}_{it} + \hat{\gamma}_{2t}[(\phi_{it} - \phi_{mt})/\phi_{mt}]$$

Period	$\hat{\gamma}_0 - R_F$	$t(\hat{\gamma}_0 - R_F)$[a]	$\hat{\beta}_0$	$\hat{\gamma}_1 - (R_M - R_F)$	$t(\hat{\gamma}_1 - (R_M - R_F))$	$\hat{\beta}_1$	$\hat{\gamma}_2$	$t(\hat{\gamma}_2)$	$\hat{\beta}_2$
1936–1975	0.00450	2.76	0.45	−0.00092	−1.00	0.75	−0.00052	−2.92	0.01
1936–1955	0.00377	1.66	0.43	−0.00060	−0.80	0.80	−0.00043	−2.12	0.01
1956–1975	0.00531	2.22	0.46	−0.00138	−0.82	0.73	−0.00062	−2.09	0.01
1936–1945	0.00121	0.30	0.63	−0.00098	−0.77	0.82	−0.00075	−2.32	−0.01
1946–1955	0.00650	2.89	0.03	−0.00021	−0.26	0.75	−0.00015	−0.65	0.06
1956–1965	0.00494	2.02	0.34	−0.00098	−0.56	0.96	−0.00039	−1.27	−0.01
1966–1975	0.00596	1.43	0.49	−0.00232	−0.80	0.69	−0.00080	−1.55	0.01

[a] $\hat{\gamma}_0 - R_F$ = mean difference between return on zero beta portfolio and risk-free rate. $\hat{\gamma}_1 - (R_M - R_F)$ = mean difference between actual risk premium ($\hat{\gamma}_1$) and risk premium stipulated by Sharpe–Lintner model ($R_M - R_F$). $\hat{\gamma}_2$ = size premium. $\hat{\beta}_i$ = actual estimated market risk of $\hat{\gamma}_i$ (theoretical values: $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 0$); all β_0, β_1 are significantly different from the theoretical values. $t(\cdot)$ = t-statistic.

110138-OPC-POD-60-7

The deviations from the theoretical betas are largest for the 'market index', smaller for the CRSP value-weighted index, and smallest for the CRSP equally-weighted index. This is due to two factors: first, even if the true covariance structure is stationary, betas with respect to a value-weighted index change whenever the weights change, since the weighted average of the betas is constrained to be equal to one. Second, the betas and their standard errors with respect to the 'market index' are much larger than for the stock indices (a typical stock beta is between two and three), which leads to larger deviations — a kind of 'leverage' effect. Thus, the results in table 1 show that the final correction for the deviation of β_0 and β_1 from their theoretical values is of crucial importance for maket proxies with changing weights.

Estimated portfolio betas and portfolio market proportions are (negatively) correlated. It is therefore possible that the errors in beta induce an error in the coefficient of the market proportion. According to Levi (1973), the probability limit of $\hat{\gamma}_1$ in the standard errors-in-the-variables model is

$$\operatorname{plim} \hat{\gamma}_1 = \gamma_1/(1 + (\sigma_u^2 \cdot \sigma_2^2)/D) < \gamma_1,$$

with

$$D = (\sigma_1^2 + \sigma_u^2) \cdot \sigma_2^2 - \sigma_{12}^2 > 0,$$

where σ_1^2, σ_2^2 are the variances of the true factors β and ϕ, respectively, σ_u^2 is the variance of the error in beta and σ_{12} is the covariance of β and ϕ. Thus, the bias in $\hat{\gamma}_1$ is unambiguously towards zero for positive γ_1. The probability limit of $\hat{\gamma}_2 - \gamma_2$ is [Levi (1973)]

$$\operatorname{plim} (\hat{\gamma}_2 - \gamma_2) = (\sigma_u^2 \cdot \sigma_{12} \cdot \gamma_1)/D.$$

We find that the bias in $\hat{\gamma}_2$ depends on the covariance between β and ϕ and the sign of γ_1. If σ_{12} has the same sign as the covariance between $\hat{\beta}$ and ϕ, i.e., $\sigma_{12} < 0$, and if $\gamma_1 > 0$, then $\operatorname{plim}(\hat{\gamma}_2 - \gamma_2) < 0$, i.e., $\operatorname{plim} \hat{\gamma}_2 < \gamma_2$. If the grouping procedure is not successful in removing the error in beta, then it is likely that the reported $\hat{\gamma}_2$ overstates the true magnitude of the size effect. If this was a serious problem in this study, the results for the different market indices should reflect the problem. In particular, using the equally-weighted stock index should then lead to the smallest size effect since, as was pointed out earlier, the error in beta problem is apparently less serious for that kind of index. In fact, we find that there is little difference between the estimates.[10]

[10] For the overall time period, $\hat{\gamma}_2$ with the equally-weighted CRSP index is -0.00044, with the value weighted CRSP index -0.00044 as well as opposed to the -0.00052 for the 'market index' reported in table 1. The estimated betas of $\hat{\gamma}_0$ and $\hat{\gamma}_1$ which reflect the degree of the error in beta problems are 0.07 and 0.91, respectively, for the equally-weighted CRSP index and 0.13 and 0.87 for the value-weighted CRSP index.

Thus, it does not appear that the size effect is just a proxy for the unobservable true beta even though the market proportion and the beta of securities are negatively correlated.

The correlation coefficient between the mean market values of the twenty-five portfolios and their betas is significantly negative, which might have introduced a multicollinearity problem. One of its possible consequences is coefficients that are very sensitive to addition or deletion of data. This effect does not appear to occur in this case: the results do not change significantly when five portfolios are dropped from the sample. Revising the grouping procedure — ranking on the basis of beta first, then ranking on the basis of market proportion — also does not lead to substantially different results.

4.2. A closer look at the results

An additional factor relevant for asset pricing — the market value of the equity of a firm — has been found. The results are based on a linear model. Linearity was assumed only for convenience and there is no theoretical reason (since there is no model) why the relationship should be linear. If it is nonlinear, the particular form of the relationship might give us a starting point for the discussion of possible causes of the size effect in the next section. An analysis of the residuals of the twenty-five portfolios is the easiest way to look at the linearity question. For each month t, the estimated residual return

$$\hat{\varepsilon}_{it} = R_{it} - \hat{\gamma}_{0t} - \hat{\gamma}_{1t}\hat{\beta}_{it} - \hat{\gamma}_{2t}[(\phi_{it} - \phi_{mt})/\phi_{mt}], \qquad i = 1, \ldots, 25, \tag{5}$$

is calculated for all portfolios. The mean residuals over the forty-five year sample period are plotted as a function of the mean market proportion in fig. 1. Since the distribution of the market proportions is very skewed, a logarithmic scale is used. The solid line connects the mean residual returns of each size group. The numbers identify the individual portfolios within each group according to beta, '1' being the one with the largest beta, '5' being the one with the smallest beta.

The figure shows clearly that the linear model is misspecified.[11] The residuals are not randomly distributed around zero. The residuals of the portfolios containing the smallest firms are all positive; the remaining ones are close to zero. As a consequence, it is impossible to use $\hat{\gamma}_2$ as a simple size premium in the cross-section. The plot also shows, however, that the misspecification is not responsible for the significance of $\hat{\gamma}_2$ since the linear model underestimates the true size effect present for very small firms. To illustrate this point, the five portfolios containing the smaller firms are

[11] The nonlinearity cannot be eliminated by defining ϕ_i as the log of the market proportion.

110138-OPC-POD-60-9

deleted from the sample and the parameters reestimated. The results, summarized in table 2, show that the $\hat{\gamma}_2$ remain essentially the same. The relationship is still not linear; the new $\hat{\gamma}_2$ still cannot be used as a size premium.

Fig. 1 suggests that the main effect occurs for very small firms. Further support for this conclusion can be obtained from a simple test. We can regress the returns of the twenty-five portfolios in each result on beta alone and examine the residuals. The regression is misspecified and the residuals contain information about the size effect. Fig. 2 shows the plot of those residuals in the same format as fig. 1. The smallest firms have, on average, very large unexplained mean returns. There is no significant difference between the residuals of the remaining portfolios.

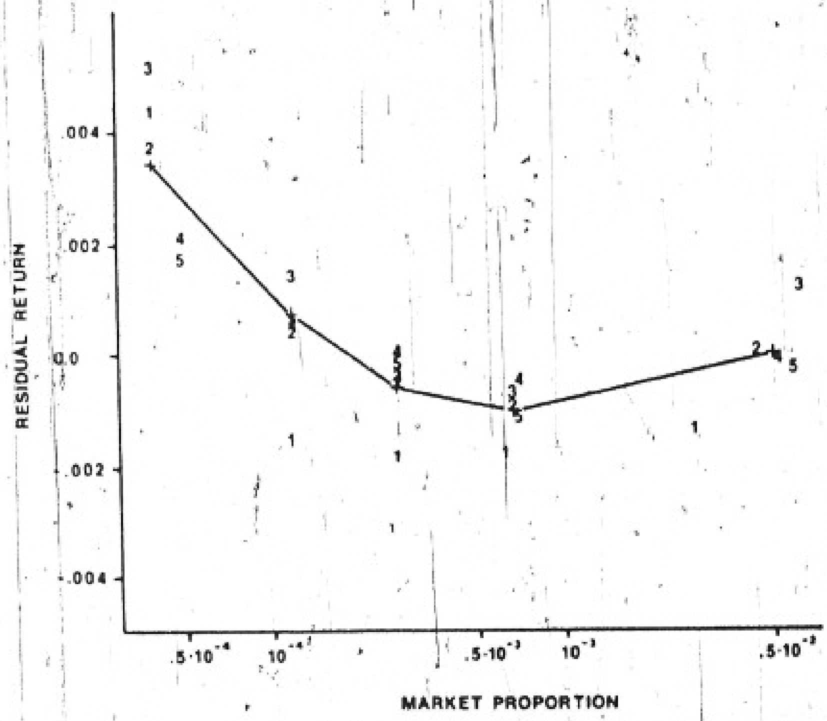


Fig. 1. Mean residual returns of portfolios (1936–1975) with equally-weighted CRSP index as market proxy. The residual is calculated with the three-factor model [eq. (3)]. The numbers 1,...,5 represent the mean residual return for the five portfolios within each size group (1: portfolio with largest beta,..., 5: portfolio with smallest beta). + represents the mean of the mean residuals of the five portfolios with similar market values.

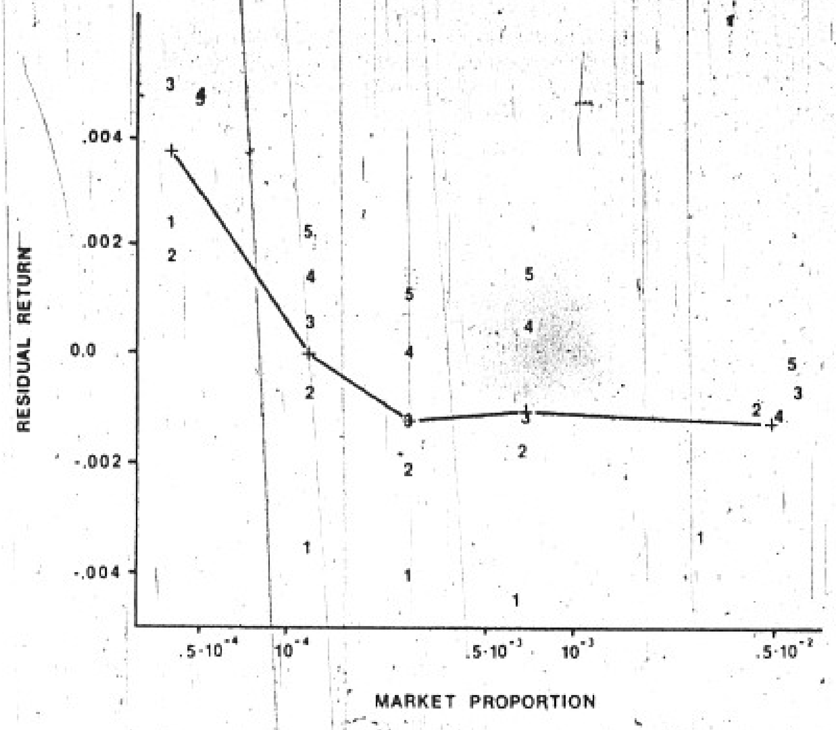


Fig. 2. Mean residual returns of portfolios (1936–1975) with equally-weighted CRSP index as market proxy. The residual is calculated with the two-factor model ($\hat{\varepsilon}_{it} = R_{it} - \hat{\gamma}_{0t} - \hat{\gamma}_{1t}\hat{\beta}_{it}$). The symbols are as defined for fig. 1.

4.3. 'Arbitrage' portfolio returns

One important empirical question still remains: How important is the size effect from a practical point of view? Fig. 2 suggests that the difference in returns between the smallest firms and the remaining ones is, on average, about 0.4 percent per month. A more dramatic result can be obtained when the securities are chosen solely on the basis of their market value.

As an illustration, consider putting equal dollar amounts into portfolios containing the smallest, largest and median-sized firms at the beginning of a year. These portfolios are to be equally weighted and contain, say, ten, twenty or fifty securities. They are to be held for five years and are rebalanced every month. They are levered or unlevered to have the same beta. We are then interested in the differences in their returns,

$$R_{1t} = R_{st} - R_{lt}, \qquad R_{2t} = R_{st} - R_{mt}, \qquad R_{3t} = R_{mt} - R_{lt}, \tag{6}$$

Table 2

Portfolio estimators for γ_2 for all 25 portfolios and for 20 portfolios (portfolios containing smallest firms deleted) based on CRSP equally weighted index with generalized least-squares estimation.[a]

	Size premium $\hat{\gamma}_2$ with	
Period	25 portfolios	20 portfolios
1936–1975	−0.00044 (−2.42)	−0.00043 (−2.54)
1936–1955	−0.00037 (−1.72)	−0.00041 (−1.88)
1956–1975	−0.00056 (−1.91)	−0.00050 (−1.91)
1936–1945	−0.00085 (−2.81)	−0.00083 (−2.48)
1946–1955	0.00003 (0.12)	−0.00003 (−0.13)
1956–1965	−0.00023) (−0.81)	−0.00017 (−0.65)
1966–1975	−0.00091 (−1.78)	−0.00085 (−1.84)

[a] t-statistic in parentheses.

where R_{st}, R_{at} and R_{lt} are the returns on the portfolios containing the smallest, median-sized and largest firms at portfolio formation time (and $R_{1t} = R_{2t} + R_{3t}$). The procedure involves (a) the calculation of the three differences in raw returns in each month and (b) running time series regressions of the differences on the excess returns of the market proxy. The intercept terms of these regressions are then interpreted as the $\bar{R}_i$, $i=1,\ldots,3$. Thus, the differences can be interpreted as 'arbitrage' returns, since, e.g., R_{1t} is the return obtained from holding the smallest firms long and the largest firms short, representing zero net investment in a zero-beta portfolio.[12] Simple equally weighted portfolios are used rather than more sophisticated minimum variance portfolios to demonstrate that the size effect is not due to some quirk in the covariance matrix.

Table 3 shows that the results of the earlier tests are fully confirmed. $\bar{R}_2$, the difference in returns between very small firms and median-size firms, is typically considerably larger than $\bar{R}_3$, the difference in returns between median-sized and very large firms. The average excess return from holding very small firms long and very large firms short is, on average, 1.52 percent

[12] No *ex post* sample bias is introduced, since monthly rebalancing includes stocks delisted during the five years. Thus, the portfolio size is generally accurate only for the first month of each period.

Table 3

Mean monthly returns on 'arbitrage' portfolios.[a]

$R_j - R_k = \hat{\alpha}_i + \hat{\beta}_i(R_m - R_f)$

$\hat{\alpha}_1$[b]	$\hat{\alpha}_2$[c]	$\hat{\alpha}_3$[d]

Table 3
Mean monthly returns on 'arbitrage' portfolios.[a]
$R_j - R_k = \hat{\alpha}_i + \hat{\beta}_i(R_m - R_F)$

	$\hat{\alpha}_1$[b]			$\hat{\alpha}_2$[c]			$\hat{\alpha}_3$[d]		
	$n=10$	$n=20$	$n=50$	$n=10$	$n=20$	$n=50$	$n=10$	$n=20$	$n=50$
Overall period									
1931–1975	0.0152	0.0148	0.0101	0.0130	0.0124	0.0089	0.0021	0.0024	0.0012
	(2.99)	(3.53)	(3.07)	(2.90)	(3.56)	(3.64)	(1.06)	(1.41)	(0.85)
Five-year subperiods									
1931–1935	0.0589	0.0597	0.0427	0.0462	0.0462	0.0326	0.0127	0.0134	0.0101
	(2.25)	(2.81)	(2.35)	(1.92)	(2.55)	(2.46)	(1.09)	(1.49)	(1.42)
1936–1940	0.0201	0.0182	0.0089	0.0118	0.0145	0.0064	0.0084	0.0037	0.0025
	(0.82)	(0.97)	(0.67)	(0.55)	(0.90)	(0.65)	(1.20)	(0.62)	(0.49)
1941–1945	0.0430	0.0408	0.0269	0.0381	0.0367	0.0228	0.0049	0.0038	0.0041
	(2.29)	(2.46)	(2.17)	(2.29)	(2.54)	(2.02)	(1.25)	(1.09)	(1.68)
1946–1950	−0.0060	−0.0046	−0.0036	−0.0058	−0.0059	−0.0029	−0.0002	−0.0104	−0.0007
	(−1.17)	(−0.97)	(−0.97)	(−1.03)	(−1.29)	(−0.83)	(−0.07)	(−0.50)	(−0.38)
1951–1955	−0.0067	−0.0011	0.0013	−0.0004	0.0026	0.0010	−0.0062	−0.0037	0.0003
	(−0.89)	(−0.21)	(0.32)	(−0.07)	(0.72)	(0.39)	(−1.29)	(−0.99)	(0.11)
1956–1960	0.0039	0.0008	0.0037	0.0007	−0.0027	0.0011	0.0031	0.0035	0.0026
	(0.67)	(0.15)	(0.89)	(0.14)	(−0.64)	(0.45)	(0.88)	(1.16)	(0.97)
1961–1965	0.0131	0.0060	0.0024	0.0096	0.0046	0.0036	0.0035	0.0014	−0.0012
	(1.38)	(0.67)	(0.31)	(1.11)	(0.72)	(0.77)	(0.59)	(0.24)	(−0.24)
1966–1970	0.0121	0.0117	0.0077	0.0129	0.0110	0.0071	0.0008	0.0007	0.0006
	(1.64)	(2.26)	(1.91)	(1.93)	(2.71)	(2.43)	(0.23)	(0.22)	(0.27)
1971–1975	0.0063	0.0108	0.0098	0.0033	0.0077	0.0083	0.0030	0.0031	0.0015
	(0.60)	(1.23)	(1.45)	(0.39)	(1.18)	(1.79)	(0.64)	(0.72)	(0.43)

[a] Equally-weighted portfolios with n securities, adjusted for differences in market risk with respect to CRSP value-weighted index, t-statistics in parentheses.
[b] Small firms held long, large firms held short.
[c] Small firms held long, median-size firms held short.
[d] Median-size firms held long, large firms held short.

110138-OPC-POD-60-13

per month or 19.8 percent on an annualized basis. This strategy, which suggests very large 'profit opportunities', leaves the investor with a poorly diversified portfolio. A portfolio of small firms has typically much larger residual risk with respect to a value-weighted index than a portfolio of very large firms with the same number of securities [Banz (1978, ch. 3)]. Since the fifty largest firms make up more than 25 percent of the total market value of NYSE stocks, it is not surprising that a larger part of the variation of the return of a portfolio of those large firms can be explained by its relation with the value-weighted market index. Table 3 also shows that the strategy would not have been successful in every five year subperiod. Nevertheless, the magnitude of the size effect during the past forty-five years is such that it is of more than just academic interest.

5. Conclusions

The evidence presented in this study suggests that the CAPM is misspecified. On average, small NYSE firms have had significantly larger risk adjusted returns than large NYSE firms over a forty year period. This size effect is not linear in the market proportion (or the log of the market proportion) but is most pronounced for the smallest firms in the sample. The effect is also not very stable through time. An analysis of the ten year subperiods show substantial differences in the magnitude of the coefficient of the size factor (table 1).

There is no theoretical foundation for such an effect. We do not even know whether the factor is size itself or whether size is just a proxy for one or more true but unknown factors correlated with size. It is possible, however, to offer some conjectures and even discuss some factors for which size is suspected to proxy. Recent work by Reinganum (1980) has eliminated one obvious candidate: the price-earnings (P/E) ratio.[13] He finds that the P/E-effect, as reported by Basu (1977), disappears for both NYSE and AMEX stocks when he controls for size but that there is a significant size effect even when he controls for the P/E-ratio, i.e., the P/E-ratio effect is a proxy for the size effect and not vice versa. Stattman (1980), who found a significant negative relationship between the ratio of book value and market value of equity and its return, also reports that this relationship is just a proxy for the size effect. Naturally, a large number of possible factors remain to be tested.[14] But the Reinganum results point out a potential problem with some of the existing negative evidence of the efficient market hypothesis. Basu believed to have identified a market inefficiency but his P/E-effect is

[13]The average correlation coefficient between P/E-ratio and market value is only 0.16 for individual stocks for thirty-eight quarters ending in 1978. But for the portfolios formed on the basis of P/E-ratio, it rises to 0.82. Recall that Basu (1977) used ten portfolios in his study.

[14]E.g., debt-equity ratios, skewness of the return distribution [Kraus and Litzenberger (1976)].

just a proxy for the size effect. Given its longevity, it is not likely that it is due to a market inefficiency but it is rather evidence of a pricing model misspecification. To the extent that tests of market efficiency use data of firms of different sizes and are based on the CAPM, their results might be at least contaminated by the size effect.

One possible explanation involving the size of the firm directly is based on a model by Klein and Bawa (1977). They find that if insufficient information is available about a subset of securities, investors will not hold these securities because of estimation risk, i.e., uncertainty about the true parameters of the return distribution. If investors differ in the amount of information available, they will limit their diversification to different subsets of all securities in the market.[15] It is likely that the amount of information generated is related to the size of the firm. Therefore, many investors would not desire to hold the common stock of very small firms. I have shown elsewhere [Banz (1978, ch. 2)] that securities sought by only a subset of the investors have higher risk-adjusted returns than those considered by all investors. Thus, lack of information about small firms leads to limited diversification and therefore to higher returns for the 'undesirable' stocks of small firms.[16] While this informal model is consistent with the empirical results, it is, nevertheless, just conjecture.

To summarize, the size effect exists but it is not at all clear why it exists. Until we find an answer, it should be interpreted with caution. It might be tempting to use the size effect, e.g., as the basis for a theory of mergers — large firms are able to pay a premium for the stock of small firms since they will be able to discount the same cash flows at a smaller discount rate. Naturally, this might turn out to be complete nonsense if size were to be shown to be just a proxy.

The preceding discussion suggests that the results of this study leave many questions unanswered. Further research should consider the relationship between size and other factors such as the dividend yield effect, and the tests should be expanded to include OTC stocks as well.

[15] Klein and Bawa (1977, p. 102).

[16] A similar result can be obtained with the introduction of fixed holding costs which lead to limited diversification as well. See Brennan (1975), Banz (1978, ch. 2) and Mayshar (1979).

References

Ball, Ray, 1978, Anomalies in relationships between securities' yields and yield surrogates, Journal of Financial Economics 6, 103–126.

Banz, Rolf W., 1978, Limited diversification and market equilibrium: An empirical analysis, Ph.D. dissertation (University of Chicago, Chicago, IL).

Basu, S., 1977, Investment performance of common stocks in relation to their price-earnings ratios: A test of market efficiency, Journal of Finance 32, June, 663–682.

Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, Journal of Business 45, July, 444–454.

Black, Fischer, and Myron Scholes, 1974, The effects of dividend yield and dividend policy on common stock prices and returns, Journal of Financial Economics 1, May, 1–22.

Black, Fischer, Michael C. Jensen and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in: M.C. Jensen, ed., Studies in the theory of capital markets (Praeger, New York) 79–121.

Brennan, Michael J., 1975, The optimal number of securities in a risky asset portfolio when there are fixed costs of transacting: Theory and some empirical evidence, Journal of Financial and Quantitative Analysis 10, Sept., 483–496.

Brenner, Menachem, 1976, A note on risk, return and equilibrium: Empirical tests, Journal of Political Economy 84, 407–409.

Fama, Eugene F., 1976, Foundations of finance (Basic Books, New York).

Fama, Eugene F. and James D. MacBeth, 1973, Risk return and equilibrium: Some empirical tests, Journal of Political Economy 71, May–June, 607–636.

Ibbotson, Roger G. and Carol L. Fall, 1979, The United States market wealth portfolio, Journal of Portfolio Management 6, 82–92.

Ibbotson, Roger G. and Rex A. Sinquefield, 1977, Stocks, bonds, bills and inflation: The past (1926–1976) and the future (1977–2000) (Financial Analysis Research Foundation).

Klein, Roger W. and Vijay S. Bawa, 1977, The effect of limited information and estimation risk on optimal portfolio diversification, Journal of Financial Economics 5, Aug., 89–111.

Kraus, Alan and Robert H. Litzenberger, 1976, Skewness preference and the valuation of risk assets, Journal of Finance 31, 1085–1100.

Levi, Maurice D., 1973, Errors in the variables bias in the presence of correctly measured variables, Econometrica 41, Sept., 985–986.

Litzenberger, Robert H. and Krishna Ramaswamy, 1979, The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence, Journal of Financial Economics 7, June, 163–195.

Mayshar, Joram, 1979, Transaction costs in a model of capital market equilibrium, Journal of Political Economy 87, 673–700.

Reinganum, Marc R., 1980, Misspecification of capital asset pricing: Empirical anomalies based on earnings yields and market values, Journal of Financial Economics, this issue.

Roll, Richard, 1977, A critique of the asset pricing theory's tests: Part I, Journal of Financial Economics 4, Jan., 120–176.

Stattman, Dennis, 1980, Book values and expected stock returns, Unpublished M.B.A. honors paper (University of Chicago, Chicago, IL).

Theil, Henri, 1971, Principles of econometrics (Wiley, New York).

U.S. Department of Commerce, Office of Business Economics, 1969, 1970, Survey of current business 49, May, 11–12; 50, May, 14.

*The Capital Asset Pricing Model: Some Empirical Tests**

FISCHER BLACK,† MICHAEL C. JENSEN,‡

AND

MYRON SCHOLES§

I. Introduction and Summary

Considerable attention has recently been given to general equilibrium models of the pricing of capital assets. Of these, perhaps the best known is the mean-variance formulation originally developed by Sharpe [1964] and Treynor [1961], and extended and clarified by Lintner [1965a, b], Mossin [1966], Fama [1968a, b], and Long [1972]. In addition Treynor [1965], Sharpe [1966], and Jensen [1968, 1969] have developed portfolio evaluation models which are either based on this asset pricing model or bear a close relation to it. In the development of the asset pricing model it is assumed that (1) all investors are single period risk-averse utility of terminal wealth maximizers and can choose among portfolios solely on the basis of mean and variance, (2) there are no taxes or

*We wish to thank Eugene Fama, John Long, David Mayers, Merton Miller, and Walter Oi for benefits obtained in conversations on these issues and D. Besenfelder, J. Shaeffer, and B. Wade for programming assistance. This research has been partially supported by the University of Rochester Systems Analysis Program under Bureau of Naval Personnel contract number N-00022-69-6-0085, The National Science Foundation under grant GS-2964, The Ford Foundation, the Wells Fargo Bank, the Manufacturers National Bank of Detroit, and the Security Trust Company. The calculations were carried out at the University of Rochester Computing Center, which is in part supported by National Science Foundation grant GJ-828.

†University of Chicago.

‡University of Rochester.

§Massachusetts Institute of Technology.

transactions costs, (3) all investors have homogeneous views regarding the parameters of the joint probability distribution of all security returns, and (4) all investors can borrow and lend at a given riskless rate of interest. The main result of the model is a statement of the relation between the expected risk premiums on individual assets and their "systematic risk." The relationship is

$$E(\tilde{R}_j) = E(\tilde{R}_M)\beta_j \qquad (1)$$

where the tildes denote random variables and

$$E(\tilde{R}_j) = \frac{E(\tilde{P}_t) - P_{t-1} + E(\tilde{D}_t)}{P_{t-1}} - r_{Ft} = \text{expected excess returns on the } j\text{th asset}$$

$\tilde{D}_t$ = dividends paid on the jth security at time t

r_{Ft} = the riskless rate of interest

$E(\tilde{R}_M)$ = expected excess returns on a "market portfolio" consisting of an investment in every asset outstanding in proportion to its value

$$\beta_j = \frac{\operatorname{cov}(\tilde{R}_j, \tilde{R}_M)}{\sigma^2(\tilde{R}_M)} = \text{the "systematic" risk of the } j\text{th asset.}$$

Relation 1 says that the expected excess return on any asset is directly proportional to its β. If we define α_j as

$$\alpha_j = E(\tilde{R}_j) - E(\tilde{R}_M)\beta_j$$

then (1) implies that the α on every asset is zero.

If empirically true, the relation given by (1) has wide-ranging implications for problems in capital budgeting, cost benefit analysis, portfolio selection, and for other economic problems requiring knowledge of the relation between risk and return. Evidence presented by Jensen [1968, 1969] on the relationship between the expected return and systematic risk of a large sample of mutual funds suggests that (1) might provide an adequate description of the relation between risk and return for securities. On the other hand, evidence presented by Douglas [1969], Lintner [1965], and most recently Miller and Scholes [1972] seems to indicate the model does not provide a complete description of the structure of security returns. In particular, the work done by Miller and Scholes suggests that the α's on individual assets depend in a systematic way on their β's: that high-beta assets tend to have negative α's, and that low-beta stocks tend to have positive α's.

Our main purpose is to present some additional tests of this asset pricing model which avoid some of the problems of earlier studies and which, we believe, provide additional insights into the nature of the structure of security returns. All previous direct tests of the model have been conducted using cross-sectional methods; primarily regression of $\bar{R}_j$, the mean excess return over a time interval for a set of securities on estimates of the systematic risk, $\hat{\beta}_j$, of each of the securities. The equation

$$\bar{R}_j = \gamma_0 + \gamma_1\hat{\beta}_j + \tilde{u}_j$$

was estimated, and contrary to the theory, γ_0 seemed to be significantly different from zero and γ_1 significantly different from $\bar{R}_M$, the slope predicted by the model. We shall show in Section III that, because of the structure of the process which appears to be generating the data, these cross-sectional tests of significance can be misleading and therefore do not provide direct tests of the validity of (1). In Section II we provide a more powerful time series test of the validity of the model, which is free of the difficulties associated with the cross-sectional tests. These results indicate that the usual form of the asset pricing model as given by (1) does not provide an accurate description of the structure of security returns. The tests indicate that the expected excess returns on high-beta assets are lower than (1) suggests and that the expected excess returns on low-beta assets are higher than (1) suggests. In other words, that high-beta stocks have negative α's and low-beta stocks have positive α's.

The data indicate that the expected return on a security can be represented by a two-factor model such as

$$E(\tilde{r}_j) = E(\tilde{r}_Z)(1-\beta_j) + E(\tilde{r}_M)\beta_j \qquad (2)$$

where the r's indicate total returns and $E(\tilde{r}_Z)$ is the expected return on a second factor, which we shall call the "beta factor," since its coefficient is a function of the asset's β. After we had observed this phenomenon, Black [1970] was able to show that relaxing the assumption of the existence of riskless borrowing and lending opportunities provides an asset pricing model which implies that, in equilibrium, the expected return on an asset will be given by (2). His results furnish an explicit definition of the beta factor, $\tilde{r}_Z$, as the return on a portfolio that has a zero covariance with the return on the market portfolio $\tilde{r}_M$. Although this model is entirely consistent with our empirical results (and provides a convenient interpretation of them), there are perhaps other plausible hypotheses consistent with the data (we shall briefly discuss several in Section V). We hasten to add that we have not attempted here to supply any direct tests of these alternative hypotheses.

The evidence presented in Section II indicates the expected excess return on an asset is not strictly proportional to its β, and we believe that this evidence, coupled with that given in Section IV, is sufficiently strong to warrant rejection of the traditional form of the model given by (1). We then show in Section III how the cross-sectional tests are subject to measurement error bias, provide a solution to this problem through grouping procedures, and show how cross-sectional methods are relevant to testing the expanded two-factor form of the model. Here we find that the evidence indicates the existance of a linear relation between risk and return and is therefore consistent with a form of the two-factor model which specifies the realized returns on each asset to be a linear function of the returns on the two factors $\tilde{r}_Z$ and $\tilde{r}_M$,

$$\tilde{r}_j = \tilde{r}_Z(1-\beta_j) + \tilde{r}_M\beta_j + \tilde{\omega}_j \qquad (2)$$

The fact that the α's of high-beta securities are negative and that the α's of low-beta securities are positive implies that the mean of the beta factor is greater than r_F. The traditional form of the capital asset pricing model as expressed by (1), could hold exactly, even if asset returns were generated by (2′), if the mean of the beta factor were equal to the risk-free rate. We show in Section IV that the mean of the beta factor has had a positive trend over the period 1931–65 and was on the order of 1.0 to 1.3% per month in the two sample intervals we examined in the period 1948–65. This seems to have been significantly different from the average risk-free rate and indeed is roughly the same size as the average market return of 1.3 and 1.2% per month over the two sample intervals in this period. This evidence seems to be sufficiently strong enough to warrant rejection of the traditional form of the model given by (1). In addition, the standard deviation of the beta factor over these two sample intervals was 2.0 and 2.2% per month, as compared with the standard deviation of the market factor of 3.6 and 3.8% per month. Thus the beta factor seems to be an important determinant of security returns.

II. *Time Series Tests of the Model*

A. Specification of the Model. Although the model of (1) which we wish to test is stated in terms of expected returns, it is possible to use realized returns to test the theory. Let us represent the returns on any security by the "market model" originally proposed by Markowitz [1959] and extended by Sharpe [1963] and Fama [1968a]

$$\tilde{R}_j = E(\tilde{R}_j) + \beta_j\tilde{R}'_M + \tilde{e}_j \qquad (3)$$

where $\tilde{R}'_M = \tilde{R}_M - E(\tilde{R}_M)$ = the "unexpected" excess market return, and $\tilde{R}'_M$ and $\tilde{e}_j$ are normally distributed random variables that satisfy:

$$E(\tilde{R}'_M) = 0 \qquad (4a)$$

$$E(\tilde{e}_j) = 0 \qquad (4b)$$

$$E(\tilde{e}_j\tilde{R}'_M) = 0 \qquad (4c)$$

The specifications of the market model, extensively tested by Fama et al. [1969] and Blume [1968], are well satisfied by the data for a large number of securities on the New York Stock Exchange. The only assumption violated to any extent is the normality assumption[1]—the estimated residuals seem to conform to the infinite variance members of the stable class of distributions rather than the normal. There are those who would explain these discrepancies from normality by certain nonstationarities in the distributions (cf. Press [1967]), which still yield finite variances. However, Wise [1963] has shown that the least-squares estimate of β_j in (3) is unbiased (although not efficient) even if the variance does not exist, and simulations by Blattberg and Sargent [1968] and Fama and Babiak [1968] also indicate that the least-squares procedures are not totally inappropriate in the presence of infinite variance stable distributions. For simplicity, therefore, we shall ignore the nonnormality issues and continue to assume normally distributed random variables where relevant.[2] However, because of these problems caution should be exercised in making literal interpretations of any significance tests.

Substituting from (1) for $E(R_j)$ in (3) we obtain

$$\tilde{R}_j = \tilde{R}_M\beta_j + \tilde{e}_j \qquad (5)$$

where $\tilde{R}_M$ is the ex post excess return on the market portfolio over the holding period of interest. If assets are priced in the market such that (1) holds over each short time interval (say a month), then we can test the traditional form of the model by adding an intercept α_j to (5) and subscripting each of the variables by t to obtain

$$\tilde{R}_{jt} = \alpha_j + \beta_j \tilde{R}_{Mt} + \tilde{e}_{jt} \tag{6}$$

which, given the assumptions of the market model, is a regression equation. If the asset pricing and the market models given by (1), (3), and (4) are valid, then the intercept α_j in (6) will be zero. Thus a direct test of the model can be obtained by estimating (6) for a security over some time period and testing to see if α_j is significantly different from zero.[3,4]

B. An Aggregation Problem. The test just proposed is simple but inefficient, since it makes use of information on only a single security whereas data is available on a large number of securities. We would like to design a test that allows us to aggregate the data on a large number of securities in an efficient manner. If the estimates of the α_j's were independent with normally distributed residuals, we could proceed along the lines outlined by Jensen [1968] and compare the frequency distributions of the "t" values for the intercepts with the theoretical distribution. However, the fact that the e_{jt} are not cross-sectionally independent, (that is, $E(\tilde{e}_{jt}\tilde{e}_{it}) \neq 0$ for $i \neq j$, cf. King [1966]); makes this procedure much more difficult.

One procedure for solving this problem which makes appropriate allowance for the effects of the nonindependence of the residuals on the standard error of estimate of the average coefficient, $\bar{\alpha}$, is to run the tests on grouped data. That is, we form portfolios (or groups) of the individual securities and estimate (6) defining $\tilde{R}_{Kt}$ to be the average return on all securities in the Kth portfolio for time t. Given this definition of $\tilde{R}_{Kt}$, $\hat{\beta}_K$ will be the average risk of the securities in the portfolio and $\hat{\alpha}_K$ will be the average intercept. Moreover, since the residual variance from this regression will incorporate the effects of any cross-sectional interdependencies in the $\tilde{e}_{jt}$ among the securities in each portfolio, the standard error of the intercept $\hat{\alpha}_K$ will appropriately incorporate the nonindependence of $\tilde{e}_{jt}$.

In addition, we wish to group our securities such that we obtain the maximum possible dispersion of the risk coefficients, β_K. If we were to construct our portfolios by using the ranked values of the $\hat{\beta}_j$, we would introduce a selection bias into the procedure. This would occur because those securities

entering the first or high-beta portfolio would tend to have positive measurement errors in their $\hat{\beta}_j$, and this would introduce positive bias in $\hat{\beta}_K$, the estimated portfolio risk coefficient. This positive bias in $\hat{\beta}_K$ will, of course, introduce a negative bias in our estimate of the intercept, $\hat{\alpha}_K$, for that portfolio. On the other hand, the opposite would occur for the lowest beta portfolio; its $\hat{\beta}_K$ would be negatively biased, and therefore our estimate of the intercept for this low-risk portfolio would be positively biased. Thus even if the traditional model were true, this selection bias would tend to cause the low-risk portfolios to exhibit positive intercepts and high-risk portfolios to exhibit negative intercepts. To avoid this bias, we need to use an instrumental variable that is highly correlated with $\hat{\beta}_j$, but that can be observed independently of $\hat{\beta}_j$. The instrumental variable we have chosen is simply an independent estimate of the β of the security obtained from past data. Thus when we estimate the group risk parameter on sample data not used in the ranking procedures, the measurement errors in these estimates will be independent of the errors in the coefficients used in the ranking and we therefore obtain unbiased estimates of $\hat{\beta}_K$ and $\hat{\alpha}_K$.

C. The Data. The data used in the tests to be described were taken from the University of Chicago Center for Research in Security Prices Monthly Price Relative File, which contains monthly price, dividend, and adjusted price and dividend information for all securities listed on the New York Stock Exchange in the period January, 1926–March, 1966. The monthly returns on the market portfolio R_{Mt} were defined as the returns that would have been earned on a portfolio consisting of an equal investment in every security listed on the NYSE at the beginning of each month. The risk-free rate was defined as the 30-day rate on U.S. Treasury Bills for the period 1948–66. For the period 1926–47 the dealer commercial paper rate[5] was used because Treasury Bill rates were not available.

D. The Grouping Procedure

1. *The ranking procedure.* Ideally we would like to assign the individual securities to the various groups on the basis of the ranked β_j (the true coefficients), but of course these are unobservable. In addition we cannot assign them on the basis of the $\hat{\beta}_j$, since this would introduce the selection bias prob-

lems discussed previously. Therefore, we must use a ranking procedure that is independent of the measurement errors in the $\hat{\beta}_j$. One way to do this is to use part of the data – in our case five years of previous monthly data – to obtain estimates $\hat{\beta}_{j0}$, of the risk measures for each security. The ranked values of the $\hat{\beta}_{j0}$ are used to assign membership to the groups. We then use data from a subsequent time period to estimate the group risk coefficients $\hat{\beta}_K$, which then contain measurement errors for the individual securities, which are independent of the errors in $\hat{\beta}_{j0}$ and hence independent of the original ranking and independent among the securities in each group.

2. *The stationarity assumptions.* The group assignment procedure just described will be satisfactory as long as the coefficients β_j are stationary through time. Evidence presented by Blume [1968] indicates this assumption is not totally inappropriate, but we have used a somewhat more complicated procedure for grouping the firms which allows for any nonstationarity in the coefficients through time.

We began by estimating the coefficient β_j, (call this estimate $\hat{\beta}_{j0}$) in (6) for the five-year period January, 1926–December, 1930 for all securities listed on the NYSE at the beginning of January 1931 for which at least 24 monthly returns were available. These securities were then ranked from high to low on the basis of the estimates $\hat{\beta}_{j0}$, and were assigned to ten portfolios[6] – the 10% with the largest $\hat{\beta}_{j0}$ to the first portfolio, and so on. The return in each of the next 12 months for each of the ten portfolios was calculated. Then the entire process was repeated for all securities listed as of January, 1932 (for which at least 24 months of previous monthly returns were available) using the immediately preceding five years of data (if available) to estimate new coefficients to be used for ranking and assignment to the ten portfolios. The monthly portfolio returns were again calculated for the next year. This process was then repeated for January, 1933, January, 1934, and so on, through January, 1965.

In this way we obtained 35 years of monthly returns on ten portfolios from the 1,952 securities in the data file. Since at each stage we used all listed securities for which at least 24 months of data were available in the immediately preceding five-year period, the total number of securities used in the analysis varied through time ranging from 582 to 1,094, and thus the number of securities contained in each portfolio changed from year to year.[7] The total number of securities

from which the portfolios were formed at the beginning of each year is given in Table 1. Each of the portfolios may be thought of as a mutual fund portfolio, which has an identity of its own, even though the stocks it contains change over time.

TABLE 1
Total Number of Securities Entering All Portfolios, by Year

Year	*Number of Securities*	*Year*	*Number of Securities*
1931	582	1949	893
1932	673	1950	928
1933	688	1951	943
1934	683	1952	966
1935	676	1953	994
1936	674	1954	1000
1937	666	1955	1006
1938	690	1956	994
1939	718	1957	994
1940	743	1958	1000
1941	741	1959	995
1942	757	1960	1021
1943	772	1961	1014
1944	778	1962	1024
1945	773	1963	1056
1946	791	1964	1081
1947	812	1965	1094
1948	842		

E. The Empirical Results

1. *The entire period.* Given the 35 years of monthly returns on each of the ten portfolios calculated as explained previously, we then calculated the least-squares estimates of the parameters α_K and β_K in (6) for each of the ten portfolios ($K = 1, \ldots, 10$) using all 35 years of monthly data (420 observations). The results are summarized in Table 2. Portfolio number 1 contains the highest-risk securities and portfolio number 10 contains the lowest-risk securities. The estimated risk coefficients range from 1.561 for portfolio 1 to 0.499 for portfolio 10. The critical intercepts, the $\hat{\alpha}_K$, are given in the second line of Table 2 and the Student "t" values are given directly below them. The correlation between the portfolio returns and the market returns, $r(\tilde{R}_K, \tilde{R}_M)$, and the autocorrelation of the residuals, $r(\tilde{e}_t, \tilde{e}_{t-1})$, are also given in Table 2. The autocorrelation appears to be quite small and the correlation between the portfolio and market returns are, as expected, quite

TABLE 2

Summary of Statistics for Time Series Tests, Entire Period (January, 1931–December, 1965)
(Sample Size for Each Regression = 420)

Item*	Portfolio Number 1	2	3	4	5	6	7	8	9	10	$\bar{R}_M$
$\hat{\beta}$	1.5614	1.3838	1.2483	1.1625	1.0572	0.9229	0.8531	0.7534	0.6291	0.4992	1.0000
$\hat{\alpha}\cdot 10^2$	−0.0829	−0.1938	−0.0649	−0.0167	−0.0543	0.0593	0.0462	0.0812	0.1968	0.2012	
$t(\hat{\alpha})$	−0.4274	−1.9935	−0.7597	−0.2468	−0.8869	0.7878	0.7050	1.1837	2.3126	1.8684	
$r(\tilde{R}, \tilde{R}_M)$	0.9625	0.9875	0.9882	0.9914	0.9915	0.9833	0.9851	0.9793	0.9560	0.8981	
$r(\tilde{e}_t, \tilde{e}_{t-1})$	0.0549	−0.0638	0.0366	0.0073	−0.0708	−0.1248	0.1294	0.1041	0.0444	0.0992	
$\sigma(\tilde{e})$	0.0393	0.0197	0.0173	0.0137	0.0124	0.0152	0.0133	0.0139	0.0172	0.0218	
$\bar{R}$	0.0213	0.0177	0.0171	0.0163	0.0145	0.0137	0.0126	0.0115	0.0109	0.0091	0.0142
σ	0.1445	0.1248	0.1126	0.1045	0.0950	0.0836	0.0772	0.0685	0.0586	0.0495	0.0891

* $\bar{R}$ = average monthly excess returns, σ = standard deviation of the monthly excess returns, r = correlation coefficient.

high. The standard deviation of the residuals $\sigma(\tilde{e}_K)$, the average monthly excess return $\bar{R}_K$, and the standard deviation of the monthly excess return, σ, are also given for each of the portfolios.

Note first that the intercepts $\hat{\alpha}$ are consistently negative for the high-risk portfolios ($\hat{\beta} > 1$) and consistently positive for the low-risk portfolios ($\hat{\beta} < 1$). Thus the high-risk securities earned less on average over this 35-year period than the amount predicted by the traditional form of the asset pricing model. At the same time, the low-risk securities earned more than the amount predicted by the model.

The significance tests given by the "t" values in Table 2 are somewhat inconclusive, since only 3 of the 10 coefficients have "t" values greater than 1.85 and, as we pointed out earlier, we should use some caution in interpreting these "t" values since the normality assumptions can be questioned. We shall see, however, that due to the existence of some nonstationarity in the relations and to the lack of more complete aggregation, these results vastly understate the significance of the departures from the traditional model.

2. *The subperiods.* In order to test the stationarity of the empirical relations, we divided the 35-year interval into four equal subperiods each containing 105 months. Table 3 presents a summary of the regression statistics of (6) calculated using the data for each of these periods for each of the ten portfolios. Note that the data for $\hat{\beta}$ in Table 3 indicate that, except for portfolios 1 and 10, the risk coefficients $\hat{\beta}_K$ were fairly stationary.

Note, however, in the sections for α and $t(\hat{\alpha})$ that the critical intercepts $\hat{\alpha}_K$, were most definitely nonstationary throughout this period. The positive α's for the high-risk portfolios in the first subperiod (January, 1931–September, 1939) indicate that these securities earned more than the amount predicted by the model, and the negative α's for the low-risk portfolios indicate they earned less than what the model predicted. In the three succeeding subperiods (October, 1939–June, 1948; July, 1948–March, 1957, and April, 1957–December, 1965) this pattern was reversed and the departures from the model seemed to become progressively larger; so much larger that six of the ten coefficients in the last subperiod seem significant. (Note that all six coefficients are those with β's most different from unity – a point we shall return to. Thus it seems unlikely that these changes were the result of chance; they most probably reflect changes in the α_K's).

TABLE 3

Summary of Coefficients for the Subperiods

Item*	Sub-period†	Portfolio Number 1	2	3	4	5	6	7	8	9	10	M_M
	1	1.5416	1.3993	1.2620	1.1813	1.0750	0.9197	0.8569	0.7510	0.6222	0.4843	1.0000
	2	1.7157	1.3196	1.1938	1.0861	0.9697	0.9254	0.8114	0.7675	0.6647	0.5626	1.0000
$\hat{\beta}$	3	1.5427	1.3598	1.1822	1.1216	1.0474	0.9851	0.9180	0.7714	0.6547	0.4868	1.0000
	4	1.4423	1.2764	1.1818	1.0655	0.9957	0.9248	0.8601	0.7800	0.6614	0.6226	1.0000
	1	0.7366	0.1902	0.3978	0.1314	−0.0650	−0.0501	−0.2190	−0.3786	−0.2128	−0.0710	
	2	−0.2197	−0.1300	−0.1224	0.0653	−0.0805	0.0914	0.1306	0.0760	0.2685	0.1478	
$\hat{\alpha} \cdot 10^2$	3	−0.4614	−0.3994	−0.1189	0.0052	0.0002	−0.0070	0.1266	0.2428	0.3032	0.2035	
	4	−0.4475	−0.2536	−0.2329	−0.0654	0.0840	0.1356	0.1218	0.3257	0.3338	0.3685	
	1	1.3881	0.6121	1.4037	0.6484	−0.3687	−0.1882	−1.0341	−1.7601	−0.7882	−0.1978	
	2	−0.4256	−0.7605	−0.8719	0.5019	−0.6288	0.8988	1.1377	0.6178	1.7853	0.8377	
$t(\hat{\alpha})$	3	−2.9030	−3.6760	−1.5160	0.0742	0.0029	−0.1010	1.8261	3.3768	3.3939	1.9879	
	4	−2.8761	−2.4603	−2.7886	−0.7722	1.1016	1.7937	1.6769	3.8772	3.0651	3.2439	
	1	0.0412	0.0326	0.0317	0.0272	0.0230	0.0197	0.0166	0.0127	0.0115	0.0099	0.0220
	2	0.0233	0.0183	0.0165	0.0168	0.0136	0.0147	0.0134	0.0122	0.0126	0.0098	0.0149
$\bar{R}$	3	0.0126	0.0112	0.0120	0.0126	0.0117	0.0109	0.0115	0.0110	0.0103	0.0075	0.0112
	4	0.0082	0.0082	0.0081	0.0087	0.0096	0.0095	0.0088	0.0101	0.0092	0.0092	0.0088
	1	0.2504	0.2243	0.2023	0.1886	0.1715	0.1484	0.1377	0.1211	0.1024	0.0850	0.1587
	2	0.1187	0.0841	0.0758	0.0690	0.0618	0.0586	0.0519	0.0494	0.0441	0.0392	0.0624
σ	3	0.0581	0.0505	0.0436	0.0413	0.0385	0.0364	0.0340	0.0289	0.0253	0.0203	0.0363
	4	0.0577	0.0503	0.0463	0.0420	0.0391	0.0365	0.0340	0.0312	0.0277	0.0265	0.0386

*$\bar{R}$ = average monthly excess returns, σ = standard deviation of monthly excess returns.

†Subperiod 1 = January, 1931–September, 1939; 2 = October, 1939–June 1948; 3 = July, 1948–March, 1957; 4 = April, 1957–December, 1965.

Note that the correlation coefficients between $\tilde{R}_{Kt}$ and $\tilde{R}_{Mt}$ given in Table 2 for each of the portfolios are all greater than 0.95 except for portfolio number 10. The lowest of the 40 coefficients in the subperiods (not shown) was 0.87, and all but two were greater than 0.90. A a result, the standard deviation of the residuals from each regression is quite small and hence so is the standard error of estimate of α, and this provides the main advantage of grouping in these tests.

III. Cross-sectional Tests of the Model

A. Tests of the Two-Factor Model. Although the time series tests discussed in Section II provide a test of the traditional form of the asset pricing model, they cannot be used to test the two-factor model directly. The cross-sectional tests, however, do furnish an opportunity to test the linearity of the relation between returns and risk implied by (2) or (2′) without making any explicit specification of the intercept. Recall that the traditional form of the model implies $\gamma_0 = 0$ and $\gamma_1 = \bar{R}_M$. The two factor model merely requires the linearity of (2) to hold for any specific cross section and allows the intercept to be nonzero. At this level of specification we shall not specify the size or even the sign of γ_0. We shall be able to make some statements on this point after a closer examination of the theory. However, we shall first examine the empirical evidence to motivate that discussion.

B. Measurement Errors and Bias in Cross-sectional Tests. We consider here the problems caused in cross-sectional tests of the model by measurement errors in the estimation of the security risk measures.[8] Let β_j represent the true (and unobservable) systematic risk of firm j and $\hat{\beta}_j = \beta_j + \tilde{\epsilon}_j$ be the measured value of the systematic risk of firm j where we assume that $\tilde{\epsilon}_j$, the measurement error, is normally distributed and for all j satisfies

$$E(\tilde{\epsilon}_j) = 0 \tag{7a}$$

$$E(\tilde{\epsilon}_j \beta_j) = 0 \tag{7b}$$

$$E(\tilde{\epsilon}_j \tilde{\epsilon}_i) = \begin{cases} 0 & i \neq j \\ \sigma^2(\tilde{\epsilon}) & i = j \end{cases} \tag{7c}$$

The traditional form of the asset pricing model and the assumptions of the market model imply that the mean excess return on a security

$$\bar{R}_j = \frac{\sum_{t=1}^{T} \tilde{R}_{jt}}{T} \qquad (8)$$

observed over T periods can be written as

$$\bar{R}_j = E(\bar{R}_j|\bar{R}_M) + \bar{e}_j = \bar{R}_M\beta_j + \bar{e}_j \qquad (9)$$

where $\bar{R}_M = \Sigma_{t=1}^{T} \tilde{R}_{Mt}/T$, $\bar{e}_j = \Sigma_{t=1}^{T} e_{jt}/T$. Now an obvious test of the traditional form of the asset pricing model is to fit

$$\bar{R}_j = \gamma_0 + \gamma_1\hat{\beta}_j + \bar{e}_j^* \qquad (10)$$

to a cross section of firms (where $\hat{\beta}_j$ is the estimated risk coefficient for each firm and $\bar{e}_j^* = \bar{e}_j - \gamma_1\bar{\epsilon}_j$) and test to see if, as implied by the theory

$$\gamma_0 = 0 \qquad \text{and} \qquad \gamma_1 = \bar{R}_M$$

There are two major difficulties with this procedure; the first involves bias due to the measurement errors in $\hat{\beta}_j$, and the second involves the apparent inadequacy of (9) as a specification of the process generating the data. The two-factor asset pricing model given by (2′) implies that γ_0 and γ_1 are random coefficients—that is, in addition to the theoretical values above, they involve a variable that is random through time. If the two-factor model is the true model, the usual significance tests on γ_0 and γ_1 are misleading, since the data from a given cross section cannot provide any evidence on the standard deviation of $\tilde{r}_Z$ and hence results in a serious underestimate of the sampling error of $\hat{\gamma}_0$ and $\hat{\gamma}_1$. Ignoring this second difficulty for the moment, we shall first consider the measurement error problems and the cross-sectional empirical evidence. The random coefficients issue and appropriate significance tests in the context of the two-factor model are discussed in more detail in Section IV.

As long as the $\hat{\beta}_j$ contain the measurement errors $\bar{\epsilon}_j$, the least-squares estimates $\hat{\gamma}_0$ and $\hat{\gamma}_1$ in (10) will be subject to the well-known errors in variables bias and will be inconsistent, (cf. Johnston [1963, Chap. VI]). That is, assuming that $\bar{e}_j$ and $\bar{\epsilon}_j$ are independent and are independent of the β_j in the cross-sectional sample,

$$\text{plim}\ \hat{\gamma}_1 = \frac{\gamma_1}{1 + \sigma^2(\bar{\epsilon})/S^2(\beta_j)} \qquad (11)$$

where $S^2(\beta_j)$ is the cross-sectional sample variance of the true risk parameters β_j. Even for large samples, then, as long as the variance of the errors in the risk measure $\sigma^2(\bar{\epsilon})$ is positive, the estimated coefficient $\hat{\gamma}_1$ will be biased toward zero and $\hat{\gamma}_0$ will therefore be biased away from zero. Hence tests of the significance of the differences $\hat{\gamma}_0 - 0$ and $\hat{\gamma}_1 - \bar{R}_M$ will be misleading.

C. The Grouping Solution to the Measurement Error Problem. We show in the Appendix that by appropriate grouping of the data to be used in estimating (10) one can substantially reduce the bias introduced through the existence of measurement errors in the $\hat{\beta}_j$. In essence the procedure amounts to systematically ordering the firms into groups (in fact by the same procedure that formed the ten portfolios used in the time series tests in Section II) and then calculating the risk measures $\hat{\beta}$ for each portfolio using the time series of portfolio returns. This procedure can greatly reduce the sampling error in the estimated risk measures; indeed, for large samples and independent errors, the sampling error is virtually eliminated. We then estimate the cross-sectional parameters of (10) using the portfolio mean returns over the relevant holding period and the risk coefficients obtained from estimation of (6) from the time series of portfolio returns. If appropriate grouping procedures are employed, this procedure will yield consistent estimates of the parameters γ_0 and γ_1 and thus will yield virtually unbiased estimates for samples in which the number of securities entering each group is large. Thus, by applying the cross-sectional test to our ten portfolios rather than to the underlying individual securities, we can virtually eliminate the measurement error problem.[9]

D. The Cross-sectional Empirical Results. Given the 35 years of monthly returns on each of the ten portfolios calculated as explained in Section II, we then estimated $\hat{\beta}_K$ and $\bar{R}_K$ $(K = 1, 2, \ldots, 10)$ for each portfolio, using all 35 years of monthly data. These estimates (see Table 2) were then used in estimating the cross-sectional relation given by (10) for various holding periods.

Figure 1 is a plot of $\bar{R}_K$ versus $\hat{\beta}_K$ for the 35-year holding period January, 1931–December, 1965. The symbol × denotes the average monthly excess return and risk of each of the ten portfolios. The symbol ☐ denotes the average excess

110138-OPC-POD-60-24

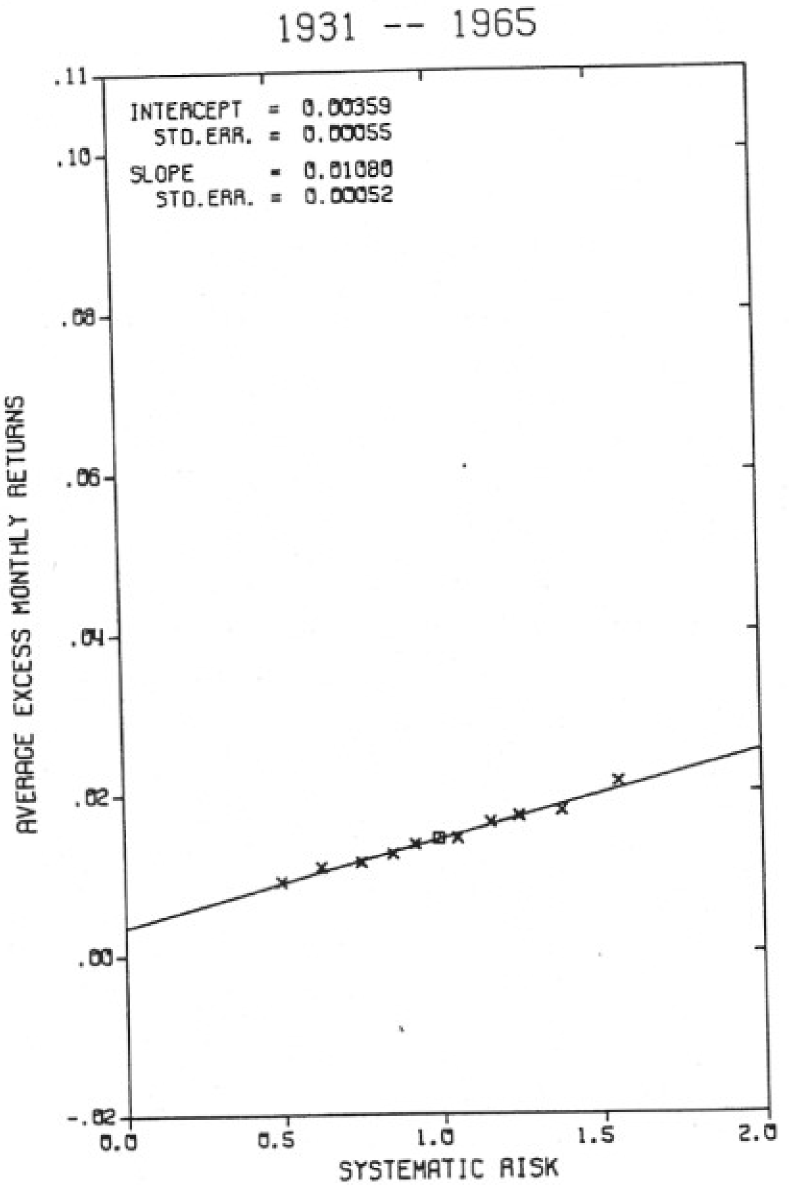


FIGURE 1 Average excess monthly returns versus systematic risk for the 35-year period 1931–65 for each of ten portfolios (denoted by ×) and the market portfolio (denoted by ☐).

return and risk of the market portfolio (which by the definition of β is equal to unity). The line represents the least-squares estimate of the relation between $\bar{R}_K$ and $\hat{\beta}_K$. The "intercept" and "slope" (with their respective standard errors given in parentheses) in the upper portion of the figure are the coefficients γ_0 and γ_1 of (10).

The traditional form of the asset pricing model implies that the intercept γ_0 in (10) should be equal to zero and the slope γ_1 should be equal to $\bar{R}_M$, the mean excess return on the market portfolio. Over this 35-year period, the average monthly excess return on the market portfolio $\bar{R}_M$, was 0.0142, and the theoretical values of the intercept and slope in Figure 1 are

$$\gamma_0 = 0 \qquad \text{and} \qquad \gamma_1 = 0.0142$$

The "t" values

$$t(\hat{\gamma}_0) = \frac{\hat{\gamma}_0}{s(\hat{\gamma}_0)} = \frac{0.00359}{0.00055} = 6.52$$

$$t(\hat{\gamma}_1) = \frac{\gamma_1 - \hat{\gamma}_1}{s(\hat{\gamma}_1)} = \frac{0.0142 - 0.0108}{0.00052} = 6.53$$

seem to indicate the observed relation is significantly different from the theoretical one. However, as we shall see, because (9) is a misspecification of the process generating the data, these tests vastly overstate the significance of the results.

We also divided the 35-year interval into four equal subperiods, and Figures 2 through 5 present the plots of the $\bar{R}_K$ versus the $\hat{\beta}_K$ for each of these intervals. In order to obtain better estimates of the risk coefficients for each of the subperiods, we used the coefficients previously estimated over the entire 35-year period.[10] The graphs indicate that the relation between return and risk is linear but that the slope is related in a nonstationary way to the theoretical slope for each period. Note that the traditional model implies that the theoretical relationship (not drawn) always passes through the two points given by the origin (0, 0) and the average market excess returns represented by ☐ in each figure. In the first subperiod (see Fig. 2) the empirical slope is steeper than the theoretical slope and then becomes successively flatter in each of the following three periods. In the last subperiod (see Fig. 5) the slope $\hat{\gamma}_1$ even has the "wrong" sign.

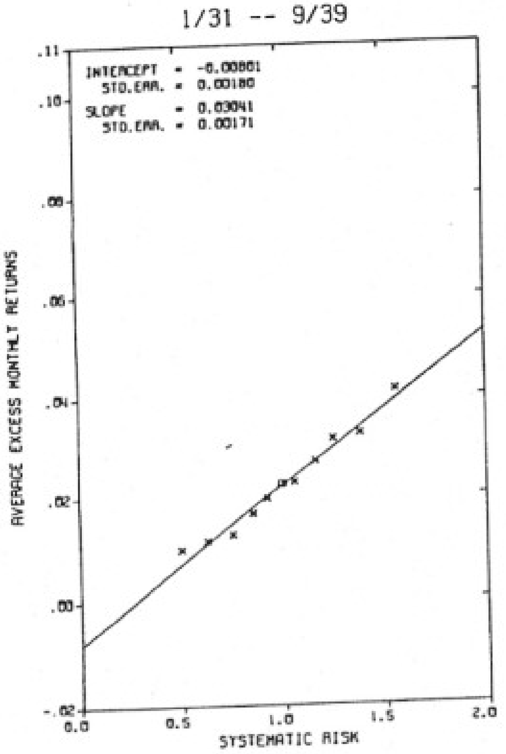


FIGURE 2 Average excess monthly returns versus systematic risk for the 105-month period January, 1931–September, 1939. Symbols as in Figure 1.

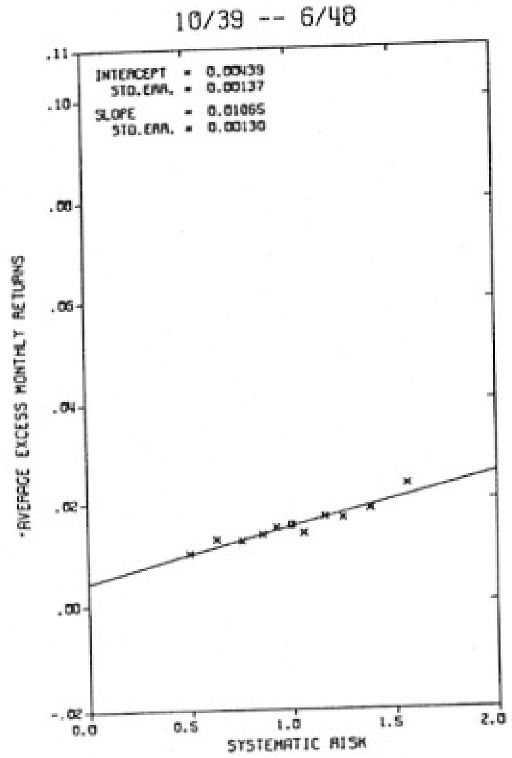


FIGURE 3 Average excess monthly returns versus systematic risk for the 105-month period October, 1939–June, 1948. Symbols as in Figure 1.

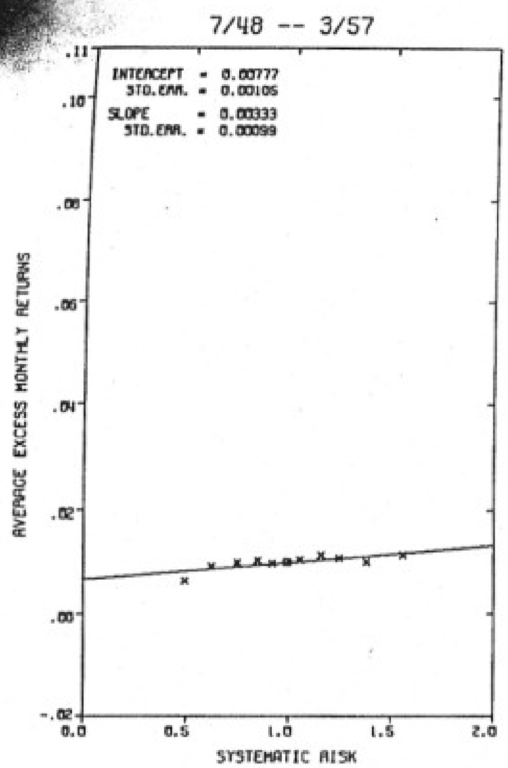


FIGURE 4 Average excess monthly returns versus systematic risk for the 105-month period July, 1948–March, 1957. Symbols as in Figure 1.

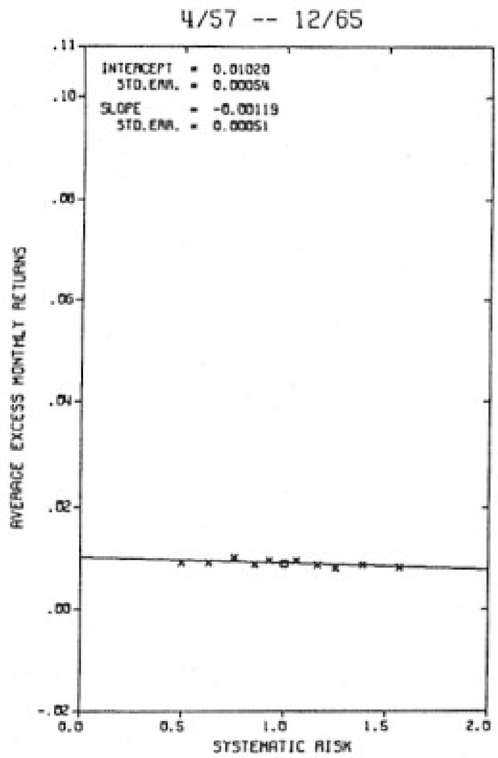


FIGURE 5 Average excess monthly returns versus systematic risk for the 105-month period April, 1957–December, 1965. Symbols as in Figure 1.

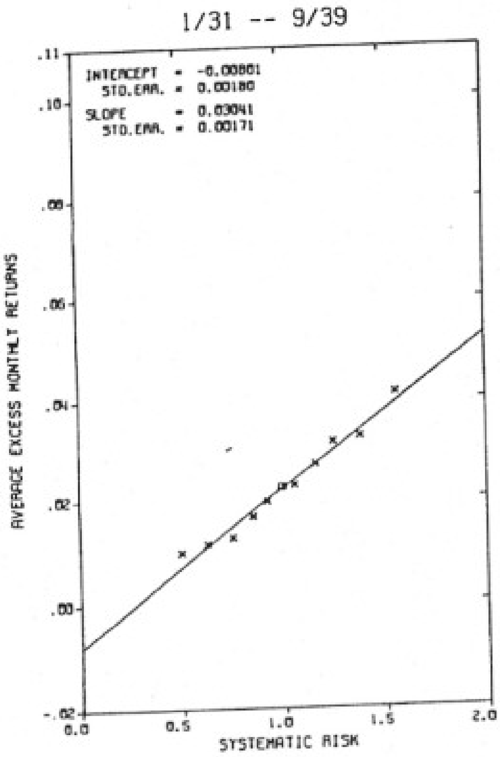


FIGURE 2 Average excess monthly returns versus systematic risk for the 105-month period January, 1931–September, 1939. Symbols as in Figure 1.

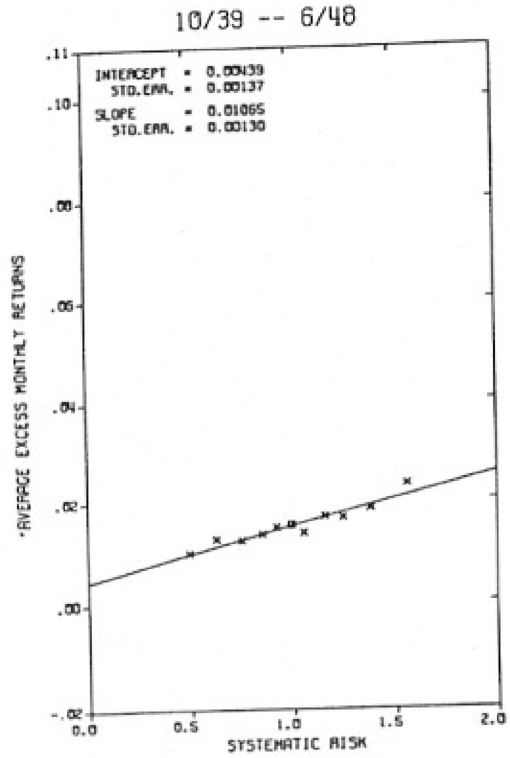


FIGURE 3 Average excess monthly returns versus systematic risk for the 105-month period October, 1939–June, 1948. Symbols as in Figure 1.

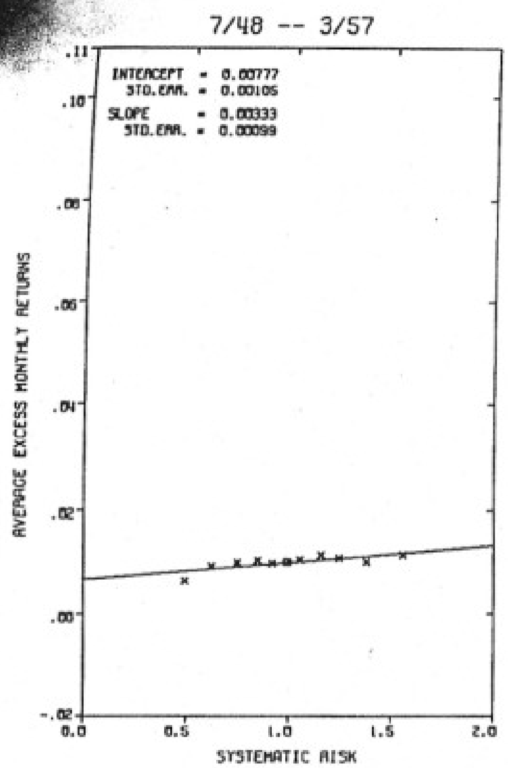


FIGURE 4 Average excess monthly returns versus systematic risk for the 105-month period July, 1948–March, 1957. Symbols as in Figure 1.

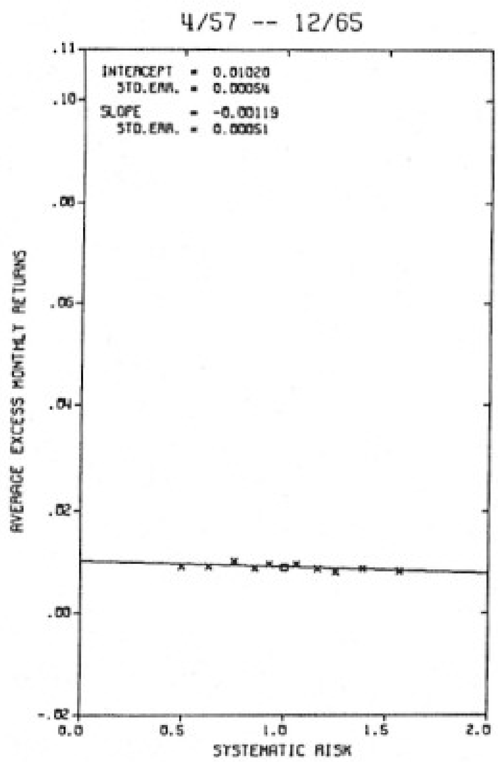


FIGURE 5 Average excess monthly returns versus systematic risk for the 105-month period April, 1957–December, 1965. Symbols as in Figure 1.

110138-OPC-POD-60-27

TABLE 4
Summary of Cross-sectional Regression Coefficients and Their t Values

	Time Period				
	Total Period	Subperiods			
	1/31–12/65	1/31–9/39	10/39–6/48	7/48–3/57	4/57–12/65
$\hat{\gamma}_0$	0.00359	−0.00801	0.00439	0.00777	0.01020
$\hat{\gamma}_1$	0.0108	0.0304	0.0107	0.0033	−0.0012
$\gamma_1 = \bar{R}_M$	0.0142	0.0220	0.0149	0.0112	0.0088
$t(\hat{\gamma}_0)$	6.52	−4.45	3.20	7.40	18.89
$t(\gamma_1 - \hat{\gamma}_1)$	6.53	−4.91	3.23	7.98	19.61

The coefficients $\hat{\gamma}_0$, $\hat{\gamma}_1$, γ_1 and the "t" values of $\hat{\gamma}_0$ and $\gamma_1 - \hat{\gamma}_1$ are summarized in Table 4 for the entire period and for each of the four subperiods. The smallest "t" value given there is 3.20, and all seem to be "significantly" different from their theoretical values. However, as we have already maintained, these "t" values are somewhat misleading because the estimated coefficients fluctuate far more in the subperiods than the estimated sampling errors indicate. This evidence suggests that the model given by (9) is misspecified. We shall now attempt to deal with this specification problem and to furnish an alternative formulation of the model.

IV. A Two-Factor Model

A. *Form of the Model.* As mentioned in the introduction, Black [1970] has shown under assumptions identical to that of the asset pricing model that, if riskless borrowing opportunities do not exist, the expected return on any asset j will be given by

$$E(\tilde{r}_j) = E(\tilde{r}_Z)(1-\beta_j) + E(\tilde{r}_M)\beta_j \tag{12}$$

where $\tilde{r}_Z$ represents the return on a "zero beta" portfolio—a portfolio whose covariance with the returns on the market portfolio $\tilde{r}_M$ is zero.[11]

Close examination of the empirical evidence from both the cross-sectional and the time series tests indicates that the results are consistent with a model that expresses the return on a security as a linear function of the market factor r_M, (with a coefficient of β_j) and a second factor r_Z, (with a coefficient of $1-\beta_j$). The function is

$$\tilde{r}_{jt} = \tilde{r}_{Zt}(1-\beta_j) + \tilde{r}_{Mt}\beta_j + \tilde{w}_{jt} \tag{13}$$

Because the coefficient of the second factor is a function of the security's β, we call this factor the beta factor. For a given holding period T, the average value of $\tilde{r}_{Zt}$ will determine the relation between $\hat{\alpha}$ and $\hat{\beta}$ for different securities or portfolios. If the data are being generated by the process given by (13) and if we estimate the single variable time series regression given by (6), then the intercept $\hat{\alpha}$ in that regression will be

$$\hat{\alpha} = (\bar{r}_Z - \bar{r}_F)(1-\hat{\beta}_j) = \bar{R}_Z(1-\hat{\beta}_j) \tag{14}$$

where $\bar{r}_Z = \sum_{t=1}^{T} \tilde{r}_{Zt}/T$ is the mean return on the beta factor over the period, $\bar{r}_F$ is the mean risk-free rate over the period, and $\bar{R}_Z$ is the difference between the two. Thus if $\bar{R}_Z$ is positive, high-beta securities will tend to have negative $\hat{\alpha}$'s, and low-beta securities will tend to have positive $\hat{\alpha}$'s. If $\bar{R}_Z$ is negative, high-beta securities will tend to have positive $\hat{\alpha}$'s, and low-beta securities will tend to have negative $\hat{\alpha}$'s.

In addition, if we estimate the cross-sectional regression given by (10), the expanded two-factor model implies that the true values of the parameters γ_0 and γ_1 will not be equal to zero and $\bar{R}_M$ but instead will be given by

$$\gamma_0 = \bar{R}_Z \quad \text{and} \quad \gamma_1 = \bar{R}_M - \bar{R}_Z$$

Hence if $\bar{R}_Z$ is positive, γ_0 will be positive and γ_1 will be less than $\bar{R}_M$. If $\bar{R}_Z$ is negative, γ_0 will be negative and γ_1 will be greater than $\bar{R}_M$.

Thus we can interpret Table 3 and Figures 2 through 5 as indicating that $\bar{R}_Z$ was negative in the first subperiod and became positive and successively larger in each of the following subperiods.

Examining (12), we see that the traditional form of the capital asset pricing model, as expressed in (1), is consistent with the present two-factor model if

$$E(\tilde{R}_Z) = 0 \tag{15}$$

and (questions of statistical efficiency aside) any test for whether α_K for a portfolio is zero is equivalent to a test for whether $E(\tilde{R}_Z)$ is zero. The results in Table 3 suggest that $E(\tilde{R}_Z)$ is not stationary through time. For example, $\hat{\alpha}_K$ for the lowest risk portfolio (number 10) is negative in the first subperiod and positive in the last subperiod, with a "t" value of 8. Thus it is unlikely that the true values of α_K were the same in

TABLE 4
Summary of Cross-sectional Regression Coefficients and Their t Values

	Time Period				
	Total Period	*Subperiods*			
	1/31–12/65	1/31–9/39	10/39–6/48	7/48–3/57	4/57–12/65
$\hat{\gamma}_0$	0.00359	−0.00801	0.00439	0.00777	0.01020
$\hat{\gamma}_1$	0.0108	0.0304	0.0107	0.0033	−0.0012
$\gamma_1 = \bar{R}_M$	0.0142	0.0220	0.0149	0.0112	0.0088
$t(\hat{\gamma}_0)$	6.52	−4.45	3.20	7.40	18.89
$t(\gamma_1 - \hat{\gamma}_1)$	6.53	−4.91	3.23	7.98	19.61

The coefficients $\hat{\gamma}_0$, $\hat{\gamma}_1$, γ_1 and the "t" values of $\hat{\gamma}_0$ and $\gamma_1 - \hat{\gamma}_1$ are summarized in Table 4 for the entire period and for each of the four subperiods. The smallest "t" value given there is 3.20, and all seem to be "significantly" different from their theoretical values. However, as we have already maintained, these "t" values are somewhat misleading because the estimated coefficients fluctuate far more in the subperiods than the estimated sampling errors indicate. This evidence suggests that the model given by (9) is misspecified. We shall now attempt to deal with this specification problem and to furnish an alternative formulation of the model.

IV. A Two-Factor Model

A. *Form of the Model.* As mentioned in the introduction, Black [1970] has shown under assumptions identical to that of the asset pricing model that, if riskless borrowing opportunities do not exist, the expected return on any asset j will be given by

$$E(\tilde{r}_j) = E(\tilde{r}_Z)(1 - \beta_j) + E(\tilde{r}_M)\beta_j \tag{12}$$

where $\tilde{r}_Z$ represents the return on a "zero beta" portfolio—a portfolio whose covariance with the returns on the market portfolio $\tilde{r}_M$ is zero.[11]

Close examination of the empirical evidence from both the cross-sectional and the time series tests indicates that the results are consistent with a model that expresses the return on a security as a linear function of the market factor r_M, (with a coefficient of β_j) and a second factor r_Z, (with a coefficient of

$1 - \beta_j$). The function is

$$\tilde{r}_{jt} = \tilde{r}_{Zt}(1 - \beta_j) + \tilde{r}_{Mt}\beta_j + \tilde{w}_{jt} \tag{13}$$

Because the coefficient of the second factor is a function of the security's β, we call this factor the beta factor. For a given holding period T, the average value of $\tilde{r}_{Zt}$ will determine the relation between $\hat{\alpha}$ and $\hat{\beta}$ for different securities or portfolios. If the data are being generated by the process given by (13) and if we estimate the single variable time series regression given by (6), then the intercept $\hat{\alpha}$ in that regression will be

$$\hat{\alpha} = (\bar{r}_Z - \bar{r}_F)(1 - \hat{\beta}_j) = \bar{R}_Z(1 - \hat{\beta}_j) \tag{14}$$

where $\bar{r}_Z = \sum_{t=1}^{T} \tilde{r}_{Zt}/T$ is the mean return on the beta factor over the period, $\bar{r}_F$ is the mean risk-free rate over the period, and $\bar{R}_Z$ is the difference between the two. Thus if $\bar{R}_Z$ is positive, high-beta securities will tend to have negative $\hat{\alpha}$'s, and low-beta securities will tend to have positive $\hat{\alpha}$'s. If $\bar{R}_Z$ is negative, high-beta securities will tend to have positive $\hat{\alpha}$'s, and low-beta securities will tend to have negative $\hat{\alpha}$'s.

In addition, if we estimate the cross-sectional regression given by (10), the expanded two-factor model implies that the true values of the parameters γ_0 and γ_1 will not be equal to zero and $\bar{R}_M$ but instead will be given by

$$\gamma_0 = \bar{R}_Z \qquad \text{and} \qquad \gamma_1 = \bar{R}_M - \bar{R}_Z$$

Hence if $\bar{R}_Z$ is positive, γ_0 will be positive and γ_1 will be less than $\bar{R}_M$. If $\bar{R}_Z$ is negative, γ_0 will be negative and γ_1 will be greater than $\bar{R}_M$.

Thus we can interpret Table 3 and Figures 2 through 5 as indicating that $\bar{R}_Z$ was negative in the first subperiod and became positive and successively larger in each of the following subperiods.

Examining (12), we see that the traditional form of the capital asset pricing model, as expressed in (1), is consistent with the present two-factor model if

$$E(\tilde{R}_Z) = 0 \tag{15}$$

and (questions of statistical efficiency aside) any test for whether α_K for a portfolio is zero is equivalent to a test for whether $E(\tilde{R}_Z)$ is zero. The results in Table 3 suggest that $E(\tilde{R}_Z)$ is not stationary through time. For example, $\hat{\alpha}_K$ for the lowest risk portfolio (number 10) is negative in the first subperiod and positive in the last subperiod, with a "t" value of 8. Thus it is unlikely that the true values of α_K were the same in the two subperiods (each of which contains 105 observations) and thus unlikely that the true values of $E(R_Z)$ were the same in the two subperiods, and we shall derive formal tests of this proposition below.

The existence of a factor $\tilde{R}_Z$ with a weight proportional to $1-\beta_j$ in most securities is also suggested by the unreasonably high "t" values[12] obtained in the cross-sectional regressions, as given in Table 4. Since γ_0 and γ_1 involve $\tilde{R}_Z$, which is a random variable from cross section to cross section, and since no single cross-sectional run can provide any information whatsoever on the variability of $\tilde{R}_Z$, this element is totally ignored in the usual calculation of the standard errors of γ_0 and γ_1. It is not surprising, therefore, that each individual cross-sectional result seems so highly significant but so totally different from any other cross-sectional relationship. Of course the presence of infinite-variance stable distributions will also contribute to this type of phenomenon.

In addition, in an attempt to determine whether the linearity observed in Figures 1 through 5 was in some way due to the averaging involved in the long periods presented there, we replicated those plots for our ten portfolios for 17 separate two-year periods from 1932 to 1965. These results, which also exhibit a remarkable linearity, are presented in Figures 6a and 6b. Since the evidence seems to indicate that the all-risky asset model describes the data better than the traditional model, and since the definition of our "riskless" interest rate was somewhat arbitrary in any case, these plots were derived from calculations on the raw return data with no reference whatsoever to the "risk-free" rate defined earlier (including the recalculation of the ten portfolios and the estimation of the β_j). Figures 7 through 11 contain a replication of Figures 1 through 5 calculated on the same basis. These results indicate that the basic findings summarized previously cannot be be attributed to misspecification of the riskless rate.

In summary, then, the empirical results suggest that the returns on different securities can be written as a linear function of two factors as given in (13), that the expected excess return on the beta factor $\tilde{R}_Z$ has in general been positive, and that the expected return on the beta factor has been higher in more recent subperiods than in earlier subperiods.

B. Explicit Estimation of the Beta Factor and a Crucial Test of the Model. Since the traditional form of the asset

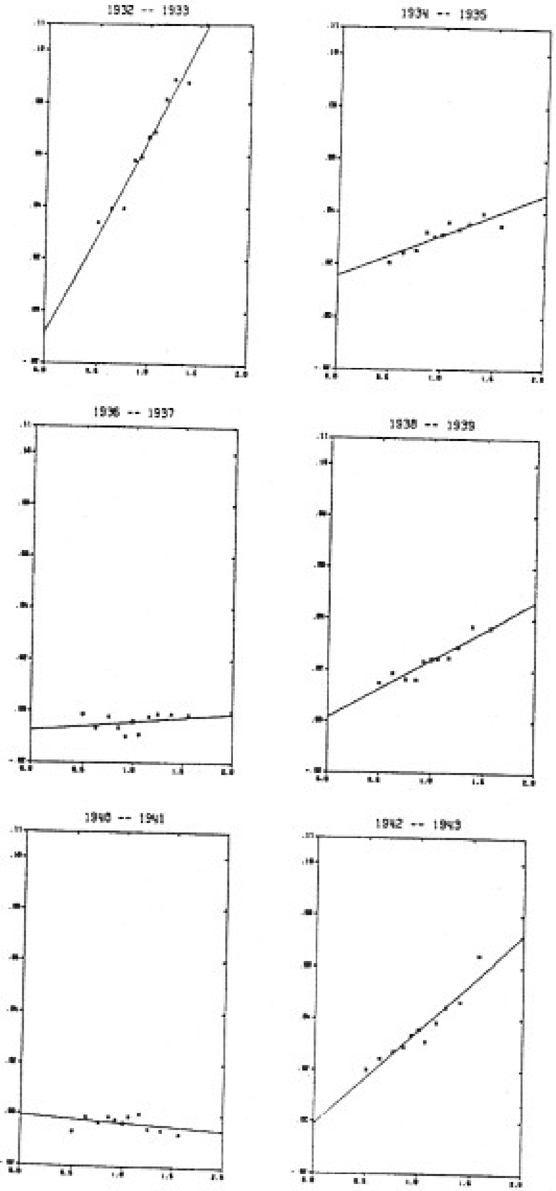
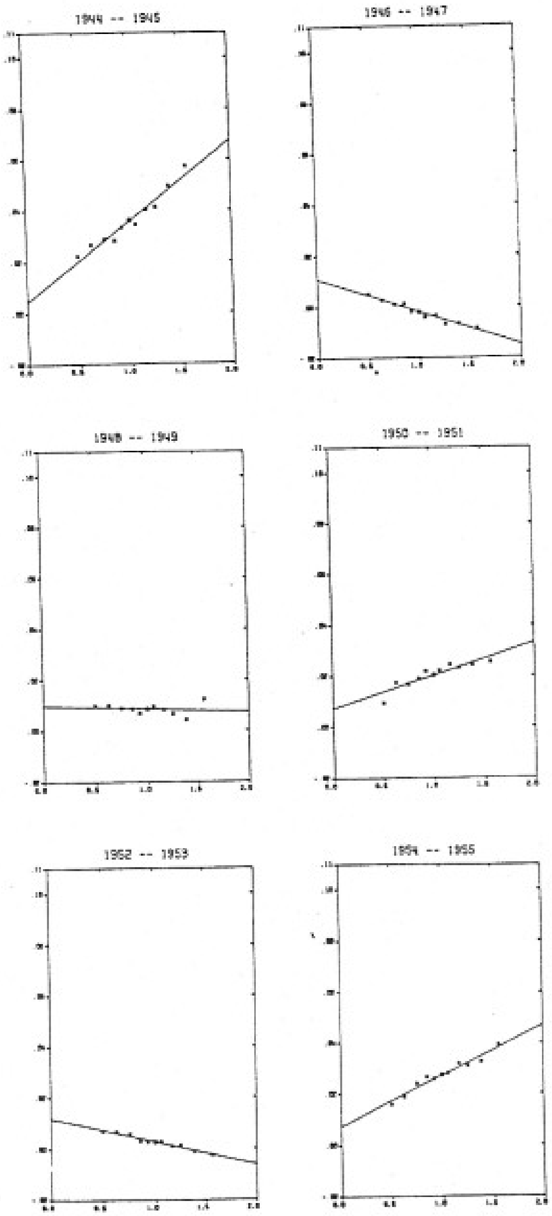


FIGURE 6 Average monthly returns versus systematic risk for 17 non-overlapping two-year periods from 1932 to 1965.

FIGURE 6 (continued)



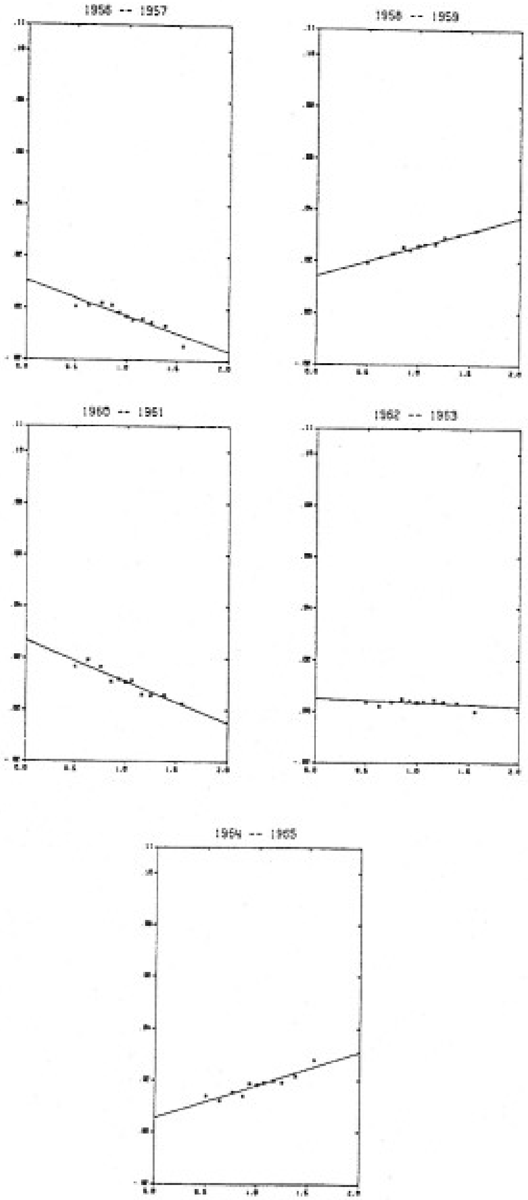


FIGURE 6 (continued)

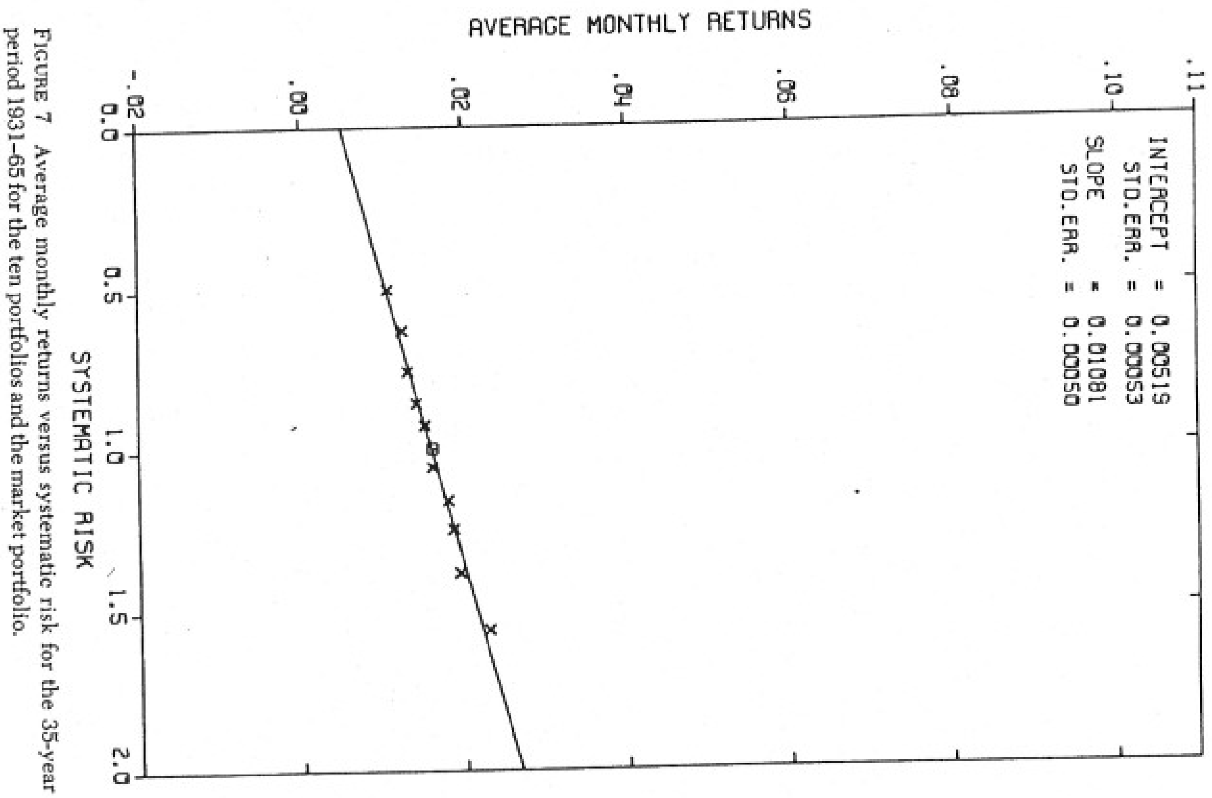


FIGURE 7 Average monthly returns versus systematic risk for the 35-year period 1931–65 for the ten portfolios and the market portfolio.

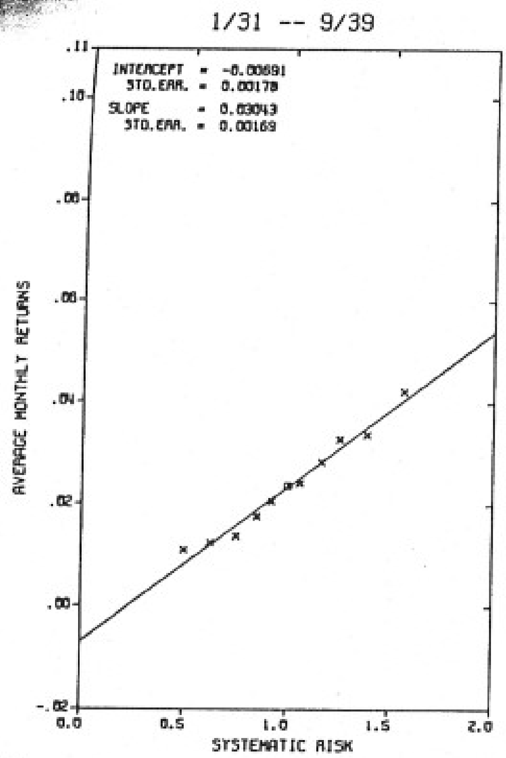


FIGURE 8 Average monthly returns versus systematic risk for the 105-month period January, 1931 – September, 1939.

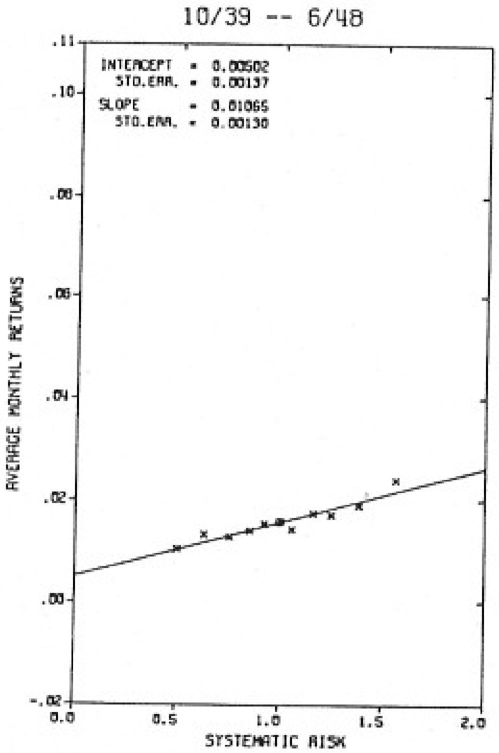


FIGURE 9 Average monthly returns versus systematic risk for the 105-month period October, 1939 – June, 1948.

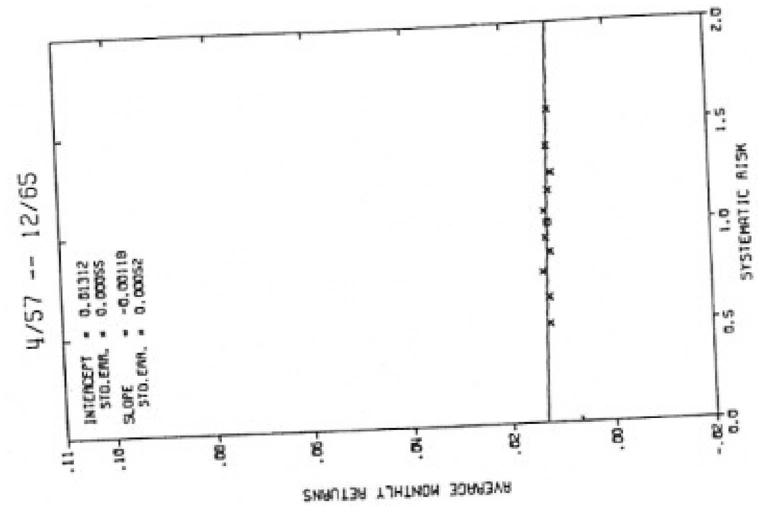


FIGURE 11 Average monthly returns versus systematic risk for the 105-month period April, 1957 – December, 1965.

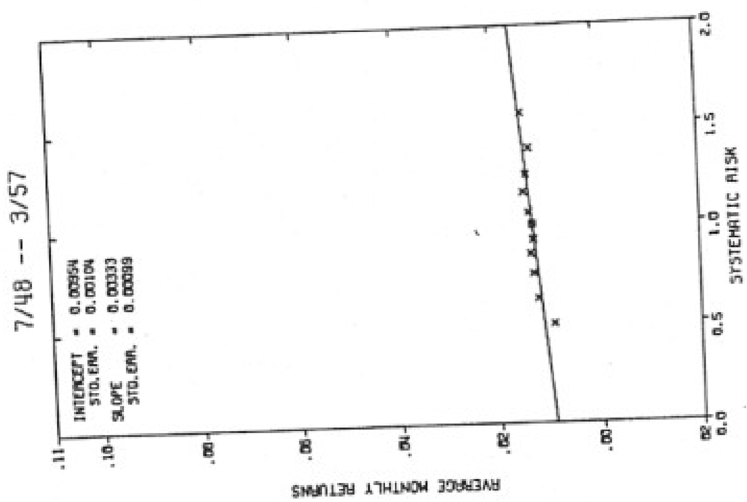


FIGURE 10 Average monthly returns versus systematic risk for the 105-month period July, 1948 – March, 1957.

pricing model is consistent with the existence of the beta factor as long as the excess returns on the beta factor have a zero mean,[13] our purpose here is to provide a procedure for explicit estimation of the time series of the factor. Given such a time series, we can then make explicit estimates of the significance of its mean excess return rather than depending mainly on an examination of the $\hat{\alpha}_j$ for high- and low-beta securities. Solving (13) for $\tilde{r}_{Zt}$ plus the error term, we have an estimate $\hat{r}_{Zjt}$, of $\tilde{r}_{Zt}$

$$\hat{r}_{Zjt} = \frac{1}{(1-\beta_j)}[\tilde{r}_j - \beta_j \tilde{r}_{Mt}] = \tilde{r}_{Zt} + \tilde{u}_{jt} \qquad (16)$$

where $\tilde{u}_{jt} = \tilde{\omega}_{jt}/(1-\beta_j)$. We subscript $\hat{r}_{Zt}$ by j to denote that this is an estimate of $\tilde{r}_{Zt}$ obtained from the jth asset or portfolio. Now, since we can obtain as many separate estimates of $\tilde{r}_{Zt}$ as we have securities or portfolios, we can formulate a combined estimate

$$r^{\circ}_{Zt} = \sum_j h_j \hat{r}_{Zjt} \qquad (17)$$

which is a linear combination of the $\hat{r}_{Zjt}$, to provide a much more efficient estimate of $\tilde{r}_{Zt}$. The problem is to find that linear combination of the $\hat{r}_{Zjt}$ which minimizes the error variance in the estimate of $\tilde{r}_{Zt}$. That is, we want to

$$\min_{h_j} E(r^{\circ}_{Zt} - \tilde{r}_{Zt})^2 = \min_{h_j} E\Big(\sum_j h_j \hat{r}_{Zjt} - \tilde{r}_{Zt}\Big)^2$$

subject to $\sum_j h_j = 1$, since we want an unbiased estimate. From the Lagrangian we obtain the first-order conditions

$$h_j \sigma^2(\tilde{u}_j) - \lambda = 0 \qquad j = 1, 2, \ldots, N \qquad (18)$$

where λ is the Lagrangian multiplier and N is the total number of securities or nonoverlapping portfolios. These conditions imply that

$$\frac{h_j}{h_i} = \frac{\sigma^2(\tilde{u}_i)}{\sigma^2(\tilde{u}_j)} \qquad \text{for all } i \text{ and } j \qquad (19)$$

which implies that the optimal weights h_j are proportional to $1/\sigma^2(\tilde{u}_j)$. That is,

$$h_j = \frac{K}{\sigma^2(\tilde{u}_j)} \qquad j = 1, 2, \ldots, N \qquad (20)$$

where $K = 1/\Sigma_j [1/\sigma^2(\tilde{u}_j)]$ is a normalizing constant. But from the definition of $\tilde{u}_j$ we know that $\sigma^2(\tilde{u}_j) = \sigma^2(\tilde{w}_j)/(1-\beta_j)^2$, so

$$h_j = \frac{K(1-\beta_j)^2}{\sigma^2(\tilde{w}_j)} \tag{21}$$

Equation (21) makes sense, for we are then weighting the estimates in proportion to $(1-\beta_j)^2$ and inversely proportional to $\sigma^2(\tilde{w}_j)$. However, since we cannot observe $\sigma^2(\tilde{w}_j)$ directly,[14] we are forced, for lack of explicit estimates, to assume that the $\sigma^2(\tilde{w}_j)$ are all identical and to use as our weights

$$h_j = K'(1-\beta_j)^2 \tag{22}$$

where $K' = 1/\Sigma_j(1-\beta_j)^2$.

Equations (17) and (22) thus provide an unbiased and (approximately) efficient procedure for estimating $\bar{r}_{Zt}$ utilizing all available information. However, there is a problem of bias involved in actually applying this procedure to the security data. The coefficient β_j is of course unobservable, and in general if we use our estimates $\hat{\beta}_j$ in the weighting procedure we will introduce bias into our estimate of $\bar{r}_{Zt}$. To understand this, recall that $\hat{\beta}_j = \beta_j + \epsilon_j$, substitute this into (13) with the necessary additions and subtractions, and solve for the estimate

$$\hat{r}_{Zjt} = \frac{\bar{r}_{jt} - \hat{\beta}_j \bar{r}_{Mt}}{(1-\hat{\beta}_j)} = \frac{\bar{r}_{Zt}(1-\beta_j) + \tilde{w}_j - \tilde{\epsilon}_j \bar{r}_{Mt}}{(1-\hat{\beta}_j)}$$

Substituting this into (17), using (22), rearranging terms, and taking the probability limit, we have

$$\operatorname*{plim}_{N\to\infty} r_{Zt}^{\circ} = \frac{C_i[S^2(\beta) + (1-\bar{\beta})^2] + \sigma^2(\tilde{\epsilon}) r_{Mt}}{[S^2(\beta) + (1-\bar{\beta})^2] + \sigma^2(\tilde{\epsilon})} \tag{23}$$

where $S^2(\beta)$ is the cross-sectional variance of the β_j and $\bar{\beta}$ is the mean. However, the average standard deviation of the measurement error $\sigma(\tilde{\epsilon}_j)$ for our portfolios is only 0.0101 (implying an average variance on the order of 0.0001), and since $S^2(\beta)$ for our ten portfolios is 0.1144 and $\bar{\beta} = 1.007$, this bias will be negligible and we shall ignore it.

To begin, let us apply the foregoing procedures to the excess return data to obtain an estimate of $\bar{R}_{Zt} = \bar{r}_{Zt} - r_{Ft}$, the excess return on the beta factor. Substituting R_{jt} for r_{jt} and R_{Mt} for r_{Mt} in (16), the $\hat{R}_{Zjt}$ were estimated for each of our ten

portfolios. These were then averaged to obtain the estimate

$$R_{Zt}^{\circ} = \sum_j h_j \hat{R}_{Zjt} = K' \sum_j (1-\beta_j)^2 \left[\frac{\bar{R}_{jt} - \hat{\beta}_j R_{Mt}}{1-\hat{\beta}_j}\right]$$

for each month t. The average of the R_{Zt}° for the entire period and for each of the four subperiods are given in Table 5, along with their t values. Table 5 also presents the serial correlation

TABLE 5

Estimated Mean Values and Serial Correlation of the Excess Returns on the Beta Factor over the Entire Periods and the Four Subperiods*

Period	$\bar{R}_Z^{\circ}$	$\sigma(R_Z^{\circ})$	$t(\bar{R}_Z^{\circ})$	$r(R_{Zt}^{\circ}, R_{Z,t-1}^{\circ})$	$t(r)$
1/31–12/65	0.00338	0.0426	1.62	0.113	2.33
1/31–9/39	−0.00849	0.0641	−1.35	0.194	1.49
10/39–6/48	0.00420	0.0455	0.946	0.208	2.19
7/48–3/57	0.00782	0.0199	4.03	−0.181	−1.87
4/57–12/65	0.00997	0.0228	4.49	0.414	4.60

*The values of $t(\bar{R}_Z^{\circ})$ were calculated under the assumption of normal distributions.

coefficients $r(R_{Zt}^{\circ}, R_{Z,t-1}^{\circ})$.[15] Note that the mean value $\bar{R}_Z^{\circ}$ of the beta factor over the whole period has a "t" value of only 1.64. However, as hypothesized earlier, it was negative in the first subperiod and positive and successively larger in each of the following subperiods. Moreover, in the last two subperiods its "t" values were 4.03 and 4.49, respectively. These results seem to us to be strong evidence favoring rejection of the traditional form of the asset pricing model which says that $\bar{R}_Z^{\circ}$ should be insignificantly different from zero.

In order to be sure that the significance levels reported in Table 5 are not spurious and due only to the misapplication of normal distribution theory to a situation in which the variables may actually be distributed according to the infinite variance members of the stable class of distributions. We have performed the significance tests using the stable distribution theory outlined by Fama and Roll [1968]. Table 6 presents the standardized variates (i.e., the "t" values) for $\bar{R}_Z^{\circ}$ for each of the sample periods given in Table 5 along with the "t" values at the 5% level of significance (two-tail) under

TABLE 6

Normalized Variate [i.e., t Value $t(\bar{R}_Z^\circ, \alpha) = \bar{R}_Z^\circ / \sigma(\bar{R}_Z^\circ\ \alpha)$] of the Excess Return on the Beta Factor Under the Assumption of Infinite Variance Symmetric Stable Distributions

	α					
Period	1.5	1.6	1.7	1.8	1.9	2.0
1/31–12/65	1.33	1.71	2.14	2.61	3.11*	3.65*
1/31–9/39	−1.11	−1.44	−1.71	−2.00	−2.29	−2.58
10/39–6/48	0.82	1.00	1.18	1.38	1.58	1.79
7/48–3/57	2.60	3.16	3.75*	4.37*	5.00*	5.66*
4/57–12/65	3.05	3.70	4.40*	5.11*	5.86*	6.63*
t Value at the 5% level of significance (two-tail)†	4.49	3.90	3.48.	3.16	2.93	2.77

Note: α = characteristic exponent, $\sigma(\bar{R}_Z^\circ, \alpha)$ = dispersion parameter of the distribution.

†Cf. Fama and Roll [1968].

alternative assumptions regarding the value of α, the characteristic exponent of the distribution. The smaller is α, the higher are the extreme tails of the probability distribution; $\alpha = 2$ corresponds to the normal distribution and $\alpha = 1$ to the Cauchy distribution. Evidence presented by Fama [1965] seems to indicate that α is probably in the range 1.7 to 1.9 for common stocks. We have not attempted to obtain explicit estimates of α for our data, since currently known estimation procedures are quite imprecise and require extremely large samples (up to 2,000 observations). Therefore we have simply presented the "t" values calculated according to the procedures suggested by Fama and Roll [1968] for six values of α ranging from 1.5 to 2.0. The coefficients in Table 6 that are significant at the 5% level are noted with an asterisk. Clearly, if α is greater than 1.7, the results confirm the impression gained from the normal tests given in Table 5.

Note that the estimates in Tables 5 and 6 were obtained from the excess return data; therefore, although the figures are of interest for testing the traditional form of the model, they do not give the appropriate level of the mean value of $\bar{r}_Z$. The estimates $\bar{r}_Z^\circ$ and $\bar{r}_M$ obtained from the total return data used in Figures 6 through 11 appear in Table 7, along with $\sigma(\bar{r}_Z^\circ)$ and $\sigma(\bar{r}_M)$ and the estimated values of γ_0 and γ_1 for the cross-sectional regressions [given by (10)] for each of the var-

TABLE 7

Mean and Standard Deviation of Returns on the Zero Beta and Market Portfolios and the Cross-sectional Regression Coefficients [from (10)] for Various Sample Periods

Time Period	$\bar{r}_Z^\circ$	$\bar{r}_M$	$\bar{r}_M - \bar{r}_Z^\circ$	$\sigma(r_Z^\circ)$	$\sigma(r_M)$	$\hat{\gamma}_0$[a]	$\hat{\gamma}_1$[a]
1931–1965	0.004980	0.015800	0.010820	0.042584	0.089054	0.005190	0.010807
1/31–9/39	−0.007393	0.023067	0.030459	0.063927	0.158707	−0.006913	0.030429
10/39–6/48	0.004833	0.015487	0.010665	0.045520	0.062414	0.005021	0.010652
7/48–3/57	0.009591	0.012915	0.003324	0.019895	0.036204	0.009537	0.003327
4/57–12/65	0.012889	0.011723	−0.001167	0.022631	0.038470	0.013115	−0.001181
1931	−0.047243	−0.037573	0.009669	0.040827	0.152924	−0.045492	0.009557
1932–1933	−0.009180	0.065574	0.074754	0.059741	0.245281	−0.008286	0.074696
1934–1935	0.015549	0.031250	0.015701	0.048551	0.097739	0.015542	0.015702
1936–1937	−0.007749	−0.004538	0.005211	0.032589	0.084786	−0.007336	0.003194
1938–1939	0.001919	0.024436	0.022517	0.100490	0.147129	0.001514	0.022543
1940–1941	−0.001308	−0.003902	−0.002596	0.043481	0.072454	−0.000646	−0.002638
1942–1943	−0.009898	0.035782	0.036780	0.066552	0.066451	−0.001069	0.036784
1944–1945	0.004511	0.036117	0.031507	0.032522	0.043560	0.004451	0.031517
1946–1947	0.010153	−0.002357	−0.013010	0.033074	0.056139	0.010946	−0.013061
1948–1949	0.009721	0.008529	−0.001192	0.019590	0.051471	0.009709	−0.001191
1950–1951	0.007163	0.020253	0.013090	0.028656	0.039764	0.007215	0.013087
1952–1953	0.012258	0.003054	−0.009204	0.014559	0.026896	0.012050	−0.009191
1954–1955	0.007432	0.027266	0.019834	0.019232	0.030804	0.007392	0.019836
1956–1957	0.010463	−0.003097	−0.013559	0.017638	0.032340	0.010555	−0.013565
1958–1959	0.014582	0.025060	0.011478	0.019982	0.028261	0.014205	0.011502
1960–1961	0.026825	0.010867	−0.015958	0.023178	0.036505	0.026753	−0.015953
1962–1963	0.004300	0.002728	−0.001571	0.026231	0.052144	0.005054	−0.001620
1964–1965	0.005032	0.017771	0.012738	0.014433	0.026761	0.005519	0.012707

[a]Cf. eq. (10).

ious sample periods portrayed in Figures 6 through 11. (Recall that the two-factor model implies $\gamma_0 = \bar{r}_Z$ and $\gamma_1 = \bar{r}_M - \bar{r}_Z$.) One additional item of interest in judging the im-portance of the beta factor in the determination of security returns is its standard deviation relative to that of the market returns. As Table 7 reveals, $\sigma(\bar{r}_Z^\circ)$ is roughly 50% as large as $\sigma(\bar{r}_M)$. Comparison of $\bar{r}_Z^\circ$ and $\bar{r}_M$ in Table 7 for the four 105-month subperiods indicates that the mean returns on the beta factor were approximately equal to the average market re-turns in the last two periods covering the interval July, 1948–December, 1965. Apparently, then, the relative magnitudes of $\bar{r}_Z^\circ$ and $\bar{r}_M$ indicate that the beta factor is economically as well as statistically significant.

V. *Conclusion*

The traditional form of the capital asset pricing model states that the expected excess return on a security is equal to its level of systematic risk, β, times the expected excess return on the market portfolio. That is, in capital market equilibrium, prices of assets adjust such that

$$E(\tilde{R}_j) = \gamma_1 \beta_j \tag{24}$$

where $\gamma_1 = E(\tilde{R}_M)$, the expected excess return on the market portfolio.

An alternative hypothesis of the pricing of capital assets arises from the relaxation of one of the assumptions of the tranditional form of the capital asset pricing model. Relaxa-tion of the assumption that riskless borrowing and lending opportunities are available leads to the formulation of the two-factor model. In equilibrium, the expected returns $E(\tilde{r}_j)$ on an asset will be given by

$$E(\tilde{r}_j) = E(\tilde{r}_Z) + [E(\tilde{r}_M) - E(\tilde{r}_Z)]\beta_j \tag{25}$$

where $E(\tilde{r}_Z)$ is the expected return on a portfolio that has a zero covariance (and thus $\beta_Z = 0$) with the return on the market portfolio $\tilde{r}_M$. In the context of this model, the return on 30-day Treasury Bills (which we have used as a proxy for a "riskless" rate) simply represents the return on a particular asset in the system. Thus, subtracting r_F from both sides of (25), we can rewrite (25) in terms of "excess" returns as

$$E(\tilde{R}_j) = \gamma_0 + \gamma_1 \beta_j \tag{26}$$

where $\gamma_0 = E(\tilde{R}_Z)$ and $\gamma_1 = E(\tilde{R}_M) - E(\tilde{R}_Z)$.

The traditional form of the asset pricing model implies that $\gamma_0 = 0$ and $\gamma_1 = E(\tilde{R}_M)$ and the two-factor model implies that $\gamma_0 = E(\tilde{R}_Z)$, which is not necessarily zero and that $\gamma_1 = E(\tilde{R}_M) - E(\tilde{R}_Z)$. In addition, several other models arise from relaxing some of the assumptions of the traditional asset pricing model which imply $\gamma_0 \neq 0$ and $\gamma_1 \neq E(R_M)$. These models involve explicit consideration of the problems of measuring R_M, the existence of nonmarketable assets, and the existence of differential taxes on capital gains and dividends, and we shall briefly outline them. Our main emphasis has been to test the strict traditional form of the asset pricing model; that is, is $\gamma_0 \neq 0$? We have made no attempt to provide direct tests of these other alternative hypotheses.

To test the traditional model, we used all securities listed on the New York Stock Exchange at any time in the interval between 1926 and 1966. The problem we faced was to obtain efficient estimates of the mean of the beta factor and its vari-ance. It would be possible to test the alternative hypotheses by selecting one security at random and estimating its beta from the time series and ascertaining whether its mean return was significantly different from that predicted by the tradi-tional form of the capital asset pricing model. However, this would be a very inefficient test procedure.

To gain efficiency, we grouped the securities into ten port-folios in such a way that the portfolios had a large spread in their β's. However, we knew that grouping the securities on the basis of their estimated β's would not give unbiased estimates of the portfolio "Beta," since the β's used to select the portfolios would contain measurement error. Such a pro-cedure would introduce a selection bias into the tests. To eliminate this bias we used an instrumental variable, the previous period's estimated beta, to select a security's port-folio grouping for the next year. Using these procedures, we constructed ten portfolios whose estimated β's were unbiased estimates of the portfolio "Beta." We found that much of the sampling variability of the β's estimated for individual securi-ties was eliminated by using the portfolio groupings. The β's of the portfolios constructed in this manner ranged from 0.49 to 1.5, and the estimates of the portfolio β's for the subperiods exhibited considerable stationarity.

The time series regressions of the portfolio excess returns on the market portfolio excess returns indicated that high-beta securities had significantly negative intercepts and low-beta securities had significantly positive intercepts, contrary to the predictions of the traditional form of the model. There was also considerable evidence that this effect became stronger through time, being strongest in the 1947–65 period. The cross-sectional plots of the mean excess returns on the portfolios against the estimated β's indicated that the relation between mean excess return and β was linear. However, the intercept and slope of the cross-sectional relation varied in different subperiods and were not consistent with the traditional form of the capital asset pricing model. In the two prewar 105-month subperiods examined, the slope was steeper in the first period than that predicted by the traditional form of the model, and it was flatter in the second period. In each of the two 105-month postwar periods it was considerably flatter than predicted. From the evidence of both the time series and cross-sectional runs, we were led to reject the hypothesis that γ_0 in (26) was equal to zero; we therefore concluded that the traditional form of the asset pricing model is not consistent with the data.

We also attempted to make explicit estimates of the time series of returns on the beta factor in order to obtain a more efficient estimate of its mean and variance and thereby enable ourselves to directly test whether or not the mean excess return on the beta factor was zero. We derived a minimum-variance, unbiased linear estimator of the returns on the β factor using our portfolio return data. We showed that, given the independence of the residuals the optimum estimator requires knowledge of the unobservable residual variances of each of the portfolios but that this problem could be avoided if they were equal. Under this assumption of equal residual variances, we estimated the time series of returns on the beta factor. However, if these assumptions (i.e., the independence of the residuals and equality of their variances) are not valid—and there is reason to believe they are not—more complicated procedures are necessary to obtain minimum-variance estimates. Such estimators, which use the complete covariance structure of the portfolio returns are available (although not derived here). However, we feel that a straightforward application of these procedures to the return data would result in the introduction of serious *ex post* bias in the estimates. Thus we have left a complete investigation of these problems, as well as more detailed tests of the two-factor model, to a future paper. In order to fully utilize the properties of the two-factor model in a number of applied problems (such as portfolio evaluation, see Jensen [1971] and various issues in valuation

theory), it will be necessary to have minimum-variance unbiased estimates of the time series of returns on the beta factor, and we hope to provide such estimates in the not-too-distant future.

The evidence obtained from the time series of returns on the beta factor indicated that the beta factor had a nonzero mean and that the mean was nonstationary over time. It seems to us that we have established the presence and significance of the beta factor in explaining security returns but, as mentioned earlier, we have not provided any direct tests aimed at explaining the existence of the beta factor. We have, however, suggested an economic rationale for why capital market equilibrium is consistent with the finding of this second factor. Black [1970] has shown that if riskless borrowing opportunities are not available, the equilibrium expected returns on an asset will be a linear function of two factors, one the β factor, the other the market factor.

In addition, Black and Jensen [1970] have demonstrated that if assets are omitted from the estimated market return, a model similar in some ways to the two-factor model would result. (Roll's analysis [1969] is relevant to this issue as well.) That is, it yields a model similar in structure to (26) and implies that $\gamma_0 \neq 0$. However, it is clear from Figures 6a and 6b and Table 7 that the beta factor (the intercept in the figures and γ_0 in Table 7) is highly variable and any alternative hypothesis must be consistent with this phenomenon. In other words, it is not sufficient for an alternative model to simply imply a nonzero but constant intercept in (26).

Others have provided alternative models that are similar in structure to the Black–Jensen results. For example, Mayers [1972] has developed an equilibrium model incorporating the existence of nonmarketable assets and has shown that the basic linear relation of the traditional model is unaltered, but the constant term γ_0 will be nonzero and γ_1 will not equal $E(R_M)$. The implications of his model for the structure of asset returns are virtually identical to those of the omitted assets model. Brennan [1970] has derived the equilibrium structure of security returns when the effects of a differential tax on dividends and capital gains are considered. He also concludes that the basic linearity of the traditional model is unchanged, but a nonzero constant term must be included and γ_1 will not equal $E(R_M)$. Black and Scholes [1970], however, have tested for the existence of dividend effects and have found that the differential tax on dividends and capital gains does not affect the structure of security returns and hence cannot explain the results reported here.

There are undoubtedly other economic hypotheses that are consistent with the findings of the existence of a second factor and consistent also with capital market equilibrium. Each hypothesis must be tested directly to determine whether it can account for the presence of the β factor. The Black–Scholes investigation of dividend effects is an example of such a test.

Appendix: The Grouping Solution to the Measurement Error Problem

Consider first the estimate $\hat{\beta}_j$ of the risk parameter in more detail. We will want to test (10) over some holding period, but we must first obtain the estimates of the risk parameter $\hat{\beta}_j$, from the time series equation given by (6). For simplicity, we shall assume that the $\tilde{e}_{jt}$ are independently distributed and have constant variance for all j and t. The least-squares estimate of β_j in (6), $\hat{\beta}_j$, is thus unbiased but subject to a sampling error $\tilde{\epsilon}_j$ as in (7), and the variance of the sampling error of the estimate $\hat{\beta}_j$ is

$$\text{var}\,(\hat{\beta}_j|\beta_j) = \sigma^2(\tilde{\epsilon}_j) = \frac{\sigma^2(\tilde{e}_j)}{\phi} = \frac{\sigma^2(\tilde{e})}{\phi} \tag{A.1}$$

since $\sigma^2(\tilde{e}_j)$ was assumed equal for all j, and where

$$\phi = \sum_{t=1}^{T} (R_{Mt} - R_M)^2 \tag{A.2}$$

is the sample sum of squared deviations of the independent variable over the T observations used in the time series estimating equation. Hence using (11) we see that

$$\text{plim}\,\hat{\gamma} = \frac{\gamma_1}{1 + \sigma^2(\tilde{e})/\phi S^2(\beta_j)} \tag{A.3}$$

Let us assume that we can order the firms on the basis of β_j or on the basis of some instrumental variable highly correlated with β_j but independent of $\tilde{\epsilon}_j$. Given the N ordered firms, we group them into M equal-size contiguous subgroups, represented by $K = 1, 2, \ldots, M$ and calculate the average return for each group for each month t according to

$$\tilde{R}_{Kt} = \frac{1}{L}\sum_{j=1}^{L} \tilde{R}_{Kjt} \qquad K = 1, 2, \ldots, M \tag{A.4}$$

$$L = \frac{N}{M} \quad \text{(assumed to be integer)} \tag{A.5}$$

where $\tilde{R}_{Kjt}$ is the return for month t for security j in group K. We then estimate the systematic risk of the group by applying least squares to

$$\tilde{R}_{Kt} = \alpha_K + \beta_K \tilde{R}_{Mt} + \tilde{e}_{Kt} \qquad \begin{cases} K = 1, 2, \ldots, M \\ t = 1, 2, \ldots, T \end{cases} \tag{A.6}$$

where

$$\tilde{e}_{Kt} = \frac{1}{L}\sum_{j=1}^{L} \tilde{e}_{Kjt} \tag{A.7}$$

and

$$\sigma^2(\tilde{e}_{Kt}) = \frac{\sigma^2(\tilde{e})}{L} \tag{A.8}$$

Equation (A.8) holds, since, by assumption, the $\tilde{e}_{Kjt}$ are independently distributed with equal variance. The least-squares estimate of β_K in (A.6) is $\hat{\beta}_K = \beta_K + \tilde{\epsilon}_K$ and its variance is

$$\text{var}\,(\hat{\beta}_K|\beta_K) = \sigma^2(\tilde{\epsilon}_K) = \frac{\sigma^2(\tilde{e})}{\phi L} \tag{A.9}$$

Now if we estimate the cross-sectional relation (10) using our M observations on $\bar{R}_K = \sum_{t=1}^{T} R_{Kt}/T$ and $\hat{\beta}_K$ for some holding period, we have

$$\bar{R}_K = \gamma_0 + \gamma_1 \hat{\beta}_K + \tilde{e}_K^{\circ} \tag{A.10}$$

where

$$\tilde{e}_K^{\circ} = \sum_{t=1}^{T} \frac{\tilde{e}_{Kt}^{\circ}}{T} = \tilde{e}_K - \gamma_1 \tilde{\epsilon}_K \tag{A.11}$$

Now the large sample estimate of γ_1 in (A.10)

$$\text{plim}\,\hat{\gamma}_1 = \frac{\gamma_1}{1 + \dfrac{\text{plim}\,\sigma^2(\tilde{\epsilon}_K)}{\text{plim}\,S^2(\beta_K)}} = \frac{\gamma_1}{1 + \dfrac{\text{plim}\,\frac{1}{L}\sigma^2(\tilde{e})}{\phi S^2(\beta_K)}} = \gamma_1 \tag{A.12}$$

since plim $\sigma^2(\tilde{e})/L = 0$ as long as $L \to \infty$ as $N \to \infty$, and this is true as long as we hold the number of groups constant. Thus these grouping procedures will result in unbiased estimates of the parameters of (10) for large samples. Note that $S^2(\beta_K)$, the cross-sectional sample variance of the true group risk coefficients, is constant with increasing L so long as securities are assigned to groups on the basis of the ranked $\hat{\beta}_j$. Note also, however, that if we randomly assigned securities to the M groups we would have plim $S^2(\beta_K) =$ plim $S^2(\beta_j)/L$ and (A.12) would thus be identical to (A.3). Therefore, random grouping would be of no help in eliminating the bias. As can be seen, the grouping procedures we have already described in the time series tests accomplish these results. While we expect these procedures to substantially reduce the bias[16] they cannot completely eliminate it in our case because the $\tilde{e}_j$ and therefore the $\tilde{\epsilon}_j$ are not independent across firms. However, as discussed in Section III, we expect the remaining bias to be trivially small.

Notes

1. Note that (4c) can be valid even though R_M is a weighted average of the R_j and therefore R'_M contains e_j. This may be clarified as follows: taking the weighted sum of (3) using the weights, X_j, of each security in the market portfolio we know by the definition of R_M that $\Sigma_j X_j R_j = R_M$, $\Sigma_j X_j \beta_j = 1$, and $\Sigma_j X_j e_j = 0$. Thus by the last equality we know $X_j e_j = -\Sigma_{i \neq j} X_i e_i$, and by substitution $E(e_j X_j e_j) = E[e_j(-\Sigma_{i \neq j} X_i e_i)] = X_j \sigma^2(e_j)$, and this implies condition (4c) since $E(e_j R'_m) = X_j \sigma^2(e_j) + E[e_j \Sigma_{i \neq j} X_i e_i] = 0$.
2. We could develop the model and tests under the assumption of infinite variance stable distributions, but this would unnecessarily complicate some of the analysis. We shall take explicit account of these distributional problems in some of the crucial tests of significance in Section IV.
3. Recall that the R_{it} and R_{Mt} are defined as excess returns. The model can be formulated with r_{Ft} omitted from (6) and therefore assumed constant (then $\alpha_j = r_F(1-\beta_j)$) or included as a variable (as we have done), which strictly requires them to be known for all t. But experiments with estimates obtained with the inclusion of r_{Ft} as a variable in (6) yield results virtually identical to those obtained with the assumption of constant r_F [and hence the exclusion of r_{Ft} as a variable in (6)], so we shall ignore this problem here. See also Roll [1969] and Miller and Scholes [1972] for a thorough discussion of the bias introduced through misspecification of the riskless rate. Miller and Scholes conclude as we do that these problems are not serious.
4. Unbiased measurement errors in $\hat{\beta}_j$ cause severe difficulties with the cross-sectional tests of the model, and it is important to note that the time series form of the tests given by (6) are free of this source of bias. Unbiased measurement errors in $\hat{\beta}_j$, which is estimated simultaneously with α_j in the time series formulation, cause errors in the estimate of α_j but no systematic bias. Measurement errors in R_{Mt} may cause difficulties in

both the cross-sectional and time series forms of the tests, but we shall ignore this issue here. For an analysis of the problems associated with measurement errors in R_{Mt}, see Black and Jensen [1970], Miller and Scholes [1972], and Roll [1969].
5. Treasury Bill rates were obtained from the Salomon Brothers & Hutzler quote sheets at the end of the previous month for the following month. Dealer commercial paper rates were obtained from Banking and Monetary Statistics, Board of Governors of the Federal Reserve System, Washington, D.C.
6. The choice of the number of portfolios is somewhat arbitrary. As we shall see below, we wanted enough portfolios to provide a continuum of observations across the risk spectrum to enable us to estimate the suspected relation between α_K and β_K.
7. Note that in order for the risk parameters of the groups β_K, to be stationary through time, our procedures require that firms leave and enter the sample symmetrically across the entire risk spectrum.
8. See also Miller and Scholes [1972], who provide a careful analysis (using procedures that are complementary to but much different from those suggested here) of many of these problems with cross-sectional tests and their implications for the interpretation of previous empirical work.
9. Intuitively one can see that the measurement error problem is virtually eliminated by these procedures because the errors in $\hat{\beta}_K$ become extremely small. Since the correlations $r(\tilde{R}_K, \tilde{R}_M)$ are so high in Table 2, the standard errors of estimate of the coefficients β_K are all less than 0.022, and nine of them are less than 0.012. The average standard error of estimate for the ten $\hat{\beta}_K$ coefficients given in Table 2 for the entire period was 0.0101 and the cross-sectional variance of the $\hat{\beta}_K$, $S^2(\hat{\beta}_K)$ was 0.1144. Hence, assuming $S^2(\hat{\beta}_K) = S^2(\beta_K)$, squaring 0.0101, and using (11), we see that our estimate of γ_1 will be greater than 99.9% of its true value.
10. The analysis was also performed where the coefficients were reestimated for each subperiod, and the results were very similar because the $\hat{\beta}_K$ were quite stable over time. We report these results since this estimation procedure seemed to result in a slightly larger spread of the $\hat{\beta}_K$ and since the increased sample sizes tends to further reduce the bias caused by the variance of the measurement error in $\hat{\beta}_K$.
11. In fact, there is an infinite number of such zero β portfolios. Of all such portfolios, however, r_Z is the return on the one with minimum variance. (We are indebted to John Long for the proof of this point.)
12. We say unreasonably high because the coefficients change from period to period by amounts ranging up to almost seven times their estimated standard errors.
13. Although the traditional form of the model is consistent with the existence of the β factor if its excess return had a zero mean, clearly it would not provide as complete an explanation of the structure of asset returns as a model that explicitly incorporated such a factor. In particular, under these circumstances the traditional form would provide an adequate description of security returns over fairly lengthy periods of time, say three years or more, but it would probably not furnish an adequate description of security returns over much shorter intervals.
14. We only observe the residual variance from the single variable regression, and, as we can see from (13), this will be equal to $(1-\beta_j)^2\sigma^2(\tilde{r}_Z) + \sigma^2(\tilde{w}_j)$. However, there are more general procedures for estimating $\tilde{r}_{Zt}$ in

the situation of nonidentical $\sigma^2(\tilde{u}_j)$ and $\text{cov}(\tilde{u}_j, \tilde{u}_i) = 0$ for $j \neq i$. But we leave an investigation of the properties of these estimates and some additional tests of the two-factor model for a future paper. If the assumption of identical $\sigma^2(\tilde{u}_j)$ made here is inappropriate, we still obtain an unbiased estimate of the $\bar{R}_z$. However, the estimated variance of $\bar{R}_z$, which is of some interest, will be greater than the true variance.

15. The serial correlation for the entire period appears significant. Indeed, the serial correlation in the last period, 0.414, seems very large and even highly significant, with a t value of 4.6. However, the coefficients in the earlier periods seem to border on significance but show an inordinately large amount of variability, thus indicating substantial nonstationarity.
16. As mentioned earlier, the choice of the number of groups is somewhat arbitrary and, for any given sample size, involves a tradeoff between the bias and the degree of sampling error in the estimates of the parameters in (10). In an unpublished study of the properties of the grouping procedures by simulation techniques, Jensen and Mendu Rao have found that, when $\sigma^2(\tilde{\epsilon}_j) = S^2(\beta_j)$, the use of ten groups with a total sample size of $N = 400$, yields estimates of the coefficient γ_1 in (10) which, on the average, are biased downward by less than 0.9% of their true value and have a standard error of estimate about 50% higher than that obtained with ungrouped data. The ungrouped sample estimates were, of course, 50% of their true values on the average [as implied by (11) for these assumed variances].

References

BLACK, FISCHER. "Capital Market Equilibrium With No Riskless Borrowing or Lending" (forthcoming in the *Journal of Business*).

———, and JENSEN, MICHAEL C. "Incomplete Measurement of Market Returns and Its Implications for Tests of the Asset Pricing Model" (unpublished manuscript, November, 1970).

———, and SCHOLES, MYRON. "Dividend Yields and Common Stock Returns: A New Methodology" (Cambridge: Sloan School of Management, Massachusetts Institute of Technology, Working Paper #488–70, September, 1970).

BLATTBERG, ROBERT, and SARGENT, THOMAS. "Regression with Paretian Disturbances: Some Sampling Results," *Econometrics*, V. 39 (May, 1971) 501–10.

BLUME, MARSHALL. "The Assessment of Portfolio Performance" (Ph.D. dissertation, University of Chicago, 1968).

DOUGLAS, GEORGE W. "Risk in the Equity Markets: An Empirical Appraisal of Market Efficiency," *Yale Economic Essays*, IX (Spring, 1969), 3–45.

FAMA, EUGENE. F. "Risk, Return, and Equilibrium: Some Clarifying Comments," *Journal of Finance* (March, 1968), 29–40.

———. "Risk, Return, and Equilibrium," *Journal of Political Economy*, LXXIX (January–February, 1971).

———, and BABIAK, HARVEY. "Dividend Policy: An Empirical Analysis," *Journal of the American Statistical Association*, LXIII (December,
———, FISHER, LAWRENCE; JENSEN, MICHAEL C.; and ROOL, RICHARD.
1968), 1132–61.
"The Adjustment of Stock Prices to New Information," *International Economic Review*, X (February, 1969), 1–26.

———, and ROLL, RICHARD. "Some Properties of Symmetric Stable Distributions," *Journal of the American Statistical Association*, LXIII (September, 1968), 817–36.

JENSEN, MICHAEL C. "The Performance of Mutual Funds in the Period 1945–64," *Journal of Finance*, XXIII (May, 1968), 389–416.

———. "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios," *Journal of Business*, XLII (April, 1969), 167–247.

———, "Optimal Utilization of Market Forecasts and The Evaluation of Investment Performance" (Working Paper No. 7109, University of Rochester School of Management, September, 1971).

KING, BENJAMIN F. "Market and Industry Factors in Stock Price Behavior," *Journal of Business*, XXXIX (January 1966, Part II), 134–90.

LINTNER, JOHN. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, XLVII (February, 1965a), 13–37.

———. "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, XX (December, 1965b), 587–616.

LONG, JOHN B., JR. "Consumption–Investment Decisions and Equilibrium in the Securities Market," this volume, 1972.

MARKOWITZ, HARRY M. Portfolio Selection: Efficient Diversification of Investments, Cowles Foundation Monograph No. 16 (New York: John Wiley & Sons, 1959).

MAYERS, DAVID. "Nonmarketable Assets and Capital Market Equilibrium Under Uncertainty," *this volume*, 1971.

MILLER, MERTON H., and SCHOLES, MYRON. "Rates of Return in Relation to Risk: A Re-examination of Some Recent Findings," this volume, 1972.

MOSSIN, JAN. "Equilibrium in a Capital Asset Market," *Econometrica*, XXXIV (October, 1966), 768–83.

PRESS, S. JAMES. "A Compound Events Model for Security Prices," *Journal of Business*, XL (July, 1967), 317–37.

ROLL, RICHARD. "Bias in Fitting the Sharpe Model to Time Series Data," *Journal of Financial and Quantitative Analysis*, IV (September, 1969), 271–89.

SHARPE, WILLIAM F. "A Simplified Model for Portfolio Analysis," *Management Science* (January, 1963), 277–93.

———. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, XIX (September, 1964), 425–42.

———. "Risk Aversion in the Stock Market," *Journal of Finance*, XX (September, 1965), 416–22.

———. "Mutual Fund Performance," *Journal of Business*, XXXIX, Part 2 (January, 1966), 119–38.

TREYNOR, JACK L. "Toward a Theory of Market Value of Risky Assets" (unpublished manuscript, 1961).

———. "How to Rate Management of Investment Funds," *Harvard Business Review*, XLIII (January–February, 1965), 63–75.

WALD, ABRAHAM. "The Fitting of Straight Lines if Both Variables are Subject to Error," *Annals of Mathematical Statistics*, II (1940), 284–300.

WISE, JOHN. "Linear Estimators for Linear Regression Systems Having Infinite Variances" (paper presented at the Berkeley–Stanford Mathematical Economics Seminar, October, 1963).

THE JOURNAL OF FINANCE • VOL XLVII, NO 2 • JUNE 1992

The Cross-Section of Expected Stock Returns

EUGENE F. FAMA and KENNETH R. FRENCH*

ABSTRACT

Two easily measured variables, size and book-to-market equity, combine to capture the cross-sectional variation in average stock returns associated with market β, size, leverage, book-to-market equity, and earnings-price ratios. Moreover, when the tests allow for variation in β that is unrelated to size, the relation between market β and average return is flat, even when β is the only explanatory variable.

THE ASSET-PRICING MODEL OF Sharpe (1964), Lintner (1965), and Black (1972) has long shaped the way academics and practitioners think about average returns and risk. The central prediction of the model is that the market portfolio of invested wealth is mean-variance efficient in the sense of Markowitz (1959). The efficiency of the market portfolio implies that (a) expected returns on securities are a positive linear function of their market βs (the slope in the regression of a security's return on the market's return), and (b) market βs suffice to describe the cross-section of expected returns.

There are several empirical contradictions of the Sharpe-Lintner-Black (SLB) model. The most prominent is the size effect of Banz (1981). He finds that market equity, ME (a stock's price times shares outstanding), adds to the explanation of the cross-section of average returns provided by market βs. Average returns on small (low ME) stocks are too high given their β estimates, and average returns on large stocks are too low.

Another contradiction of the SLB model is the positive relation between leverage and average return documented by Bhandari (1988). It is plausible that leverage is associated with risk and expected return, but in the SLB model, leverage risk should be captured by market β. Bhandari finds, however, that leverage helps explain the cross-section of average stock returns in tests that include size (ME) as well as β.

Stattman (1980) and Rosenberg, Reid, and Lanstein (1985) find that average returns on U.S. stocks are positively related to the ratio of a firm's book value of common equity, BE, to its market value, ME. Chan, Hamao, and Lakonishok (1991) find that book-to-market equity, BE/ME, also has a strong role in explaining the cross-section of average returns on Japanese stocks.

* Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637. We acknowledge the helpful comments of David Booth, Nai-fu Chen, George Constantinides, Wayne Ferson, Edward George, Campbell Harvey, Josef Lakonishok, Rex Sinquefield, René Stulz, Mark Zmijeweski, and an anonymous referee. This research is supported by the National Science Foundation (Fama) and the Center for Research in Security Prices (French).

Finally, Basu (1983) shows that earnings-price ratios (E/P) help explain the cross-section of average returns on U.S. stocks in tests that also include size and market β. Ball (1978) argues that E/P is a catch-all proxy for unnamed factors in expected returns; E/P is likely to be higher (prices are lower relative to earnings) for stocks with higher risks and expected returns, whatever the unnamed sources of risk.

Ball's proxy argument for E/P might also apply to size (ME), leverage, and book-to-market equity. All these variables can be regarded as different ways to scale stock prices, to extract the information in prices about risk and expected returns (Keim (1988)). Moreover, since E/P, ME, leverage, and BE/ME are all scaled versions of price, it is reasonable to expect that some of them are redundant for describing average returns. Our goal is to evaluate the joint roles of market β, size, E/P, leverage, and book-to-market equity in the cross-section of average returns on NYSE, AMEX, and NASDAQ stocks.

Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) find that, as predicted by the SLB model, there is a positive simple relation between average stock returns and β during the pre-1969 period. Like Reinganum (1981) and Lakonishok and Shapiro (1986), we find that the relation between β and average return disappears during the more recent 1963–1990 period, even when β is used alone to explain average returns. The appendix shows that the simple relation between β and average return is also weak in the 50-year 1941–1990 period. In short, our tests do not support the most basic prediction of the SLB model, that average stock returns are positively related to market βs.

Unlike the simple relation between β and average return, the univariate relations between average return and size, leverage, E/P, and book-to-market equity are strong. In multivariate tests, the negative relation between size and average return is robust to the inclusion of other variables. The positive relation between book-to-market equity and average return also persists in competition with other variables. Moreover, although the size effect has attracted more attention, book-to-market equity has a consistently stronger role in average returns. Our bottom-line results are: (a) β does not seem to help explain the cross-section of average stock returns, and (b) the combination of size and book-to-market equity seems to absorb the roles of leverage and E/P in average stock returns, at least during our 1963–1990 sample period.

If assets are priced rationally, our results suggest that stock risks are multidimensional. One dimension of risk is proxied by size, ME. Another dimension of risk is proxied by BE/ME, the ratio of the book value of common equity to its market value.

It is possible that the risk captured by BE/ME is the relative distress factor of Chan and Chen (1991). They postulate that the earning prospects of firms are associated with a risk factor in returns. Firms that the market judges to have poor prospects, signaled here by low stock prices and high ratios of book-to-market equity, have higher expected stock returns (they are penalized with higher costs of capital) than firms with strong prospects. It is also possible, however, that BE/ME just captures the unraveling (regression toward the mean) of irrational market whims about the prospects of firms.

Whatever the underlying economic causes, our main result is straightforward. Two easily measured variables, size (ME) and book-to-market equity (BE/ME), provide a simple and powerful characterization of the cross-section of average stock returns for the 1963–1990 period.

In the next section we discuss the data and our approach to estimating β. Section II examines the relations between average return and β and between average return and size. Section III examines the roles of E/P, leverage, and book-to-market equity in average returns. In sections IV and V, we summarize, interpret, and discuss applications of the results.

I. Preliminaries

A. Data

We use all nonfinancial firms in the intersection of (a) the NYSE, AMEX, and NASDAQ return files from the Center for Research in Security Prices (CRSP) and (b) the merged COMPUSTAT annual industrial files of income-statement and balance-sheet data, also maintained by CRSP. We exclude financial firms because the high leverage that is normal for these firms probably does not have the same meaning as for nonfinancial firms, where high leverage more likely indicates distress. The CRSP returns cover NYSE and AMEX stocks until 1973 when NASDAQ returns also come on line. The COMPUSTAT data are for 1962–1989. The 1962 start date reflects the fact that book value of common equity (COMPUSTAT item 60), is not generally available prior to 1962. More important, COMPUSTAT data for earlier years have a serious selection bias; the pre-1962 data are tilted toward big historically successful firms.

To ensure that the accounting variables are known before the returns they are used to explain, we match the accounting data for all fiscal yearends in calendar year $t - 1$ (1962–1989) with the returns for July of year t to June of $t + 1$. The 6-month (minimum) gap between fiscal yearend and the return tests is conservative. Earlier work (e.g., Basu (1983)) often assumes that accounting data are available within three months of fiscal yearends. Firms are indeed required to file their 10-K reports with the SEC within 90 days of their fiscal yearends, but on average 19.8% do not comply. In addition, more than 40% of the December fiscal yearend firms that do comply with the 90-day rule file on March 31, and their reports are not made public until April. (See Alford, Jones, and Zmijewski (1992).)

We use a firm's market equity at the end of December of year $t - 1$ to compute its book-to-market, leverage, and earnings-price ratios for $t - 1$, and we use its market equity for June of year t to measure its size. Thus, to be included in the return tests for July of year t, a firm must have a CRSP stock price for December of year $t - 1$ and June of year t. It must also have monthly returns for at least 24 of the 60 months preceding July of year t (for "pre-ranking" β estimates, discussed below). And the firm must have COMPUSTAT data on total book assets (A), book equity (BE), and earnings (E), for its fiscal year ending in (any month of) calendar year $t - 1$.

Our use of December market equity in the E/P, BE/ME, and leverage ratios is objectionable for firms that do not have December fiscal yearends because the accounting variable in the numerator of a ratio is not aligned with the market value in the denominator. Using ME at fiscal yearends is also problematic; then part of the cross-sectional variation of a ratio for a given year is due to market-wide variation in the ratio during the year. For example, if there is a general fall in stock prices during the year, ratios measured early in the year will tend to be lower than ratios measured later. We can report, however, that the use of fiscal-yearend MEs, rather than December MEs, in the accounting ratios has little impact on our return tests.

Finally, the tests mix firms with different fiscal yearends. Since we match accounting data for all fiscal yearends in calendar year $t - 1$ with returns for July of t to June of $t + 1$, the gap between the accounting data and the matching returns varies across firms. We have done the tests using the smaller sample of firms with December fiscal yearends with similar results.

B. Estimating Market βs

Our asset-pricing tests use the cross-sectional regression approach of Fama and MacBeth (1973). Each month the cross-section of returns on stocks is regressed on variables hypothesized to explain expected returns. The time-series means of the monthly regression slopes then provide standard tests of whether different explanatory variables are on average priced.

Since size, E/P, leverage, and BE/ME are measured precisely for individual stocks, there is no reason to smear the information in these variables by using portfolios in the Fama-MacBeth (FM) regressions. Most previous tests use portfolios because estimates of market βs are more precise for portfolios. Our approach is to estimate βs for portfolios and then assign a portfolio's β to each stock in the portfolio. This allows us to use individual stocks in the FM asset-pricing tests.

B.1. β Estimation: Details

In June of each year, all NYSE stocks on CRSP are sorted by size (ME) to determine the NYSE decile breakpoints for ME. NYSE, AMEX, and NASDAQ stocks that have the required CRSP-COMPUSTAT data are then allocated to 10 size portfolios based on the NYSE breakpoints. (If we used stocks from all three exchanges to determine the ME breakpoints, most portfolios would include only small stocks after 1973, when NASDAQ stocks are added to the sample.)

We form portfolios on size because of the evidence of Chan and Chen (1988) and others that size produces a wide spread of average returns and βs. Chan and Chen use only size portfolios. The problem this creates is that size and the βs of size portfolios are highly correlated (-0.988 in their data), so asset-pricing tests lack power to separate size from β effects in average returns.

To allow for variation in β that is unrelated to size, we subdivide each size decile into 10 portfolios on the basis of pre-ranking βs for individual stocks. The pre-ranking βs are estimated on 24 to 60 monthly returns (as available) in the 5 years before July of year t. We set the β breakpoints for each size decile using only NYSE stocks that satisfy our COMPUSTAT-CRSP data requirements for year $t-1$. Using NYSE stocks ensures that the β breakpoints are not dominated after 1973 by the many small stocks on NASDAQ. Setting β breakpoints with stocks that satisfy our COMPUSTAT-CRSP data requirements guarantees that there are firms in each of the 100 size-β portfolios.

After assigning firms to the size-β portfolios in June, we calculate the equal-weighted monthly returns on the portfolios for the next 12 months, from July to June. In the end, we have post-ranking monthly returns for July 1963 to December 1990 on 100 portfolios formed on size and pre-ranking βs. We then estimate βs using the full sample (330 months) of post-ranking returns on each of the 100 portfolios, with the CRSP value-weighted portfolio of NYSE, AMEX, and (after 1972) NASDAQ stocks used as the proxy for the market. We have also estimated βs using the value-weighted or the equal-weighted portfolio of NYSE stocks as the proxy for the market. These βs produce inferences on the role of β in average returns like those reported below.

We estimate β as the sum of the slopes in the regression of the return on a portfolio on the current and prior month's market return. (An additional lead and lag of the market have little effect on these sum βs.) The sum βs are meant to adjust for nonsynchronous trading (Dimson (1979)). Fowler and Rorke (1983) show that sum βs are biased when the market return is autocorrelated. The 1st- and 2nd-order autocorrelations of the monthly market returns for July 1963 to December 1990 are 0.06 and -0.05, both about 1 standard error from 0. If the Fowler-Rorke corrections are used, they lead to trivial changes in the βs. We stick with the simpler sum βs. Appendix Table AI shows that using sum βs produces large increases in the βs of the smallest ME portfolios and small declines in the βs of the largest ME portfolios.

Chan and Chen (1988) show that full-period β estimates for portfolios can work well in tests of the SLB model, even if the true βs of the portfolios vary through time, if the variation in the βs is proportional,

$$\beta_{jt} - \beta_j = k_t(\beta_j - \beta), \tag{1}$$

where β_{jt} is the true β for portfolio j at time t, β_j is the mean of β_{jt} across t, and β is the mean of the β_j. The Appendix argues that (1) is a good approximation for the variation through time in the true βs of portfolios (j) formed on size and β. For diehard β fans, sure to be skeptical of our results on the weak role of β in average stock returns, we can also report that the results stand up to robustness checks that use 5-year pre-ranking βs, or 5-year post-ranking βs, instead of the full-period post-ranking βs.

We allocate the full-period post-ranking β of a size-β portfolio to each stock in the portfolio. These are the βs that will be used in the Fama-MacBeth cross-sectional regressions for individual stocks. We judge that the precision of the full-period post-ranking portfolio βs, relative to the imprecise β estimates that would be obtained for individual stocks, more than makes up for the fact that true βs are not the same for all stocks in a portfolio. And note that assigning full-period portfolio βs to stocks does not mean that a stock's β is constant. A stock can move across portfolios with year-to-year changes in the stock's size (ME) and in the estimates of its β for the preceding 5 years.

B.2. β Estimates

Table I shows that forming portfolios on size and pre-ranking βs, rather than on size alone, magnifies the range of full-period post-ranking βs. Sorted on size alone, the post-ranking βs range from 1.44 for the smallest ME portfolio to 0.92 for the largest. This spread of βs across the 10 size deciles is smaller than the spread of post-ranking βs produced by the β sort of *any* size decile. For example, the post-ranking βs for the 10 portfolios in the smallest size decile range from 1.05 to 1.79. Across all 100 size-β portfolios, the post-ranking βs range from 0.53 to 1.79, a spread 2.4 times the spread, 0.52, obtained with size portfolios alone.

Two other facts about the βs are important. First, in each size decile the post-ranking βs closely reproduce the ordering of the pre-ranking βs. We take this to be evidence that the pre-ranking β sort captures the ordering of true post-ranking βs. (The appendix gives more evidence on this important issue.) Second, the β sort is not a refined size sort. In any size decile, the average values of ln(ME) are similar across the β-sorted portfolios. Thus the pre-ranking β sort achieves its goal. It produces strong variation in post-ranking βs that is unrelated to size. This is important in allowing our tests to distinguish between β and size effects in average returns.

II. β and Size

The Sharpe-Lintner-Black (SLB) model plays an important role in the way academics and practitioners think about risk and the relation between risk and expected return. We show next that when common stock portfolios are formed on size alone, there seems to be evidence for the model's central prediction: average return is positively related to β. The βs of size portfolios are, however, almost perfectly correlated with size, so tests on size portfolios are unable to disentangle β and size effects in average returns. Allowing for variation in β that is unrelated to size breaks the logjam, but at the expense of β. Thus, when we subdivide size portfolios on the basis of pre-ranking βs, we find a strong relation between average return and size, but no relation between average return and β.

A. Informal Tests

Table II shows post-ranking average returns for July 1963 to December 1990 for portfolios formed from one-dimensional sorts of stocks on size or β. The portfolios are formed at the end of June each year and their equal-weighted returns are calculated for the next 12 months. We use returns for July to June to match the returns in later tests that use the accounting data. When we sort on just size or 5-year pre-ranking βs, we form 12 portfolios. The middle 8 cover deciles of size or β. The 4 extreme portfolios (1A, 1B, 10A, and 10B) split the bottom and top deciles in half.

Table II shows that when portfolios are formed on size alone, we observe the familiar strong negative relation between size and average return (Banz (1981)), and a strong positive relation between average return and β. Average returns fall from 1.64% per month for the smallest ME portfolio to 0.90% for the largest. Post-ranking βs also decline across the 12 size portfolios, from 1.44 for portfolio 1A to 0.90 for portfolio 10B. Thus, a simple size sort seems to support the SLB prediction of a positive relation between β and average return. But the evidence is muddied by the tight relation between size and the βs of size portfolios.

The portfolios formed on the basis of the ranked market βs of stocks in Table II produce a wider range of βs (from 0.81 for portfolio 1A to 1.73 for 10B) than the portfolios formed on size. Unlike the size portfolios, the β-sorted portfolios do not support the SLB model. There is little spread in average returns across the β portfolios, and there is no obvious relation between β and average returns. For example, although the two extreme portfolios, 1A and 10B, have much different βs, they have nearly identical average returns (1.20% and 1.18% per month). These results for 1963–1990 confirm Reinganum's (1981) evidence that for β-sorted portfolios, there is no relation between average return and β during the 1964–1979 period.

The 100 portfolios formed on size and then pre-ranking β in Table I clarify the contradictory evidence on the relation between β and average return produced by portfolios formed on size or β alone. Specifically, the two-pass sort gives a clearer picture of the separate roles of size and β in average returns. Contrary to the central prediction of the SLB model, the second-pass β sort produces little variation in average returns. Although the post-ranking βs in Table I increase strongly in each size decile, average returns are flat or show a slight tendency to decline. In contrast, within the columns of the average return and β matrices of Table I, average returns and βs decrease with increasing size.

The two-pass sort on size and β in Table I says that variation in β that is tied to size is positively related to average return, but variation in β unrelated to size is not compensated in the average returns of 1963–1990. The proper inference seems to be that there is a relation between size and average return, but controlling for size, there is no relation between β and average return. The regressions that follow confirm this conclusion, and they produce another that is stronger. The regressions show that when one allows

Table I

Average Returns, Post-Ranking βs and Average Size For Portfolios Formed on Size and then β: Stocks Sorted on ME (Down) then Pre-Ranking β (Across): July 1963 to December 1990

Portfolios are formed yearly. The breakpoints for the size (ME, price times shares outstanding) deciles are determined in June of year t (t = 1963–1990) using all NYSE stocks on CRSP. All NYSE, AMEX, and NASDAQ stocks that meet the CRSP-COMPUSTAT data requirements are allocated to the 10 size portfolios using the NYSE breakpoints. Each size decile is subdivided into 10 β portfolios using pre-ranking βs of individual stocks, estimated with 2 to 5 years of monthly returns (as available) ending in June of year t. We use only NYSE stocks that meet the CRSP-COMPUSTAT data requirements to establish the β breakpoints. The equal-weighted monthly returns on the resulting 100 portfolios are then calculated for July of year t to June of year $t + 1$.

The post-ranking βs use the full (July 1963 to December 1990) sample of post-ranking returns for each portfolio. The pre- and post-ranking βs (here and in all other tables) are the sum of the slopes from a regression of monthly returns on the current and prior month's returns on the value-weighted portfolio of NYSE, AMEX, and (after 1972) NASDAQ stocks. The average return is the time-series average of the monthly equal-weighted portfolio returns, in percent. The average size of a portfolio is the time-series average of monthly averages of ln(ME) for stocks in the portfolio at the end of June of each year, with ME denominated in millions of dollars

The average number of stocks per month for the size-β portfolios in the smallest size decile varies from 70 to 177. The average number of stocks for the size-β portfolios in size deciles 2 and 3 is between 15 and 41, and the average number for the largest 7 size deciles is between 11 and 22.

The All column shows statistics for equal-weighted size-decile (ME) portfolios. The All row shows statistics for equal-weighted portfolios of the stocks in each β group.

	All	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 (in Percent)											
All	1.25	1.34	1.29	1.36	1.31	1.33	1.28	1.24	1.21	1.25	1.14
Small-ME	1.52	1.71	1.57	1.79	1.61	1 50	1.50	1.37	1.63	1.50	1.42
ME-2	1.29	1.25	1.42	1.36	1.39	1.65	1.61	1.37	1 31	1.34	1.11
ME-3	1.24	1.12	1.31	1.17	1.70	1.29	1.10	1.31	1.36	1.26	0.76
ME-4	1.25	1.27	1.13	1.54	1.06	1.34	1.06	1.41	1.17	1.35	0.98
ME-5	1.29	1.34	1.42	1.39	1.48	1.42	1.18	1.13	1.27	1.18	1.08
ME-6	1.17	1.08	1.53	1.27	1.15	1.20	1.21	1.18	1.04	1.07	1.02
ME-7	1.07	0.95	1.21	1.26	1.09	1.18	1.11	1.24	0.62	1.32	0.76
ME-8	1.10	1.09	1.05	1.37	1.20	1.27	0.98	1.18	1.02	1.01	0.94
ME-9	0.95	0.98	0.88	1.02	1.14	1.07	1.23	0.94	0.82	0.88	0.59
Large-ME	0.89	1.01	0.93	1.10	0.94	0.93	0.89	1.03	0.71	0.74	0.56

Table I—*Continued*

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
					Panel B: Post-Ranking βs						
All		0.87	0.99	1.09	1.16	1.26	1.29	1.35	1.45	1.52	1.72
Small-ME	1.44	1.05	1.18	1.28	1.32	1.40	1.40	1.49	1.61	1.64	1.79
ME-2	1.39	0.91	1.15	1.17	1.24	1.36	1.41	1.43	1.50	1.66	1.76
ME-3	1.35	0.97	1.13	1.13	1.21	1.26	1.28	1.39	1.50	1.51	1.75
ME-4	1.34	0.78	1.03	1.17	1.16	1.29	1.37	1.46	1.51	1.64	1.71
ME-5	1.25	0.66	0.85	1.12	1.15	1.16	1.26	1.30	1.43	1.59	1.68
ME-6	1.23	0.61	0.78	1.05	1.16	1.22	1.28	1.36	1.46	1.49	1.70
ME-7	1.17	0.57	0.92	1.01	1.11	1.14	1.26	1.24	1.39	1.34	1.60
ME-8	1.09	0.53	0.74	0.94	1.02	1.13	1.12	1.18	1.26	1.35	1.52
ME-9	1.03	0.58	0.74	0.80	0.95	1.06	1.15	1.14	1.21	1.22	1.42
Large-ME	0.92	0.57	0.71	0.78	0.89	0.95	0.92	1.02	1.01	1.11	1.32
					Panel C: Average Size (ln(ME))						
All	4.11	3.86	4.26	4.33	4.41	4.27	4.32	4.26	4.19	4.03	3.77
Small-ME	2.24	2.12	2.27	2.30	2.30	2.28	2.29	2.30	2.32	2.25	2.15
ME-2	3.63	3.65	3.68	3.70	3.72	3.69	3.70	3.69	3.69	3.70	3.68
ME-3	4.10	4.14	4.18	4.12	4.15	4.16	4.16	4.18	4.14	4.15	4.15
ME-4	4.50	4.53	4.53	4.57	4.54	4.56	4.55	4.52	4.58	4.52	4.56
ME-5	4.89	4.91	4.91	4.93	4.95	4.93	4.92	4.93	4.92	4.92	4.95
ME-6	5.30	5.30	5.33	5.34	5.34	5.33	5.33	5.33	5.33	5.34	5.36
ME-7	5.73	5.73	5.75	5.77	5.76	5.73	5.77	5.77	5.76	5.72	5.76
ME-8	6.24	6.26	6.27	6.26	6.24	6.24	6.27	6.24	6.24	6.24	6.26
ME-9	6.82	6.82	6.84	6.82	6.82	6.81	6.81	6.81	6.81	6.80	6.83
Large-ME	7.93	7.94	8.04	8.10	8.04	8.02	8.02	7.94	7.80	7.75	7.62

Table II

Properties of Portfolios Formed on Size or Pre-Ranking β: July 1963 to December 1990

At the end of June of each year t, 12 portfolios are formed on the basis of ranked values of size (ME) or pre-ranking β. The pre-ranking βs use 2 to 5 years (as available) of monthly returns ending in June of t. Portfolios 2–9 cover deciles of the ranking variables. The bottom and top 2 portfolios (1A, 1B, 10A, and 10B) split the bottom and top deciles in half. The breakpoints for the ME portfolios are based on ranked values of ME for all NYSE stocks on CRSP. NYSE breakpoints for pre-ranking βs are also used to form the β portfolios. NYSE, AMEX, and NASDAQ stocks are then allocated to the size or β portfolios using the NYSE breakpoints. We calculate each portfolio's monthly equal-weighted return for July of year t to June of year $t+1$, and then reform the portfolios in June of $t+1$.

BE is the book value of common equity plus balance-sheet deferred taxes, A is total book assets, and E is earnings (income before extraordinary items, plus income-statement deferred taxes, minus preferred dividends). BE, A, and E are for each firm's latest fiscal year ending in calendar year $t-1$. The accounting ratios are measured using market equity ME in December of year $t-1$. Firm size ln(ME) is measured in June of year t, with ME denominated in millions of dollars.

The average return is the time-series average of the monthly equal-weighted portfolio returns, in percent. ln(ME), ln(BE/ME), ln(A/ME), ln(A/BE), E/P, and E/P dummy are the time-series averages of the monthly average values of these variables in each portfolio. Since the E/P dummy is 0 when earnings are positive, and 1 when earnings are negative, E/P dummy gives the average proportion of stocks with negative earnings in each portfolio.

β is the time-series average of the monthly portfolio βs. Stocks are assigned the post-ranking β of the size-β portfolio they are in at the end of June of year t (Table I). These individual-firm βs are averaged to compute the monthly βs for each portfolio for July of year t to June of year $t+1$.

Firms is the average number of stocks in the portfolio each month.

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Panel A: Portfolios Formed on Size												
Return	1.64	1.16	1.29	1.24	1.25	1.29	1.17	1.07	1.10	0.95	0.88	0.90
β	1.44	1.44	1.39	1.34	1.33	1.24	1.22	1.16	1.08	1.02	0.95	0.90
ln(ME)	1.98	3.18	3.63	4.10	4.50	4.89	5.30	5.73	6.24	6.82	7.39	8.44
ln(BE/ME)	−0.01	−0.21	−0.23	−0.26	−0.32	−0.36	−0.36	−0.44	−0.40	−0.42	−0.51	−0.65
ln(A/ME)	0.73	0.50	0.46	0.43	0.37	0.32	0.32	0.24	0.29	0.27	0.17	−0.03
ln(A/BE)	0.75	0.71	0.69	0.69	0.68	0.67	0.68	0.67	0.69	0.70	0.68	0.62
E/P dummy	0.26	0.14	0.11	0.09	0.06	0.04	0.04	0.03	0.03	0.02	0.02	0.01
E(+)/P	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0 10	0.10	0.09	0.09
Firms	772	189	236	170	144	140	128	125	119	114	60	64

Table II—*Continued*

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
	Panel B: Portfolios Formed on Pre-Ranking β											
Return	1.20	1 20	1.32	1.26	1.31	1.30	1.30	1.23	1.23	1.33	1.34	1.18
β	0.81	0.79	0.92	1.04	1.13	1.19	1.26	1.32	1.41	1.52	1 63	1.73
ln(ME)	4.21	4.86	4.75	4.68	4.59	4.48	4.36	4.25	3.97	3.78	3.52	3.15
ln(BE/ME)	−0.18	−0.13	−0.22	−0.21	−0.23	−0.22	−0.22	−0.25	−0.23	−0.27	−0.31	−0.50
ln(A/ME)	0.60	0.66	0.49	0.45	0.42	0.42	0.45	0.42	0.47	0 46	0.46	0.31
ln(A/BE)	0.78	0.79	0.71	0.66	0.64	0.65	0.67	0.67	0.70	0.73	0.77	0 81
E/P dummy	0.12	0.06	0.09	0.09	0.08	0.09	0.10	0.12	0.12	0.14	0.17	0.23
E(+)/P	0.11	0.12	0.10	0.10	0.10	0.10	0.10	0.09	0.10	0.09	0 09	0.08
Firms	116	80	185	181	179	182	185	205	227	267	165	291

110138-OPC-POD-60-51

for variation in β that is unrelated to size, the relation between β and average return is flat, even when β is the only explanatory variable.

B. Fama-MacBeth Regressions

Table III shows time-series averages of the slopes from the month-by-month Fama-MacBeth (FM) regressions of the cross-section of stock returns on size, β, and the other variables (leverage, E/P, and book-to-market equity) used to explain average returns. The average slopes provide standard FM tests for determining which explanatory variables on average have non-zero expected premiums during the July 1963 to December 1990 period.

Like the average returns in Tables I and II, the regressions in Table III say that size, ln(ME), helps explain the cross-section of average stock returns. The average slope from the monthly regressions of returns on size alone is -0.15%, with a t-statistic of -2.58. This reliable negative relation persists no matter which other explanatory variables are in the regressions; the average slopes on ln(ME) are always close to or more than 2 standard errors from 0. The size effect (smaller stocks have higher average returns) is thus robust in the 1963–1990 returns on NYSE, AMEX, and NASDAQ stocks.

In contrast to the consistent explanatory power of size, the FM regressions show that market β does not help explain average stock returns for 1963–1990. In a shot straight at the heart of the SLB model, the average slope from the regressions of returns on β alone in Table III is 0.15% per month and only 0.46 standard errors from 0. In the regressions of returns on size and β, size has explanatory power (an average slope -3.41 standard errors from 0), but the average slope for β is negative and only 1.21 standard errors from 0. Lakonishok and Shapiro (1986) get similar results for NYSE stocks for 1962–1981. We can also report that β shows no power to explain average returns (the average slopes are typically less than 1 standard error from 0) in FM regressions that use various combinations of β with size, book-to-market equity, leverage, and E/P.

C. Can β Be Saved?

What explains the poor results for β? One possibility is that other explanatory variables are correlated with true βs, and this obscures the relation between average returns and measured βs. But this line of attack cannot explain why β has no power when used alone to explain average returns. Moreover, leverage, book-to-market equity, and E/P do not seem to be good proxies for β. The averages of the monthly cross-sectional correlations between β and the values of these variables for individual stocks are all within 0.15 of 0.

Another hypothesis is that, as predicted by the SLB model, there is a positive relation between β and average return, but the relation is obscured by noise in the β estimates. However, our full-period post-ranking βs do not seem to be imprecise. Most of the standard errors of the βs (not shown) are

Table III

Average Slopes (t-Statistics) from Month-by-Month Regressions of Stock Returns on β, Size, Book-to-Market Equity, Leverage, and E/P: July 1963 to December 1990

Stocks are assigned the post-ranking β of the size-β portfolio they are in at the end of June of year t (Table I). BE is the book value of common equity plus balance-sheet deferred taxes, A is total book assets, and E is earnings (income before extraordinary items, plus income-statement deferred taxes, minus preferred dividends) BE, A, and E are for each firm's latest fiscal year ending in calendar year $t - 1$. The accounting ratios are measured using market equity ME in December of year $t - 1$. Firm size ln(ME) is measured in June of year t. In the regressions, these values of the explanatory variables for individual stocks are matched with CRSP returns for the months from July of year t to June of year $t + 1$. The gap between the accounting data and the returns ensures that the accounting data are available prior to the returns. If earnings are positive, E(+)/P is the ratio of total earnings to market equity and E/P dummy is 0. If earnings are negative, E(+)/P is 0 and E/P dummy is 1.

The average slope is the time-series average of the monthly regression slopes for July 1963 to December 1990, and the t-statistic is the average slope divided by its time-series standard error.

On average, there are 2267 stocks in the monthly regressions. To avoid giving extreme observations heavy weight in the regressions, the smallest and largest 0.5% of the observations on E(+)/P, BE/ME, A/ME, and A/BE are set equal to the next largest or smallest values of the ratios (the 0.005 and 0.995 fractiles). This has no effect on inferences.

β	ln(ME)	ln(BE/ME)	ln(A/ME)	ln(A/BE)	E/P Dummy	E(+)/P
0.15						
(0.46)						
	−0.15					
	(−2 58)					
−0.37	−0.17					
(−1.21)	(−3.41)					
		0.50				
		(5.71)				
			0.50	−0.57		
			(5 69)	(−5.34)		
					0.57	4.72
					(2.28)	(4.57)
	−0.11	0.35				
	(−1.99)	(4.44)				
	−0 11		0.35	−0 50		
	(−2 06)		(4.32)	(−4.56)		
	−0.16				0.06	2.99
	(−3.06)				(0 38)	(3.04)
	−0.13	0.33			−0.14	0.87
	(−2.47)	(4.46)			(−0.90)	(1.23)
	−0.13		0.32	−0.46	−0.08	1.15
	(−2.47)		(4.28)	(−4.45)	(−0.56)	(1.57)

0.05 or less, only 1 is greater than 0.1, and the standard errors are small relative to the range of the βs (0.53 to 1.79).

The β-sorted portfolios in Tables I and II also provide strong evidence against the β-measurement-error story. When portfolios are formed on pre-ranking βs alone (Table II), the post-ranking βs for the portfolios almost perfectly reproduce the ordering of the pre-ranking βs. Only the β for portfolio 1B is out of line, and only by 0.02. Similarly, when portfolios are formed on size and then pre-ranking βs (Table I), the post-ranking βs in each size decile closely reproduce the ordering of the pre-ranking βs.

The correspondence between the ordering of the pre-ranking and post-ranking βs for the β-sorted portfolios in Tables I and II is evidence that the post-ranking βs are informative about the ordering of the true βs. The problem for the SLB model is that there is no similar ordering in the average returns on the β-sorted portfolios. Whether one looks at portfolios sorted on β alone (Table II) or on size and then β (Table I), average returns are flat (Table II) or decline slightly (Table I) as the post-ranking βs increase.

Our evidence on the robustness of the size effect and the absence of a relation between β and average return is so contrary to the SLB model that it behooves us to examine whether the results are special to 1963–1990. The appendix shows that NYSE returns for 1941–1990 behave like the NYSE, AMEX, and NASDAQ returns for 1963–1990; there is a reliable size effect over the full 50-year period, but little relation between β and average return. Interestingly, there is a reliable simple relation between β and average return during the 1941–1965 period. These 25 years are a major part of the samples in the early studies of the SLB model of Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973). Even for the 1941–1965 period, however, the relation between β and average return disappears when we control for size.

III. Book-to-Market Equity, E/P, and Leverage

Tables I to III say that there is a strong relation between the average returns on stocks and size, but there is no reliable relation between average returns and β. In this section we show that there is also a strong cross-sectional relation between average returns and book-to-market equity. If anything, this book-to-market effect is more powerful than the size effect. We also find that the combination of size and book-to-market equity absorbs the apparent roles of leverage and E/P in average stock returns.

A. Average Returns

Table IV shows average returns for July 1963 to December 1990 for portfolios formed on ranked values of book-to-market equity (BE/ME) or earnings-price ratio (E/P). The BE/ME and E/P portfolios in Table IV are formed in the same general way (one-dimensional yearly sorts) as the size and β portfolios in Table II. (See the tables for details.)

The relation between average return and E/P has a familiar U-shape (e.g., Jaffe, Keim, and Westerfield (1989) for U.S. data, and Chan, Hamao, and Lakonishok (1991) for Japan). Average returns decline from 1.46% per month for the negative E/P portfolio to 0.93% for the firms in portfolio 1B that have low but positive E/P. Average returns then increase monotonically, reaching 1.72% per month for the highest E/P portfolio.

The more striking evidence in Table IV is the strong positive relation between average return and book-to-market equity. Average returns rise from 0.30% for the lowest BE/ME portfolio to 1.83% for the highest, a difference of 1.53% per month. This spread is twice as large as the difference of 0.74% between the average monthly returns on the smallest and largest size portfolios in Table II. Note also that the strong relation between book-to-market equity and average return is unlikely to be a β effect in disguise; Table IV shows that post-ranking market βs vary little across portfolios formed on ranked values of BE/ME.

On average, only about 50 (out of 2317) firms per year have negative book equity, BE. The negative BE firms are mostly concentrated in the last 14 years of the sample, 1976–1989, and we do not include them in the tests. We can report, however, that average returns for negative BE firms are high, like the average returns of high BE/ME firms. Negative BE (which results from persistently negative earnings) and high BE/ME (which typically means that stock prices have fallen) are both signals of poor earning prospects. The similar average returns of negative and high BE/ME firms are thus consistent with the hypothesis that book-to-market equity captures cross-sectional variation in average returns that is related to relative distress.

B. Fama-MacBeth Regressions

B.1. BE/ME

The FM regressions in Table III confirm the importance of book-to-market equity in explaining the cross-section of average stock returns. The average slope from the monthly regressions of returns on ln(BE/ME) alone is 0.50%, with a t-statistic of 5.71. This book-to-market relation is stronger than the size effect, which produces a t-statistic of -2.58 in the regressions of returns on ln(ME) alone. But book-to-market equity does not replace size in explaining average returns. When both ln(ME) and ln(BE/ME) are included in the regressions, the average size slope is still -1.99 standard errors from 0; the book-to-market slope is an impressive 4.44 standard errors from 0.

B.2. Leverage

The FM regressions that explain returns with leverage variables provide interesting insight into the relation between book-to-market equity and average return. We use two leverage variables, the ratio of book assets to market equity, A/ME, and the ratio of book assets to book equity, A/BE. We interpret A/ME as a measure of market leverage, while A/BE is a measure

Table IV

Properties of Portfolios Formed on Book-to-Market Equity (BE/ME) and Earnings-Price Ratio (E/P): July 1963 to December 1990

At the end of each year $t-1$, 12 portfolios are formed on the basis of ranked values of BE/ME or E/P. Portfolios 2–9 cover deciles of the ranking variables. The bottom and top 2 portfolios (1A, 1B, 10A, and 10B) split the bottom and top deciles in half. For E/P, there are 13 portfolios; portfolio 0 is stocks with negative E/P. Since BE/ME and E/P are not strongly related to exchange listing, their portfolio breakpoints are determined on the basis of the ranked values of the variables for all stocks that satisfy the CRSP-COMPUSTAT data requirements. BE is the book value of common equity plus balance-sheet deferred taxes, A is total book assets, and E is earnings (income before extraordinary items, plus income-statement deferred taxes, minus preferred dividends). BE, A, and E are for each firm's latest fiscal year ending in calendar year $t-1$. The accounting ratios are measured using market equity ME in December of year $t-1$. Firm size ln(ME) is measured in June of year t, with ME denominated in millions of dollars. We calculate each portfolio's monthly equal-weighted return for July of year t to June of year $t+1$, and then reform the portfolios at the end of year t.

Return is the time-series average of the monthly equal-weighted portfolio returns (in percent). ln(ME), ln(BE/ME), ln(A/ME), ln(A/BE), E(+)/P, and E/P dummy are the time-series averages of the monthly average values of these variables in each portfolio Since the E/P dummy is 0 when earnings are positive, and 1 when earnings are negative, E/P dummy gives the average proportion of stocks with negative earnings in each portfolio.

β is the time-series average of the monthly portfolio βs. Stocks are assigned the post-ranking β of the size-β portfolio they are in at the end of June of year t (Table I). These individual-firm βs are averaged to compute the monthly βs for each portfolio for July of year t to June of year $t+1$. Firms is the average number of stocks in the portfolio each month.

Portfolio	0	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Panel A: Stocks Sorted on Book-to-Market Equity (BE/ME)													
Return		0.30	0.67	0.87	0.97	1.04	1.17	1.30	1.44	1.50	1.59	1.92	1.83
β		1.36	1.34	1.32	1.30	1.28	1.27	1.27	1.27	1.27	1.29	1.33	1.35
ln(ME)		4.53	4.67	4.69	4.56	4.47	4.38	4.23	4.06	3.85	3.51	3.06	2.65
ln(BE/ME)		−2.22	−1.51	−1.09	−0.75	−0.51	−0 32	−0.14	0.03	0.21	0.42	0.66	1.02
ln(A/ME)		−1.24	−0.79	−0.40	−0.05	0.20	0.40	0.56	0.71	0.91	1.12	1.35	1.75
ln(A/BE)		0.94	0.71	0.68	0.70	0.71	0.71	0.70	0.68	0.70	0.70	0.70	0.73
E/P dummy		0.29	0.15	0.10	0.08	0.08	0.08	0.09	0.09	0.11	0.15	0.22	0.36
E(+)/P		0.03	0.04	0.06	0.08	0.09	0.10	0.11	0.11	0.12	0.12	0.11	0.10
Firms		89	98	209	222	226	230	235	237	239	239	120	117

Table IV—*Continued*

Portfolio	0	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Panel B: Stocks Sorted on Earnings-Price Ratio (E/P)													
Return	1.46	1.04	0.93	0.94	1.03	1.18	1.22	1.33	1.42	1.46	1.57	1.74	1.72
β	1.47	1.40	1.35	1.31	1.28	1.26	1.25	1.26	1.24	1.23	1.24	1.28	1.31
ln(ME)	2.48	3.64	4.33	4.61	4.64	4.63	4.58	4.49	4.37	4.28	4.07	3.82	3.52
ln(BE/ME)	−0.10	−0.76	−0.91	−0.79	−0.61	−0.47	−0.33	−0.21	-0.08	0.02	0.15	0.26	0.40
ln(A/ME)	0.90	−0.05	−0.27	−0.16	0.03	0.18	0.31	0.44	0.58	0.70	0.85	1.01	1.25
ln(A/BE)	0.99	0.70	0.63	0.63	0.64	0.65	0.64	0.65	0.66	0.68	0 71	0.75	0.86
E/P dummy	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E(+)/P	0.00	0.01	0.03	0.05	0.06	0.08	0.09	0.11	0.12	0.14	0.16	0.20	0.28
Firms	355	88	90	182	190	193	196	194	197	195	195	95	91

110138-OPC-POD-60-57

of book leverage. The regressions use the natural logs of the leverage ratios, ln(A/ME) and ln(A/BE), because preliminary tests indicated that logs are a good functional form for capturing leverage effects in average returns. Using logs also leads to a simple interpretation of the relation between the roles of leverage and book-to-market equity in average returns.

The FM regressions of returns on the leverage variables (Table III) pose a bit of a puzzle. The two leverage variables are related to average returns, but with opposite signs. As in Bhandari (1988), higher market leverage is associated with higher average returns; the average slopes for ln(A/ME) are always positive and more than 4 standard errors from 0. But higher book leverage is associated with lower average returns; the average slopes for ln(A/BE) are always negative and more than 4 standard errors from 0.

The puzzle of the opposite slopes on ln(A/ME) and ln(A/BE) has a simple solution. The average slopes for the two leverage variables are opposite in sign but close in absolute value, e.g., 0.50 and −0.57. Thus it is the difference between market and book leverage that helps explain average returns. But the difference between market and book leverage is book-to-market equity, ln(BE/ME) = ln(A/ME) − ln(A/BE). Table III shows that the average book-to-market slopes in the FM regressions are indeed close in absolute value to the slopes for the two leverage variables.

The close links between the leverage and book-to-market results suggest that there are two equivalent ways to interpret the book-to-market effect in average returns. A high ratio of book equity to market equity (a low stock price relative to book value) says that the market judges the prospects of a firm to be poor relative to firms with low BE/ME. Thus BE/ME may capture the relative-distress effect postulated by Chan and Chen (1991). A high book-to-market ratio also says that a firm's market leverage is high relative to its book leverage; the firm has a large amount of market-imposed leverage because the market judges that its prospects are poor and discounts its stock price relative to book value. In short, our tests suggest that the relative-distress effect, captured by BE/ME, can also be interpreted as an involuntary leverage effect, which is captured by the difference between A/ME and A/BE.

B.3. E/P

Ball (1978) posits that the earnings-price ratio is a catch-all for omitted risk factors in expected returns. If current earnings proxy for expected future earnings, high-risk stocks with high expected returns will have low prices relative to their earnings. Thus, E/P should be related to expected returns, whatever the omitted sources of risk. This argument only makes sense, however, for firms with positive earnings. When current earnings are negative, they are not a proxy for the earnings forecasts embedded in the stock price, and E/P is not a proxy for expected returns. Thus, the slope for E/P in the FM regressions is based on positive values; we use a dummy variable for E/P when earnings are negative.

The U-shaped relation between average return and E/P observed in Table IV is also apparent when the E/P variables are used alone in the FM regressions in Table III. The average slope on the E/P dummy variable (0.57% per month, 2.28 standard errors from 0) confirms that firms with negative earnings have higher average returns. The average slope for stocks with positive E/P (4.72% per month, 4.57 standard errors from 0) shows that average returns increase with E/P when it is positive.

Adding size to the regressions kills the explanatory power of the E/P dummy. Thus the high average returns of negative E/P stocks are better captured by their size, which Table IV says is on average small. Adding both size and book-to-market equity to the E/P regressions kills the E/P dummy and lowers the average slope on E/P from 4.72 to 0.87 (t = 1.23). In contrast, the average slopes for ln(ME) and ln(BE/ME) in the regressions that include E/P are similar to those in the regressions that explain average returns with only size and book-to-market equity. The results suggest that most of the relation between (positive) E/P and average return is due to the positive correlation between E/P and ln(BE/ME), illustrated in Table IV; firms with high E/P tend to have high book-to-market equity ratios.

IV. A Parsimonious Model for Average Returns

The results to here are easily summarized:

(1) When we allow for variation in β that is unrelated to size, there is no reliable relation between β and average return.
(2) The opposite roles of market leverage and book leverage in average returns are captured well by book-to-market equity.
(3) The relation between E/P and average return seems to be absorbed by the combination of size and book-to-market equity.

In a nutshell, market β seems to have no role in explaining the average returns on NYSE, AMEX, and NASDAQ stocks for 1963–1990, while size and book-to-market equity capture the cross-sectional variation in average stock returns that is related to leverage and E/P.

A. Average Returns, Size and Book-to-Market Equity

The average return matrix in Table V gives a simple picture of the two-dimensional variation in average returns that results when the 10 size deciles are each subdivided into 10 portfolios based on ranked values of BE/ME for individual stocks. Within a size decile (across a row of the average return matrix), returns typically increase strongly with BE/ME: on average, the returns on the lowest and highest BE/ME portfolios in a size decile differ by 0.99% (1.63% − 0.64%) per month. Similarly, looking down the columns of the average return matrix shows that there is a negative relation between average return and size: on average, the spread of returns across the size portfolios in a BE/ME group is 0.58% per month. The average return matrix gives life to the conclusion from the regressions that,

Table V

Average Monthly Returns on Portfolios Formed on Size and Book-to-Market Equity; Stocks Sorted by ME (Down) and then BE/ME (Across): July 1963 to December 1990

In June of each year t, the NYSE, AMEX, and NASDAQ stocks that meet the CRSP-COMPUSTAT data requirements are allocated to 10 size portfolios using the NYSE size (ME) breakpoints. The NYSE, AMEX, and NASDAQ stocks in each size decile are then sorted into 10 BE/ME portfolios using the book-to-market ratios for year $t-1$. BE/ME is the book value of common equity plus balance-sheet deferred taxes for fiscal year $t-1$, over market equity for December of year $t-1$. The equal-weighted monthly portfolio returns are then calculated for July of year t to June of year $t+1$.

Average monthly return is the time-series average of the monthly equal-weighted portfolio returns (in percent).

The All column shows average returns for equal-weighted size decile portfolios. The All row shows average returns for equal-weighted portfolios of the stocks in each BE/ME group.

		Book-to-Market Portfolios									
	All	Low	2	3	4	5	6	7	8	9	High
All	1.23	0.64	0.98	1.06	1.17	1.24	1.26	1.39	1.40	1.50	1.63
Small-ME	1.47	0.70	1.14	1.20	1.43	1.56	1.51	1.70	1.71	1.82	1.92
ME-2	1.22	0.43	1.05	0.96	1.19	1.33	1.19	1.58	1.28	1.43	1.79
ME-3	1.22	0.56	0.88	1.23	0.95	1.36	1.30	1.30	1.40	1.54	1.60
ME-4	1.19	0.39	0.72	1.06	1.36	1.13	1.21	1.34	1.59	1.51	1.47
ME-5	1.24	0.88	0.65	1.08	1.47	1.13	1.43	1.44	1.26	1.52	1.49
ME-6	1.15	0.70	0.98	1.14	1.23	0.94	1.27	1.19	1.19	1.24	1.50
ME-7	1.07	0.95	1.00	0.99	0.83	0.99	1.13	0.99	1.16	1.10	1.47
ME-8	1.08	0.66	1.13	0.91	0.95	0.99	1.01	1.15	1.05	1.29	1 55
ME-9	0.95	0.44	0.89	0.92	1.00	1.05	0.93	0.82	1.11	1.04	1.22
Large-ME	0.89	0.93	0.88	0.84	0.71	0.79	0.83	0.81	0.96	0.97	1.18

controlling for size, book-to-market equity captures strong variation in average returns, and controlling for book-to-market equity leaves a size effect in average returns.

B. The Interaction between Size and Book-to-Market Equity

The average of the monthly correlations between the cross-sections of ln(ME) and ln(BE/ME) for individual stocks is -0.26. The negative correlation is also apparent in the average values of ln(ME) and ln(BE/ME) for the portfolios sorted on ME or BE/ME in Tables II and IV. Thus, firms with low market equity are more likely to have poor prospects, resulting in low stock prices and high book-to-market equity. Conversely, large stocks are more likely to be firms with stronger prospects, higher stock prices, lower book-to-market equity, and lower average stock returns.

The correlation between size and book-to-market equity affects the regressions in Table III. Including ln(BE/ME) moves the average slope on ln(ME) from -0.15 ($t = -2.58$) in the univariate regressions to -0.11 ($t = -1.99$) in the bivariate regressions. Similarly, including ln(ME) in the regressions lowers the average slope on ln(BE/ME) from 0.50 to 0.35 (still a healthy 4.44 standard errors from 0). Thus, part of the size effect in the simple regressions is due to the fact that small ME stocks are more likely to have high book-to-market ratios, and part of the simple book-to-market effect is due to the fact that high BE/ME stocks tend to be small (they have low ME).

We should not, however, exaggerate the links between size and book-to-market equity. The correlation (−0.26) between ln(ME) and ln(BE/ME) is not extreme, and the average slopes in the bivariate regressions in Table III show that ln(ME) and ln(BE/ME) are both needed to explain the cross-section of average returns. Finally, the 10×10 average return matrix in Table V provides concrete evidence that, (a) controlling for size, book-to-market equity captures substantial variation in the cross-section of average returns, and (b) within BE/ME groups average returns are related to size.

C. Subperiod Averages of the FM Slopes

The message from the average FM slopes for 1963–1990 (Table III) is that size on average has a negative premium in the cross-section of stock returns, book-to-market equity has a positive premium, and the average premium for market β is essentially 0. Table VI shows the average FM slopes for two roughly equal subperiods (July 1963–December 1976 and January 1977–December 1990) from two regressions: (a) the cross-section of stock returns on size, ln(ME), and book-to-market equity, ln(BE/ME), and (b) returns on β, ln(ME), and ln(BE/ME). For perspective, average returns on the value-weighted and equal-weighted (VW and EW) portfolios of NYSE stocks are also shown.

In FM regressions, the intercept is the return on a standard portfolio (the weights on stocks sum to 1) in which the weighted averages of the explanatory variables are 0 (Fama (1976), chapter 9). In our tests, the intercept is weighted toward small stocks (ME is in millions of dollars so ln(ME) = 0 implies ME = $1 million) and toward stocks with relatively high book-to-market ratios (Table IV says that ln(BE/ME) is negative for the typical firm, so ln(BE/ME) = 0 is toward the high end of the sample ratios). Thus it is not surprising that the average intercepts are always large relative to their standard errors and relative to the returns on the NYSE VW and EW portfolios.

Like the overall period, the subperiods do not offer much hope that the average premium for β is economically important. The average FM slope for β is only slightly positive for 1963–1976 (0.10% per month, $t = 0.25$), and it is negative for 1977–1990 (−0.44% per month, $t = -1.17$). There is a hint that the size effect is weaker in the 1977–1990 period, but inferences about the average size slopes for the subperiods lack power.

Unlike the size effect, the relation between book-to-market equity and average return is so strong that it shows up reliably in both the 1963–1976 and the 1977–1990 subperiods. The average slopes for ln(BE/ME) are all more than 2.95 standard errors from 0, and the average slopes for the

Table VI

Subperiod Average Monthly Returns on the NYSE Equal-Weighted and Value-Weighted Portfolios and Subperiod Means of the Intercepts and Slopes from the Monthly FM Cross-Sectional Regressions of Returns on (a) Size (ln(ME)) and Book-to-Market Equity (ln(BE/ME)), and (b) β, ln(ME), and ln(BE/ME)

Mean is the time-series mean of a monthly return, Std is its time-series standard deviation, and t(Mn) is Mean divided by its time-series standard error.

	7/63–12/90 (330 Mos.)			7/63–12/76 (162 Mos.)			1/77–12/90 (168 Mos.)		
Variable	Mean	Std	t(Mn)	Mean	Std	t(Mn)	Mean	Std	t(Mn)
	NYSE Value-Weighted (VW) and Equal-Weighted (EW) Portfolio Returns								
VW	0.81	4.47	3.27	0.56	4.26	1.67	1.04	4.66	2.89
EW	0.97	5.49	3.19	0.77	5.70	1.72	1.15	5.28	2.82
	$R_{it} = a + b_{2t}\ln(ME_{it}) + b_{3t}\ln(BE/ME_{it}) + e_{it}$								
a	1.77	8.51	3.77	1.86	10.10	2.33	1.69	6.67	3.27
b_2	−0.11	1.02	−1.99	−0.16	1.25	−1.62	−0.07	0.73	−1.16
b_3	0.35	1.45	4.43	0.36	1.53	2.96	0.35	1.37	3.30
	$R_{it} = a + b_{1t}\beta_{it} + b_{2t}\ln(ME_{it}) + b_{3t}\ln(BE/ME_{it}) + e_{it}$								
a	2.07	5.75	6.55	1.73	6.22	3.54	2.40	5.25	5.92
b_1	−0.17	5.12	−0.62	0.10	5.33	0.25	−0.44	4.91	−1.17
b_2	−0.12	0.89	−2.52	−0.15	1.03	−1.91	−0.09	0.74	−1.64
b_3	0.33	1.24	4.80	0.34	1.36	3.17	0.31	1.10	3.67

subperiods (0.36 and 0.35) are close to the average slope (0.35) for the overall period. The subperiod results thus support the conclusion that, among the variables considered here, book-to-market equity is consistently the most powerful for explaining the cross-section of average stock returns.

Finally, Roll (1983) and Keim (1983) show that the size effect is stronger in January. We have examined the monthly slopes from the FM regressions in Table VI for evidence of a January seasonal in the relation between book-to-market equity and average return. The average January slopes for ln(BE/ME) are about twice those for February to December. Unlike the size effect, however, the strong relation between book-to-market equity and average return is not special to January. The average monthly February-to-December slopes for ln(BE/ME) are about 4 standard errors from 0, and they are close to (within 0.05 of) the average slopes for the whole year. Thus, there is a January seasonal in the book-to-market equity effect, but the positive relation between BE/ME and average return is strong throughout the year.

D. β and the Market Factor: Caveats

Some caveats about the negative evidence on the role of β in average returns are in order. The average premiums for β, size, and book-to-market equity depend on the definitions of the variables used in the regressions. For example, suppose we replace book-to-market equity (ln(BE/ME)) with book equity (ln(BE)). As long as size (ln(ME)) is also in the regression, this change will not affect the intercept, the fitted values or the R^2. But the change, in variables increases the average slope (and the t-statistic) on ln(ME). In other words, it increases the risk premium associated with size. Other redefinitions of the β, size, and book-to-market variables will produce different regression slopes and perhaps different inferences about average premiums, including possible resuscitation of a role for β. And, of course, at the moment, we have no theoretical basis for choosing among different versions of the variables.

Moreover, the tests here are restricted to stocks. It is possible that including other assets will change the inferences about the average premiums for β, size, and book-to-market equity. For example, the large average intercepts for the FM regressions in Table VI suggest that the regressions will not do a good job on Treasury bills, which have low average returns and are likely to have small loadings on the underlying market, size, and book-to-market factors in returns. Extending the tests to bills and other bonds may well change our inferences about average risk premiums, including the revival of a role for market β.

We emphasize, however, that different approaches to the tests are not likely to revive the Sharpe-Lintner-Black model. Resuscitation of the SLB model requires that a better proxy for the market portfolio (a) overturns our evidence that the simple relation between β and average stock returns is flat and (b) leaves β as the only variable relevant for explaining average returns. Such results seem unlikely, given Stambaugh's (1982) evidence that tests of the SLB model do not seem to be sensitive to the choice of a market proxy. Thus, if there is a role for β in average returns, it is likely to be found in a multi-factor model that transforms the flat simple relation between average return and β into a positively sloped conditional relation.

V. Conclusions and Implications

The Sharpe-Lintner-Black model has long shaped the way academics and practitioners think about average return and risk. Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) find that, as predicted by the model, there is a positive simple relation between average return and market β during the early years (1926–1968) of the CRSP NYSE returns file. Like Reinganum (1981) and Lakonishok and Shapiro (1986), we find that this simple relation between β and average return disappears during the more recent 1963–1990 period. The appendix that follows shows that the relation between β and average return is also weak in the last half century (1941–1990) of returns on NYSE stocks. In short, our tests do not support the central prediction of the SLB model, that average stock returns are positively related to market β.

Banz (1981) documents a strong negative relation between average return and firm size. Bhandari (1988) finds that average return is positively related to leverage, and Basu (1983) finds a positive relation between average return and E/P. Stattman (1980) and Rosenberg, Reid, and Lanstein (1985) document a positive relation between average return and book-to-market equity for U.S. stocks, and Chan, Hamao, and Lakonishok (1992) find that BE/ME is also a powerful variable for explaining average returns on Japanese stocks.

Variables like size, E/P, leverage, and book-to-market equity are all scaled versions of a firm's stock price. They can be regarded as different ways of extracting information from stock prices about the cross-section of expected stock returns (Ball (1978), Keim (1988)). Since all these variables are scaled versions of price, it is reasonable to expect that some of them are redundant for explaining average returns. Our main result is that for the 1963–1990 period, size and book-to-market equity capture the cross-sectional variation in average stock returns associated with size, E/P, book-to-market equity, and leverage.

A. Rational Asset-Pricing Stories

Are our results consistent with asset-pricing theory? Since the FM intercept is constrained to be the same for all stocks, FM regressions always impose a linear factor structure on returns and expected returns that is consistent with the multifactor asset-pricing models of Merton (1973) and Ross (1976). Thus our tests impose a rational asset-pricing framework on the relation between average return and size and book-to-market equity.

Even if our results are consistent with asset-pricing theory, they are not economically satisfying. What is the economic explanation for the roles of size and book-to-market equity in average returns? We suggest several paths of inquiry.

(a) The intercepts and slopes in the monthly FM regressions of returns on ln(ME) and ln(BE/ME) are returns on portfolios that mimic the underlying common risk factors in returns proxied by size and book-to-market equity (Fama (1976), chapter 9). Examining the relations between the returns on these portfolios and economic variables that measure variation in business conditions might help expose the nature of the economic risks captured by size and book-to-market equity.

(b) Chan, Chen, and Hsieh (1985) argue that the relation between size and average return proxies for a more fundamental relation between expected returns and economic risk factors. Their most powerful factor in explaining the size effect is the difference between the monthly returns on low- and high-grade corporate bonds, which in principle captures a kind of default risk in returns that is priced. It would be interesting to test whether loadings on this or other economic factors, such as those of Chen, Roll, and Ross (1986), can explain the roles of size and book-to-market equity in our tests.

(c) In a similar vein, Chan and Chen (1991) argue that the relation between size and average return is a relative-prospects effect. The earning prospects of distressed firms are more sensitive to economic conditions. This results in a distress factor in returns that is priced in expected returns. Chan and Chen construct two mimicking portfolios for the distress factor, based on dividend changes and leverage. It would be interesting to check whether loadings on their distress factors absorb the size and book-to-market equity effects in average returns that are documented here.

(d) In fact, if stock prices are rational, BE/ME, the ratio of the book value of a stock to the market's assessment of its value, should be a direct indicator of the relative prospects of firms. For example, we expect that high BE/ME firms have low earnings on assets relative to low BE/ME firms. Our work (in progress) suggests that there is indeed a clean separation between high and low BE/ME firms on various measures of economic fundamentals. Low BE/ME firms are persistently strong performers, while the economic performance of high BE/ME firms is persistently weak.

B. Irrational Asset-Pricing Stories

The discussion above assumes that the asset-pricing effects captured by size and book-to-market equity are rational. For BE/ME, our most powerful expected-return variable, there is an obvious alternative. The cross-section of book-to-market ratios might result from market overreaction to the relative prospects of firms. If overreaction tends to be corrected, BE/ME will predict the cross-section of stock returns.

Simple tests do not confirm that the size and book-to-market effects in average returns are due to market overreaction, at least of the type posited by DeBondt and Thaler (1985). One overreaction measure used by DeBondt and Thaler is a stock's most recent 3-year return. Their overreaction story predicts that 3-year losers have strong post-ranking returns relative to 3-year winners. In FM regressions (not shown) for individual stocks, the 3-year lagged return shows no power even when used alone to explain average returns. The univariate average slope for the lagged return is negative, -6 basis points per month, but less than 0.5 standard errors from 0.

C. Applications

Our main result is that two easily measured variables, size and book-to-market equity, seem to describe the cross-section of average stock returns. Prescriptions for using this evidence depend on (a) whether it will persist, and (b) whether it results from rational or irrational asset-pricing.

It is possible that, by chance, size and book-to-market equity happen to describe the cross-section of average returns in our sample, but they were and are unrelated to expected returns. We put little weight on this possibility, especially for book-to-market equity. First, although BE/ME has long been touted as a measure of the return prospects of stocks, there is no evidence that its explanatory power deteriorates through time. The 1963–1990 relation between BE/ME and average return is strong, and remarkably similar for the 1963–1976 and 1977–1990 subperiods. Second, our preliminary work on economic fundamentals suggests that high-BE/ME firms tend to be persistently poor earners relative to low-BE/ME firms. Similarly, small firms have a long period of poor earnings during the 1980s not shared with big firms. The systematic patterns in fundamentals give us some hope that size and book-to-market equity proxy for risk factors in returns, related to relative earning prospects, that are rationally priced in expected returns.

If our results are more than chance, they have practical implications for portfolio formation and performance evaluation by investors whose primary concern is long-term average returns. If asset-pricing is rational, size and BE/ME must proxy for risk. Our results then imply that the performance of managed portfolios (e.g., pension funds and mutual funds) can be evaluated by comparing their average returns with the average returns of benchmark portfolios with similar size and BE/ME characteristics. Likewise, the expected returns for different portfolio strategies can be estimated from the historical average returns of portfolios with matching size and BE/ME properties.

If asset-pricing is irrational and size and BE/ME do not proxy for risk, our results might still be used to evaluate portfolio performance and measure the expected returns from alternative investment strategies. If stock prices are irrational, however, the likely persistence of the results is more suspect.

Appendix
Size Versus β: 1941–1990

Our results on the absence of a relation between β and average stock returns for 1963–1990 are so contrary to the tests of the Sharpe-Lintner-Black model by Black, Jensen, and Scholes (1972), Fama and MacBeth (1973), and (more recently) Chan and Chen (1988), that further tests are appropriate. We examine the roles of size and β in the average returns on NYSE stocks for the half-century 1941–1990, the longest available period that avoids the high volatility of returns in the Great Depression. We do not include the accounting variables in the tests because of the strong selection bias (toward successful firms) in the COMPUSTAT data prior to 1962.

We first replicate the results of Chan and Chen (1988). Like them, we find that when portfolios are formed on size alone, there are strong relations between average return and either size or β; average return increases with β and decreases with size. For size portfolios, however, size (ln(ME)) and β are almost perfectly correlated (−0.98), so it is difficult to distinguish between the roles of size and β in average returns.

One way to generate strong variation in β that is unrelated to size is to form portfolios on size and then on β. As in Tables I to III, we find that the resulting independent variation in β just about washes out the positive simple relation between average return and β observed when portfolios are formed on size alone. The results for NYSE stocks for 1941–1990 are thus much like those for NYSE, AMEX, and NASDAQ stocks for 1963–1990.

This appendix also has methodological goals. For example, the FM regressions in Table III use returns on individual stocks as the dependent variable. Since we allocate portfolio βs to individual stocks but use firm-specific values of other variables like size, β may be at a disadvantage in the regressions for individual stocks. This appendix shows, however, that regressions for portfolios, which put β and size on equal footing, produce results comparable to those for individual stocks.

A. Size Portfolios

Table AI shows average monthly returns and market βs for 12 portfolios of NYSE stocks formed on the basis of size (ME) at the end of each year from 1940 to 1989. For these size portfolios, there is a strong positive relation between average return and β. Average returns fall from 1.96% per month for the smallest ME portfolio (1A) to 0.93% for the largest (10B) and β falls from 1.60 to 0.95. (Note also that, as claimed earlier, estimating β as the sum of the slopes in the regression of a portfolio's return on the current and prior month's NYSE value-weighted return produces much larger βs for the smallest ME portfolios and slightly smaller βs for the largest ME portfolios.)

The FM regressions in Table AI confirm the positive simple relation between average return and β for size portfolios. In the regressions of the size-portfolio returns on β alone, the average premium for a unit of β is 1.45% per month. In the regressions of individual stock returns on β (where stocks are assigned the β of their size portfolio), the premium for a unit of β is 1.39%. Both estimates are about 3 standard errors from 0. Moreover, the βs of size portfolios do not leave a residual size effect; the average residuals from the simple regressions of returns on β in Table AI show no relation to size. These positive SLB results for 1941–1990 are like those obtained by Chan and Chen (1988) in tests on size portfolios for 1954–1983.

There is, however, evidence in Table AI that all is not well with the βs of the size portfolios. They do a fine job on the relation between size and average return, but they do a lousy job on their main task, the relation between β and average return. When the residuals from the regressions of returns on β are grouped using the pre-ranking βs of individual stocks, the average residuals are strongly positive for low-β stocks (0.51% per month for group 1A) and negative for high-β stocks (−1.05% for 10B). Thus the market lines estimated with size-portfolio βs exaggerate the tradeoff of average return for β; they underestimate average returns on low-β stocks and overestimate average returns on high-β stocks. This pattern in the β-sorted average residuals for individual stocks suggests that (a) there is variation in β across stocks that is lost in the size portfolios, and (b) this variation in β is not rewarded as well as the variation in β that is related to size.

B. Two-Pass Size-β Portfolios

Like Table I, Table AII shows that subdividing size deciles using the (pre-ranking) βs of individual stocks results in strong variation in β that is

Table AI

Average Returns, Post-Ranking βs and Fama-MacBeth Regression Slopes for Size Portfolios of NYSE Stocks: 1941–1990

At the end of each year $t-1$, stocks are assigned to 12 portfolios using ranked values of ME. Included are all NYSE stocks that have a CRSP price and shares for December of year $t-1$ and returns for at least 24 of the 60 months ending in December of year $t-1$ (for pre-ranking β estimates). The middle 8 portfolios cover size deciles 2 to 9. The 4 extreme portfolios (1A, 1B, 10A, and 10B) split the smallest and largest deciles in half. We compute equal-weighted returns on the portfolios for the 12 months of year t using all surviving stocks. Average Return is the time-series average of the monthly portfolio returns for 1941–1990, in percent. Average firms is the average number of stocks in the portfolios each month. The simple βs are estimated by regressing the 1941–1990 sample of post-ranking monthly returns for a size portfolio on the current month's value-weighted NYSE portfolio return. The sum βs are the sum of the slopes from a regression of the post-ranking monthly returns on the current and prior month's VW NYSE returns.

The independent variables in the Fama-MacBeth regressions are defined for each firm at the end of December of each year $t-1$. Stocks are assigned the post-ranking (sum) β of the size portfolio they are in at the end of year $t-1$. ME is price times shares outstanding at the end of year $t-1$. In the individual-stock regressions, these values of the explanatory variables are matched with CRSP returns for each of the 12 months of year t. The portfolio regressions match the equal-weighted portfolio returns with the equal-weighted averages of β and ln(ME) for the surviving stocks in each month of year t. Slope is the average of the (600) monthly FM regression slopes and SE is the standard error of the average slope. The residuals from the monthly regressions for year t are grouped into 12 portfolios on the basis of size (ME) or pre-ranking β (estimated with 24 to 60 months of data, as available) at the end of year $t-1$. The average residuals are the time-series averages of the monthly equal-weighted portfolio residuals, in percent. The average residuals for regressions (1) and (2) (not shown) are quite similar to those for regressions (4) and (5) (shown).

	Portfolios Formed on Size											
	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Ave. return	1.96	1.59	1.44	1.36	1.28	1.24	1.23	1.17	1.15	1.13	0.97	0.93
Ave. firms	57	56	110	107	107	108	111	113	115	118	59	59
Simple β	1.29	1.24	1.21	1.19	1.16	1 13	1.13	1.12	1.09	1.05	1.00	0.98
Standard error	0.07	0.05	0.04	0.03	0.02	0 02	0.02	0.02	0.01	0.01	0.01	0.01
Sum β	1.60	1.44	1.37	1.32	1.26	1.23	1.19	1.17	1.12	1.06	0.99	0.95
Standard error	0.10	0.06	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.01	0.01	0.01

Table AI—*Continued*

	Portfolio Regressions				Individual Stock Regressions			
	(1) β	(2) ln(ME)	(3) β and ln(ME)		(4) β	(5) ln(ME)	(6) β and ln(ME)	
Slope	1.45	−0.137	3 05	0.149	1.39	−0.133	0.71	−0 060
SE	0.47	0.044	1.51	0 115	0.46	0.043	0.81	0 062

Average Residuals for Stocks Grouped on Size

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Regression (4)	0.17	0.00	−0.04	−0.06	−0.05	−0.04	0 00	−0 03	0.03	0 08	0.01	0 04
Standard error	0.11	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.05	0.06
Regression (5)	0.30	0.02	−0.05	−0 06	−0.08	−0.07	−0.03	−0.04	0.02	0.08	0.01	0.13
Standard error	0.14	0.07	0 04	0.04	0 04	0 04	0.04	0.03	0.03	0.03	0.04	0 07
Regression (6)	0.20	0.02	−0.05	−0.07	−0.08	−0 06	−0 01	−0 02	0.04	0 09	0.00	0.06
Standard error	0.10	0.06	0.04	0 04	0 04	0.04	0.03	0.03	0.03	0.03	0 05	0 05

Average Residuals for Stocks Grouped on Pre-Ranking β

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Regression (4)	0.51	0.61	0.38	0.32	0.16	0.12	0.03	−0.10	−0.27	−0.31	−0.66	−1.05
Standard error	0.21	0.19	0.13	0.08	0 04	0.03	0.04	0 05	0.09	0 11	0 18	0.23
Regression (5)	−0.10	0.00	0.02	0.09	0.05	0.07	0.05	0.00	−0.03	−0.01	−0.11	−0 33
Standard error	0.11	0.10	0.07	0.05	0.04	0.03	0.03	0.04	0.05	0.07	0 10	0 13
Regression (6)	0.09	0.25	0.13	0.19	0.11	0.14	0.09	0.01	−0.11	−0.12	−0.38	−0 70
Standard error	0.41	0.37	0.24	0.14	0 07	0.04	0.04	0.09	0.16	0.21	0 34	0.43

Table AII

Properties of Portfolios Formed on Size and Pre-Ranking β: NYSE Stocks Sorted by ME (Down) then Pre-Ranking β (Across): 1941-1990

At the end of year $t - 1$, the NYSE stocks on CRSP are assigned to 10 size (ME) portfolios Each size decile is subdivided into 10 β portfolios using pre-ranking βs of individual stocks, estimated with 24 to 60 monthly returns (as available) ending in December of year $t - 1$. The equal-weighted monthly returns on the resulting 100 portfolios are then calculated for year t. The average returns are the time-series averages of the monthly returns, in percent. The post-ranking βs use the full 1941-1990 sample of post-ranking returns for each portfolio. The pre- and post-ranking βs are the sum of the slopes from a regression of monthly returns on the current and prior month's NYSE value-weighted market return. The average size for a portfolio is the time-series average of each month's average value of ln(ME) for stocks in the portfolio. ME is denominated in millions of dollars. There are, on average, about 10 stocks in each size-β portfolio each month. The All column shows parameter values for equal-weighted size-decile (ME) portfolios. The All rows show parameter values for equal-weighted portfolios of the stocks in each β group.

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
Panel A: Average Monthly Return (in Percent)											
All		1.22	1.30	1.32	1.35	1.36	1.34	1.29	1.34	1.14	1.10
Small-ME	1.78	1.74	1.76	2.08	1.91	1.92	1.72	1.77	1.91	1.56	1 46
ME-2	1.44	1.41	1.35	1.33	1.61	1.72	1.59	1.40	1.62	1.24	1.11
ME-3	1.36	1.21	1.40	1.22	1.47	1.34	1.51	1.33	1.57	1.33	1 21
ME-4	1.28	1.26	1.29	1.19	1.27	1.51	1.30	1.19	1.56	1.18	1.00
ME-5	1.24	1.22	1.30	1.28	1.33	1.21	1.37	1.41	1.31	0.92	1.06
ME-6	1.23	1.21	1.32	1 37	1.09	1.34	1.10	1.40	1.21	1.22	1.08
ME-7	1.17	1.08	1.23	1.37	1.27	1.19	1.34	1.10	1.11	0.87	1.17
ME-8	1.15	1.06	1.18	1.26	1.25	1.26	1.17	1.16	1.05	1.08	1.04
ME-9	1.13	0.99	1.13	1.00	1.24	1.28	1.31	1.15	1.11	1.09	1.05
Large-ME	0.95	0.99	1.01	1.12	1.01	0.89	0.95	0.95	1.00	0.90	0.68

Table AII—*Continued*

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
					Panel B: Post-Ranking β						
All		0.76	0.95	1.05	1.14	1 22	1.26	1 34	1.38	1.49	1 69
Small-ME	1 52	1.17	1.40	1.31	1.50	1.46	1.50	1 69	1.60	1.75	1.92
ME-2	1.37	0.86	1.09	1.12	1.24	1 39	1.42	1.48	1.60	1.69	1.91
ME-3	1.32	0.88	0.96	1.18	1.19	1.33	1.40	1.43	1.56	1 64	1.74
ME-4	1.26	0.69	0.95	1.06	1.15	1 24	1 29	1 46	1.43	1.64	1.83
ME-5	1.23	0.70	0 95	1.04	1.10	1.22	1.32	1.34	1.41	1.56	1.72
ME-6	1.19	0.68	0.86	1.04	1.13	1.20	1.20	1 35	1.36	1.48	1.70
ME-7	1.17	0 67	0.88	0.95	1 14	1.18	1.26	1.27	1 32	1.44	1.68
ME-8	1.12	0.64	0.83	0 99	1.06	1.14	1.14	1.21	1.26	1.39	1.58
ME-9	1.06	0.68	0.81	0.94	0.96	1.06	1.11	1.18	1.22	1.25	1.46
Large-ME	0 97	0 65	0.73	0 90	0.91	0.97	1.01	1.01	1.07	1.12	1.38
					Panel C: Average Size (ln(ME))						
All		4.39	4.39	4.40	4 40	4.39	4.40	4 38	4.37	4 37	4.34
Small-ME	1 93	2.04	1.99	2.00	1 96	1.92	1.92	1 91	1.90	1 87	1.80
ME-2	2.80	2.81	2.79	2.81	2.83	2 80	2 79	2.80	2.80	2.79	2.79
ME-3	3.27	3.28	3.27	3 28	3.27	3.27	3.28	3.29	3.27	3.27	3.26
ME-4	3.67	3.67	3.67	3.67	3.68	3.68	3.67	3.68	3.66	3.67	3.67
ME-5	4.06	4.07	4.06	4.05	4.06	4.07	4.06	4 05	4.05	4.06	4.06
ME-6	4 45	4.45	4.44	4.46	4.45	4.45	4 45	4.45	4.44	4.45	4.45
ME-7	4.87	4.86	4.87	4.86	4 87	4.87	4.88	4.87	4.87	4 85	4.87
ME-8	5.36	5.38	5.38	5.38	5.35	5.36	5.37	5.37	5.36	5 35	5.34
ME-9	5.98	5.96	5.98	5.99	6 00	5.98	5.98	5.97	5.95	5 96	5.96
Large-ME	7.12	7.10	7.12	7.16	7.17	7 20	7.29	7.14	7.09	7 04	6.83

independent of size. The β sort of a size decile always produces portfolios with similar average ln(ME) but much different (post-ranking) βs. Table AII also shows, however, that investors are not compensated for the variation in β that is independent of size. Despite the wide range of βs in each size decile, average returns show no tendency to increase with β. AII

The FM regressions in Table AIII formalize the roles of size and β in NYSE average returns for 1941–1990. The regressions of returns on β alone show that using the βs of the portfolios formed on size and β, rather than size alone, causes the average slope on β to fall from about 1.4% per month (Table AI) to about 0.23% (about 1 standard error from 0). Thus, allowing for variation in β that is unrelated to size flattens the relation between average return and β, to the point where it is indistinguishable from no relation at all.

The flatter market lines in Table AIII succeed, however, in erasing the negative relation between β and average residuals observed in the regressions of returns on β alone in Table AI. Thus, forming portfolios on size and β (Table AIII) produces a better description of the simple relation between average return and β than forming portfolios on size alone (Table AI). This improved description of the relation between average return and β is evidence that the β estimates for the two-pass size-β portfolios capture variation in true βs that is missed when portfolios are formed on size alone.

Unfortunately, the flatter market lines in Table AIII have a cost, the emergence of a residual size effect. Grouped on the basis of ME for individual stocks, the average residuals from the univariate regressions of returns on the βs of the 100 size-β portfolios are strongly positive for small stocks and negative for large stocks (0.60% per month for the smallest ME group, 1A, and -0.27% for the largest, 10B). Thus, when we allow for variation in β that is independent of size, the resulting βs leave a large size effect in average returns. This residual size effect is much like that observed by Banz (1981) with the βs of portfolios formed on size and β.

The correlation between size and β is -0.98 for portfolios formed on size alone. The independent variation in β obtained with the second-pass sort on β lowers the correlation to -0.50. The lower correlation means that bivariate regressions of returns on β and ln(ME) are more likely to distinguish true size effects from true β effects in average returns.

The bivariate regressions (Table AIII) that use the βs of the size-β portfolios are more bad news for β. The average slopes for ln(ME) are close to the values in the univariate size regressions, and almost 4 standard errors from 0, but the average slopes for β are negative and less than 1 standard error from 0. The message from the bivariate regressions is that there is a strong relation between size and average return. But like the regressions in Table AIII that explain average returns with β alone, the bivariate regressions say that there is no reliable relation between β and average returns when the tests use βs that are not close substitutes for size. These uncomfortable SLB results for NYSE stocks for 1941–1990 are much like those for NYSE, AMEX, and NASDAQ stocks for 1963–1990 in Table III.

C. Subperiod Diagnostics

Our results for 1941–1990 seem to contradict the evidence in Black, Jensen, and Scholes (BJS) (1972) and Fama and MacBeth (FM) (1973) that there is a reliable positive relation between average return and β. The βs in BJS and FM are from portfolios formed on β alone, and the market proxy is the NYSE equal-weighted portfolio. We use the βs of portfolios formed on size and β, and our market is the value-weighted NYSE portfolio. We can report, however, that our inference that there isn't much relation between β and average return is unchanged when (a) the market proxy is the NYSE EW portfolio, (b) portfolios are formed on just (pre-ranking) βs, or (c) the order of forming the size-β portfolios is changed from size then β to β then size.

A more important difference between our results and the earlier studies is the sample periods. The tests in BJS and FM end in the 1960s. Table AIV shows that when we split the 50-year 1941–1990 period in half, the univariate FM regressions of returns on β produce an average slope for 1941–1965 (0.50% per month, $t = 1.82$) more like that of the earlier studies. In contrast, the average slope on β for 1966–1990 is close to 0 (-0.02, $t = 0.06$).

But Table AIV also shows that drawing a distinction between the results for 1941–1965 and 1966–1990 is misleading. The stronger tradeoff of average return for β in the simple regressions for 1941–1965 is due to the first 10 years, 1941–1950. This is the only period in Table AIV that produces an average premium for β (1.26% per month) that is both positive and more than 2 standard errors from 0. Conversely, the weak relation between β and average return for 1966–1990 is largely due to 1981–1990. The strong negative average slope in the univariate regressions of returns on β for 1981–1990 (-1.01, $t = -2.10$) offsets a positive slope for 1971–1980 (0.82, $t = 1.27$).

The subperiod variation in the average slopes from the FM regressions of returns on β alone seems moot, however, given the evidence in Table AIV that adding size always kills any positive tradeoff of average return for β in the subperiods. Adding size to the regressions for 1941–1965 causes the average slope for β to drop from 0.50 ($t = 1.82$) to 0.07 ($t = 0.28$). In contrast, the average slope on size in the bivariate regressions (-0.16, $t = -2.97$) is close to its value (-0.17, $t = -2.88$) in the regressions of returns on ln(ME) alone. Similar comments hold for 1941–1950. In short, any evidence of a positive average premium for β in the subperiods seems to be a size effect in disguise.

D. Can the SLB Model Be Saved?

Before concluding that β has no explanatory power, it is appropriate to consider other explanations for our results. One possibility is that the variation in β produced by the β sorts of size deciles in just sampling error. If so, it is not surprising that the variation in β within a size decile is unrelated to average return, or that size dominates β in bivariate tests. The standard errors of the βs suggest, however, that this explanation cannot save the SLB

Table AIII

Average Slopes, Their Standard Errors (SE), and Average Residuals from Monthly FM Regressions for Individual NYSE Stocks and for Portfolios Formed on Size and Pre-Ranking β: 1941–1990

Stocks are assigned the post-ranking β of the size-β portfolio they are in at the end of year $t-1$ (Table AII). ln(ME) is the natural log of price times shares outstanding at the end of year $t-1$. In the individual-stock regressions, these values of the explanatory variables are matched with CRSP returns for each of the 12 months in year t. The portfolio regressions match the equal-weighted portfolio returns for the size-β portfolios (Table AII) with the equal-weighted averages of β and ln(ME) for the surviving stocks in each month of year t. Slope is the time-series average of the monthly regression slopes from 1941–1990 (600 months); SE is the time-series standard error of the average slope.

The residuals from the monthly regressions in year t are grouped into 12 portfolios on the basis of size or pre-ranking β (estimated with 24 to 60 months of returns, as available) as of the end of year $t-1$. The average residuals are the time-series averages of the monthly equal-weighted averages of the residuals in percent. The average residuals (not shown) from the FM regressions (1) to (3) that use the returns on the 100 size-β portfolios as the dependent variable are always within 0.01 of those from the regressions for individual stock returns. This is not surprising given that the correlation between the time-series of 1941–1990 monthly FM slopes on β or ln(ME) for the comparable portfolio and individual stock regressions is always greater than 0.99.

	Portfolio Regressions				Individual Stock Regressions			
	(1) β	(2) ln(ME)	(3) β and ln(ME)		(4) β	(5) ln(ME)	(6) β and ln(ME)	
Slope	0.22	−0.128	−0.13	−0.143	0.24	−0.133	−0.14	−0.147
SE	0.24	0.043	0.21	0.039	0.23	0.043	0.21	0.039

Average Residuals for Stocks Grouped on Size

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Regression (4)	0.60	0.26	0.13	0.06	−0.01	−0.03	−0.03	−0.09	−0.10	−0.11	−0.25	−0.27
Standard error	0.21	0.10	0.06	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.06	0.08
Regression (5)	0.30	0.02	−0.05	−0.06	−0.08	−0.07	−0.03	−0.04	0.02	0.08	0.01	0.13
Standard error	0.14	0.07	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.07
Regression (6)	0.31	0.02	−0.05	−0.06	−0.09	−0.07	−0.03	−0.04	0.02	0.08	0.01	0.13
Standard error	0.14	0.07	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.07

Table AIII—*Continued*

	Portfolio Regressions			Individual Stock Regressions		
	(1) β	(2) ln(ME)	(3) β and ln(ME)	(4) β	(5) ln(ME)	(6) β and ln(ME)

Average Residuals for Stocks Grouped on Pre-Ranking β

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Regression (4)	−0.08	0.03	−0.01	0.08	0.04	0.08	0.04	0.02	−0.03	0.02	−0.11	−0.32
Standard error	0.07	0.05	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.06	0.07
Regression (5)	−0.10	0.00	0.02	0.09	0.05	0.07	0.05	0.00	−0.03	−0.01	−0.11	−0.33
Standard error	0.11	0.10	0.07	0.05	0.04	0.03	0.03	0.04	0.05	0.07	0.10	0.13
Regression (6)	−0.17	−0.07	−0.02	0.07	0.04	0.06	0.05	0.03	0.00	0.04	−0.04	−0.23
Standard error	0.05	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.06	0.07

Table AIV

Subperiod Average Returns on the NYSE Value-Weighted and Equal-Weighted Portfolios and Average Values of the Intercepts and Slopes for the FM Cross-Sectional Regressions of Individual Stock Returns on β and Size (ln(ME))

Mean is the average VW or EW return or an average slope from the *monthly* cross-sectional regressions of individual stock returns on β and/or ln(ME). Std is the standard deviation of the time-series of returns or slopes, and t(Mn) is Mean over its time-series standard error. The average slopes (not shown) from the FM regressions that use the returns on the 100 size-β portfolios of Table AII as the dependent variable are quite close to those for individual stock returns. (The correlation between the 1941–1990 month-by-month slopes on β or ln(ME) for the comparable portfolio and individual stock regressions is always greater than 0.99.)

Panel A

Variable	1941–1990 (600 Mos.)			1941–1965 (300 Mos.)			1966–1990 (300 Mos.)		
	Mean	Std	t(Mn)	Mean	Std	t(Mn)	Mean	Std	t(Mn)
	NYSE Value-Weighted (VW) and Equal-Weighted (EW) Portfolio Returns								
VW	0.93	4.15	5 49	1.10	3.58	5.30	0.76	4.64	2.85
EW	1.12	5.10	5.37	1.33	4.42	5.18	0.91	5.70	2.77
	$R_{it} = a + b_{1t}\beta_{it} + e_{it}$								
a	0.98	3.93	6.11	0.84	3.18	4.56	1.13	4.57	4.26
b_1	0.24	5.52	1.07	0.50	4.75	1.82	−0.02	6.19	−0 06
	$R_{it} = a + b_{2t}\ln(ME_{it}) + e_{it}$								
a	1.70	8.24	5.04	1.88	6.43	5.06	1.51	9.72	2.69
b_2	−0.13	1.06	−3.07	−0.17	1.01	−2.88	−0.10	1.11	−1.54
	$R_{it} = a + b_{1t}\beta_{it} + b_{2t}\ln(ME_{it}) + e_{it}$								
a	1.97	6.16	7.84	1.80	4.77	6.52	2.14	7.29	5.09
b_1	−0.14	5.05	−0.66	0.07	4.15	0.28	−0.34	5.80	−1.01
b_2	−0.15	0.96	−3.75	−0.16	0.94	−2.97	−0.13	0.99	−2.34

Table AIV—*Continued*

Panel B:

Return	1941-1950 Mean	1941-1950 t(Mn)	1951-1960 Mean	1951-1960 t(Mn)	1961-1970 Mean	1961-1970 t(Mn)	1971-1980 Mean	1971-1980 t(Mn)	1981-1990 Mean	1981-1990 t(Mn)
	NYSE Value-Weighted (VW) and Equal-Weighted (EW) Portfolio Returns									
VW	1.05	2.88	1.18	3.95	0.66	1.84	0.72	1.67	1.04	2.40
EW	1.59	3.16	1.13	3.76	0.88	1.96	1.04	1.82	0.95	2.01
	$R_{it} = a + b_{1t}\beta_{it} + e_{it}$									
a	0.24	0.66	1.41	6.36	0.64	1.94	0.27	0.62	2.35	5.99
b_1	1.26	2.20	−0.19	−0.63	0.32	0.72	0.82	1.27	−1.01	−2.10
	$R_{it} = a + b_{2t}\ln(ME_{it}) + e_{it}$									
a	2.63	3.47	1.08	2.73	1.78	2.50	2.18	2.03	0.82	1.20
b_2	−0.37	−2.90	0.03	0.53	−0.17	−2.19	−0.20	−1.57	0.04	0.57
	$R_{it} = a + b_{1t}\beta_{it} + b_{2t}\ln(ME_{it}) + e_{it}$									
a	2.14	3.93	1.38	4.03	2.01	4.16	1.50	2.12	2.84	4.25
b_1	0.34	0.75	−0.17	−0.53	−0 11	−0.27	0.41	0.75	−1.14	−2.16
b_2	−0.34	−2.92	0.01	0.20	−0.18	−2.89	−0.16	−1.50	−0.07	−0.84

model. The standard errors for portfolios formed on size and β are only slightly larger (0.02 to 0.11) than those for portfolios formed on size alone (0.01 to 0.10, Table AI). And the range of the post-ranking βs within a size decile is always large relative to the standard errors of the βs.

Another possibility is that the proportionality condition (1) for the variation through time in true βs, that justifies the use of full-period post-ranking βs in the FM tests, does not work well for portfolios formed on size and β. If this is a problem, post-ranking βs for the size-β portfolios should not be highly correlated across subperiods. The correlation between the half-period (1941–1965 and 1966–1990) βs of the size-β portfolios is 0.91, which we take to be good evidence that the full-period β estimates for these portfolios are informative about true βs. We can also report that using 5-year βs (pre- or post-ranking) in the FM regressions does not change our negative conclusions about the role of β in average returns, as long as portfolios are formed on β as well as size, or on β alone.

Any attempt to salvage the simple positive relation between β and average return predicted by the SLB model runs into three damaging facts, clear in Table AII. (a) Forming portfolios on size and pre-ranking βs produces a wide range of post-ranking βs in every size decile. (b) The post-ranking βs closely reproduce (in deciles 2 to 10 they exactly reproduce) the ordering of the pre-ranking βs used to form the β-sorted portfolios. It seems safe to conclude that the increasing pattern of the post-ranking βs in every size decile captures the ordering of the true βs. (c) Contrary to the SLB model, the β sorts do not produce a similar ordering of average returns. Within the rows (size deciles) of the average return matrix in Table AII, the high-β portfolios have average returns that are close to or less than the low-β portfolios.

But the most damaging evidence against the SLB model comes from the univariate regressions of returns on β in Table AIII. They say that when the tests allow for variation in β that is unrelated to size, the relation between β and average return for 1941–1990 is weak, perhaps nonexistent, even when β is the only explanatory variable. We are forced to conclude that the SLB model does not describe the last 50 years of average stock returns.

REFERENCES

Alford, Andrew, Jennifer J. Jones, and Mark E. Zmijewski, 1992, Extensions and violations of the statutory SEC Form 10-K filing date, Unpublished manuscript, University of Chicago, Chicago, IL.

Ball, Ray, 1978, Anomalies in relationships between securities' yields and yield-surrogates, *Journal of Financial Economics* 6, 103–126.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Basu, Sanjoy, 1983, The relationship between earnings yield, market value, and return for NYSE common stocks: Further evidence, *Journal of Financial Economics* 12, 129–156.

Bhandari, Laxmi Chand, 1988, Debt/Equity ratio and expected common stock returns: Empirical evidence, *Journal of Finance* 43, 507–528.

Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, *Journal of Business* 45, 444–455.

———, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: some empirical tests, in M. Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger).

Chan, Louis K., Yasushi Hamao, and Josef Lakonishok, 1991, Fundamentals and stock returns in Japan, *Journal of Finance* 46, 1739-1789

Chan, K. C. and Nai-fu Chen, 1988, An unconditional asset-pricing test and the role of firm size as an instrumental variable for risk, *Journal of Finance* 43, 309-325.

———, and Nai-fu Chen, 1991, Structural and return characteristics of small and large firms, *Journal of Finance* 46, 1467-1484.

———, Nai-fu Chen, and David A. Hsieh, 1985, An exploratory investigation of the firm size effect, *Journal of Financial Economics* 14, 451-471.

Chen, Nai-fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 56, 383-403.

DeBondt, Werner F. M., and Richard H. Thaler, 1985, Does the stock market overreact, *Journal of Finance* 40, 557-581.

Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 197-226.

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

———, and James MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.

Fowler, David J. and C. Harvey Rorke, 1983, Risk measurement when shares are subject to infrequent trading: Comment, *Journal of Financial Economics* 12, 279-283.

Jaffe, Jeffrey, Donald B. Keim, and Randolph Westerfield, 1989, Earnings yields, market values, and stock returns, *Journal of Finance* 44, 135-148.

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

———, 1988, Stock market regularities: A synthesis of the evidence and explanations, in Elroy Dimson, ed.: *Stock Market Anomalies* (Cambridge University Press, Cambridge).

Lakonishok, Josef, and Alan C. Shapiro, 1986, Systematic risk, total risk and size as determinants of stock market returns, *Journal of Banking and Finance* 10, 115-132.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13-37.

Markowitz, Harry, 1959, *Portfolio Selection: Efficient Diversification of Investments* (Wiley, New York).

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867-887.

Reinganum, Marc R., 1981, A new empirical perspective on the CAPM, *Journal of Financial and Quantitative Analysis* 16, 439-462.

Roll, Richard, 1983, Vas ist Das? The turn-of-the-year effect and the return premia of small firms, *Journal of Portfolio Management* 9, 18-28.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9-17.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341-360.

Sharpe, William F., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425-442.

Stambaugh, Robert F., 1982, On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis, *Journal of Financial Economics* 10, 237-268.

Stattman, Dennis, 1980, Book values and stock returns, *The Chicago MBA: A Journal of Selected Papers* 4, 25-45.

Journal of Economic Perspectives—Volume 18, Number 3—Summer 2004—Pages 25–46

The Capital Asset Pricing Model: Theory and Evidence

Eugene F. Fama and Kenneth R. French

The capital asset pricing model (CAPM) of William Sharpe (1964) and John Lintner (1965) marks the birth of asset pricing theory (resulting in a Nobel Prize for Sharpe in 1990). Four decades later, the CAPM is still widely used in applications, such as estimating the cost of capital for firms and evaluating the performance of managed portfolios. It is the centerpiece of MBA investment courses. Indeed, it is often the only asset pricing model taught in these courses.[1]

The attraction of the CAPM is that it offers powerful and intuitively pleasing predictions about how to measure risk and the relation between expected return and risk. Unfortunately, the empirical record of the model is poor—poor enough to invalidate the way it is used in applications. The CAPM's empirical problems may reflect theoretical failings, the result of many simplifying assumptions. But they may also be caused by difficulties in implementing valid tests of the model. For example, the CAPM says that the risk of a stock should be measured relative to a comprehensive "market portfolio" that in principle can include not just traded financial assets, but also consumer durables, real estate and human capital. Even if we take a narrow view of the model and limit its purview to traded financial assets, is it

[1] Although every asset pricing model is a capital asset pricing model, the finance profession reserves the acronym CAPM for the specific model of Sharpe (1964), Lintner (1965) and Black (1972) discussed here. Thus, throughout the paper we refer to the Sharpe-Lintner-Black model as the CAPM.

■ *Eugene F. Fama is Robert R. McCormick Distinguished Service Professor of Finance, Graduate School of Business, University of Chicago, Chicago, Illinois. Kenneth R. French is Carl E. and Catherine M. Heidt Professor of Finance, Tuck School of Business, Dartmouth College, Hanover, New Hampshire. Their e-mail addresses are ⟨eugene.fama@gsb.uchicago.edu⟩ and ⟨kfrench@dartmouth.edu⟩, respectively.*

legitimate to limit further the market portfolio to U.S. common stocks (a typical choice), or should the market be expanded to include bonds, and other financial assets, perhaps around the world? In the end, we argue that whether the model's problems reflect weaknesses in the theory or in its empirical implementation, the failure of the CAPM in empirical tests implies that most applications of the model are invalid.

We begin by outlining the logic of the CAPM, focusing on its predictions about risk and expected return. We then review the history of empirical work and what it says about shortcomings of the CAPM that pose challenges to be explained by alternative models.

The Logic of the CAPM

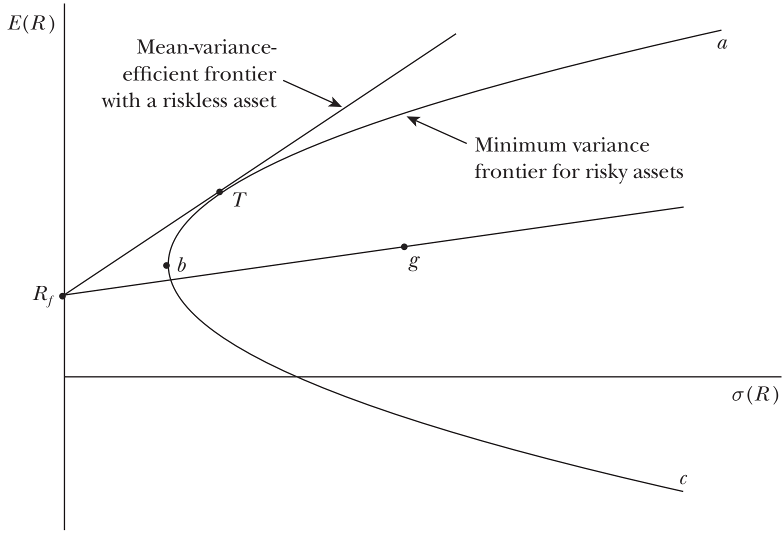
The CAPM builds on the model of portfolio choice developed by Harry Markowitz (1959). In Markowitz's model, an investor selects a portfolio at time $t - 1$ that produces a stochastic return at t. The model assumes investors are risk averse and, when choosing among portfolios, they care only about the mean and variance of their one-period investment return. As a result, investors choose "mean-variance-efficient" portfolios, in the sense that the portfolios 1) minimize the variance of portfolio return, given expected return, and 2) maximize expected return, given variance. Thus, the Markowitz approach is often called a "mean-variance model."

The portfolio model provides an algebraic condition on asset weights in mean-variance-efficient portfolios. The CAPM turns this algebraic statement into a testable prediction about the relation between risk and expected return by identifying a portfolio that must be efficient if asset prices are to clear the market of all assets.

Sharpe (1964) and Lintner (1965) add two key assumptions to the Markowitz model to identify a portfolio that must be mean-variance-efficient. The first assumption is *complete agreement*: given market clearing asset prices at $t - 1$, investors agree on the joint distribution of asset returns from $t - 1$ to t. And this distribution is the true one—that is, it is the distribution from which the returns we use to test the model are drawn. The second assumption is that there is *borrowing and lending at a risk-free rate*, which is the same for all investors and does not depend on the amount borrowed or lent.

Figure 1 describes portfolio opportunities and tells the CAPM story. The horizontal axis shows portfolio risk, measured by the standard deviation of portfolio return; the vertical axis shows expected return. The curve *abc*, which is called the minimum variance frontier, traces combinations of expected return and risk for portfolios of risky assets that minimize return variance at different levels of expected return. (These portfolios do not include risk-free borrowing and lending.) The tradeoff between risk and expected return for minimum variance portfolios is apparent. For example, an investor who wants a high expected return, perhaps at point *a*, must accept high volatility. At point *T*, the investor can have an interme-

Figure 1
Investment Opportunities



diate expected return with lower volatility. If there is no risk-free borrowing or lending, only portfolios above b along abc are mean-variance-efficient, since these portfolios also maximize expected return, given their return variances.

Adding risk-free borrowing and lending turns the efficient set into a straight line. Consider a portfolio that invests the proportion x of portfolio funds in a risk-free security and $1 - x$ in some portfolio g. If all funds are invested in the risk-free security—that is, they are loaned at the risk-free rate of interest—the result is the point R_f in Figure 1, a portfolio with zero variance and a risk-free rate of return. Combinations of risk-free lending and positive investment in g plot on the straight line between R_f and g. Points to the right of g on the line represent borrowing at the risk-free rate, with the proceeds from the borrowing used to increase investment in portfolio g. In short, portfolios that combine risk-free lending or borrowing with some risky portfolio g plot along a straight line from R_f through g in Figure 1.[2]

[2] Formally, the return, expected return and standard deviation of return on portfolios of the risk-free asset f and a risky portfolio g vary with x, the proportion of portfolio funds invested in f, as

$$R_p = xR_f + (1 - x)R_g,$$

$$E(R_p) = xR_f + (1 - x)E(R_g),$$

$$\sigma(R_p) = (1 - x)\sigma(R_g),\ x \leq 1.0,$$

which together imply that the portfolios plot along the line from R_f through g in Figure 1.

To obtain the mean-variance-efficient portfolios available with risk-free borrowing and lending, one swings a line from R_f in Figure 1 up and to the left as far as possible, to the tangency portfolio T. We can then see that all efficient portfolios are combinations of the risk-free asset (either risk-free borrowing or lending) and a single risky tangency portfolio, T. This key result is Tobin's (1958) "separation theorem."

The punch line of the CAPM is now straightforward. With complete agreement about distributions of returns, all investors see the same opportunity set (Figure 1), and they combine the same risky tangency portfolio T with risk-free lending or borrowing. Since all investors hold the same portfolio T of risky assets, it must be the value-weight market portfolio of risky assets. Specifically, each risky asset's weight in the tangency portfolio, which we now call M (for the "market"), must be the total market value of all outstanding units of the asset divided by the total market value of all risky assets. In addition, the risk-free rate must be set (along with the prices of risky assets) to clear the market for risk-free borrowing and lending.

In short, the CAPM assumptions imply that the market portfolio M must be on the minimum variance frontier if the asset market is to clear. This means that the algebraic relation that holds for any minimum variance portfolio must hold for the market portfolio. Specifically, if there are N risky assets,

$$\text{(Minimum Variance Condition for } M) \quad E(R_i) = E(R_{ZM}) + [E(R_M) - E(R_{ZM})]\beta_{iM}, \; i = 1, \ldots, N.$$

In this equation, $E(R_i)$ is the expected return on asset i, and β_{iM}, the market beta of asset i, is the covariance of its return with the market return divided by the variance of the market return,

$$\text{(Market Beta)} \quad \beta_{iM} = \frac{\text{cov}(R_i, R_M)}{\sigma^2(R_M)}.$$

The first term on the right-hand side of the minimum variance condition, $E(R_{ZM})$, is the expected return on assets that have market betas equal to zero, which means their returns are uncorrelated with the market return. The second term is a risk premium—the market beta of asset i, β_{iM}, times the premium per unit of beta, which is the expected market return, $E(R_M)$, minus $E(R_{ZM})$.

Since the market beta of asset i is also the slope in the regression of its return on the market return, a common (and correct) interpretation of beta is that it measures the sensitivity of the asset's return to variation in the market return. But there is another interpretation of beta more in line with the spirit of the portfolio model that underlies the CAPM. The risk of the market portfolio, as measured by the variance of its return (the denominator of β_{iM}), is a weighted average of the covariance risks of the assets in M (the numerators of β_{iM} for different assets).

Thus, β_{iM} is the covariance risk of asset i in M measured relative to the average covariance risk of assets, which is just the variance of the market return.[3] In economic terms, β_{iM} is proportional to the risk each dollar invested in asset i contributes to the market portfolio.

The last step in the development of the Sharpe-Lintner model is to use the assumption of risk-free borrowing and lending to nail down $E(R_{ZM})$, the expected return on zero-beta assets. A risky asset's return is uncorrelated with the market return—its beta is zero—when the average of the asset's covariances with the returns on other assets just offsets the variance of the asset's return. Such a risky asset is riskless in the market portfolio in the sense that it contributes nothing to the variance of the market return.

When there is risk-free borrowing and lending, the expected return on assets that are uncorrelated with the market return, $E(R_{ZM})$, must equal the risk-free rate, R_f. The relation between expected return and beta then becomes the familiar Sharpe-Lintner CAPM equation,

$$\text{(Sharpe-Lintner CAPM)} \quad E(R_i) = R_f + [E(R_M) - R_f)]\beta_{iM}, \; i = 1, \ldots, N.$$

In words, the expected return on any asset i is the risk-free interest rate, R_f, plus a risk premium, which is the asset's market beta, β_{iM}, times the premium per unit of beta risk, $E(R_M) - R_f$.

Unrestricted risk-free borrowing and lending is an unrealistic assumption. Fischer Black (1972) develops a version of the CAPM without risk-free borrowing or lending. He shows that the CAPM's key result—that the market portfolio is mean-variance-efficient—can be obtained by instead allowing unrestricted short sales of risky assets. In brief, back in Figure 1, if there is no risk-free asset, investors select portfolios from along the mean-variance-efficient frontier from a to b. Market clearing prices imply that when one weights the efficient portfolios chosen by investors by their (positive) shares of aggregate invested wealth, the resulting portfolio is the market portfolio. The market portfolio is thus a portfolio of the efficient portfolios chosen by investors. With unrestricted short selling of risky assets, portfolios made up of efficient portfolios are themselves efficient. Thus, the market portfolio is efficient, which means that the minimum variance condition for M given above holds, and it is the expected return-risk relation of the Black CAPM.

The relations between expected return and market beta of the Black and Sharpe-Lintner versions of the CAPM differ only in terms of what each says about $E(R_{ZM})$, the expected return on assets uncorrelated with the market. The Black version says only that $E(R_{ZM})$ must be less than the expected market return, so the

[3] Formally, if x_{iM} is the weight of asset i in the market portfolio, then the variance of the portfolio's return is

$$\sigma^2(R_M) = Cov(R_M, R_M) = Cov\left(\sum_{i=1}^{N} x_{iM}R_i,\, R_M\right) = \sum_{i=1}^{N} x_{iM}Cov(R_i, R_M).$$

premium for beta is positive. In contrast, in the Sharpe-Lintner version of the model, $E(R_{ZM})$ must be the risk-free interest rate, R_f, and the premium per unit of beta risk is $E(R_M) - R_f$.

The assumption that short selling is unrestricted is as unrealistic as unrestricted risk-free borrowing and lending. If there is no risk-free asset and short sales of risky assets are not allowed, mean-variance investors still choose efficient portfolios—points above *b* on the *abc* curve in Figure 1. But when there is no short selling of risky assets and no risk-free asset, the algebra of portfolio efficiency says that portfolios made up of efficient portfolios are not typically efficient. This means that the market portfolio, which is a portfolio of the efficient portfolios chosen by investors, is not typically efficient. And the CAPM relation between expected return and market beta is lost. This does not rule out predictions about expected return and betas with respect to other efficient portfolios—if theory can specify portfolios that must be efficient if the market is to clear. But so far this has proven impossible.

In short, the familiar CAPM equation relating expected asset returns to their market betas is just an application to the market portfolio of the relation between expected return and portfolio beta that holds in any mean-variance-efficient portfolio. The efficiency of the market portfolio is based on many unrealistic assumptions, including complete agreement and either unrestricted risk-free borrowing and lending or unrestricted short selling of risky assets. But all interesting models involve unrealistic simplifications, which is why they must be tested against data.

Early Empirical Tests

Tests of the CAPM are based on three implications of the relation between expected return and market beta implied by the model. First, expected returns on all assets are linearly related to their betas, and no other variable has marginal explanatory power. Second, the beta premium is positive, meaning that the expected return on the market portfolio exceeds the expected return on assets whose returns are uncorrelated with the market return. Third, in the Sharpe-Lintner version of the model, assets uncorrelated with the market have expected returns equal to the risk-free interest rate, and the beta premium is the expected market return minus the risk-free rate. Most tests of these predictions use either cross-section or time-series regressions. Both approaches date to early tests of the model.

Tests on Risk Premiums

The early cross-section regression tests focus on the Sharpe-Lintner model's predictions about the intercept and slope in the relation between expected return and market beta. The approach is to regress a cross-section of average asset returns on estimates of asset betas. The model predicts that the intercept in these regressions is the risk-free interest rate, R_f, and the coefficient on beta is the expected return on the market in excess of the risk-free rate, $E(R_M) - R_f$.

Two problems in these tests quickly became apparent. First, estimates of beta for individual assets are imprecise, creating a measurement error problem when they are used to explain average returns. Second, the regression residuals have common sources of variation, such as industry effects in average returns. Positive correlation in the residuals produces downward bias in the usual ordinary least squares estimates of the standard errors of the cross-section regression slopes.

To improve the precision of estimated betas, researchers such as Blume (1970), Friend and Blume (1970) and Black, Jensen and Scholes (1972) work with portfolios, rather than individual securities. Since expected returns and market betas combine in the same way in portfolios, if the CAPM explains security returns it also explains portfolio returns.[4] Estimates of beta for diversified portfolios are more precise than estimates for individual securities. Thus, using portfolios in cross-section regressions of average returns on betas reduces the critical errors in variables problem. Grouping, however, shrinks the range of betas and reduces statistical power. To mitigate this problem, researchers sort securities on beta when forming portfolios; the first portfolio contains securities with the lowest betas, and so on, up to the last portfolio with the highest beta assets. This sorting procedure is now standard in empirical tests.

Fama and MacBeth (1973) propose a method for addressing the inference problem caused by correlation of the residuals in cross-section regressions. Instead of estimating a single cross-section regression of average monthly returns on betas, they estimate month-by-month cross-section regressions of monthly returns on betas. The times-series means of the monthly slopes and intercepts, along with the standard errors of the means, are then used to test whether the average premium for beta is positive and whether the average return on assets uncorrelated with the market is equal to the average risk-free interest rate. In this approach, the standard errors of the average intercept and slope are determined by the month-to-month variation in the regression coefficients, which fully captures the effects of residual correlation on variation in the regression coefficients, but sidesteps the problem of actually estimating the correlations. The residual correlations are, in effect, captured via repeated sampling of the regression coefficients. This approach also becomes standard in the literature.

Jensen (1968) was the first to note that the Sharpe-Lintner version of the

[4] Formally, if x_{ip}, $i = 1, \ldots, N$, are the weights for assets in some portfolio p, the expected return and market beta for the portfolio are related to the expected returns and betas of assets as

$$E(R_p) = \sum_{i=1}^{N} x_{ip}E(R_i), \text{ and } \beta_{pM} = \sum_{i=1}^{N} x_{ip}\beta_{pM}.$$

Thus, the CAPM relation between expected return and beta,

$$E(R_i) = E(R_f) + [E(R_M) - E(R_f)]\beta_{iM},$$

holds when asset i is a portfolio, as well as when i is an individual security.

relation between expected return and market beta also implies a time-series regression test. The Sharpe-Lintner CAPM says that the expected value of an asset's excess return (the asset's return minus the risk-free interest rate, $R_{it} - R_{ft}$) is completely explained by its expected CAPM risk premium (its beta times the expected value of $R_{Mt} - R_{ft}$). This implies that "Jensen's alpha," the intercept term in the time-series regression,

$$\text{(Time-Series Regression)} \quad R_{it} - R_{ft} = \alpha_i + \beta_{iM}(R_{Mt} - R_{ft}) + \varepsilon_{it},$$

is zero for each asset.

The early tests firmly reject the Sharpe-Lintner version of the CAPM. There is a positive relation between beta and average return, but it is too "flat." Recall that, in cross-section regressions, the Sharpe-Lintner model predicts that the intercept is the risk-free rate and the coefficient on beta is the expected market return in excess of the risk-free rate, $E(R_M) - R_f$. The regressions consistently find that the intercept is greater than the average risk-free rate (typically proxied as the return on a one-month Treasury bill), and the coefficient on beta is less than the average excess market return (proxied as the average return on a portfolio of U.S. common stocks minus the Treasury bill rate). This is true in the early tests, such as Douglas (1968), Black, Jensen and Scholes (1972), Miller and Scholes (1972), Blume and Friend (1973) and Fama and MacBeth (1973), as well as in more recent cross-section regression tests, like Fama and French (1992).

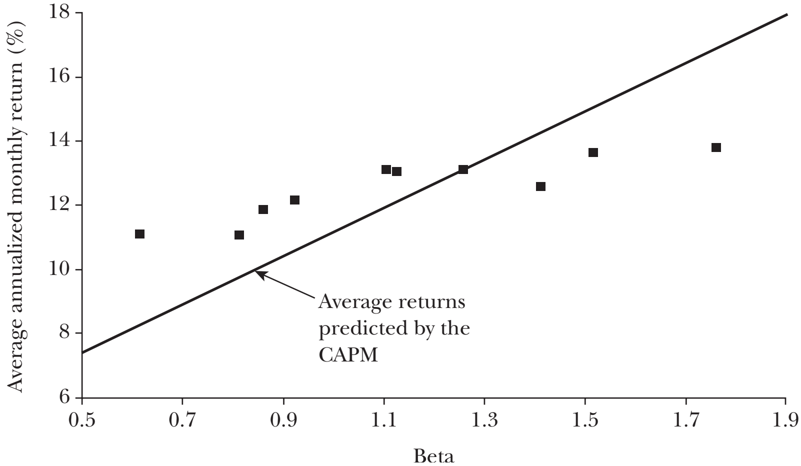
The evidence that the relation between beta and average return is too flat is confirmed in time-series tests, such as Friend and Blume (1970), Black, Jensen and Scholes (1972) and Stambaugh (1982). The intercepts in time-series regressions of excess asset returns on the excess market return are positive for assets with low betas and negative for assets with high betas.

Figure 2 provides an updated example of the evidence. In December of each year, we estimate a preranking beta for every NYSE (1928–2003), AMEX (1963–2003) and NASDAQ (1972–2003) stock in the CRSP (Center for Research in Security Prices of the University of Chicago) database, using two to five years (as available) of prior monthly returns.[5] We then form ten value-weight portfolios based on these preranking betas and compute their returns for the next twelve months. We repeat this process for each year from 1928 to 2003. The result is 912 monthly returns on ten beta-sorted portfolios. Figure 2 plots each portfolio's average return against its postranking beta, estimated by regressing its monthly returns for 1928–2003 on the return on the CRSP value-weight portfolio of U.S. common stocks.

The Sharpe-Lintner CAPM predicts that the portfolios plot along a straight

[5] To be included in the sample for year t, a security must have market equity data (price times shares outstanding) for December of $t - 1$, and CRSP must classify it as ordinary common equity. Thus, we exclude securities such as American Depository Receipts (ADRs) and Real Estate Investment Trusts (REITs).

Figure 2
Average Annualized Monthly Return versus Beta for Value Weight Portfolios Formed on Prior Beta, 1928–2003



line, with an intercept equal to the risk-free rate, R_f, and a slope equal to the expected excess return on the market, $E(R_M) - R_f$. We use the average one-month Treasury bill rate and the average excess CRSP market return for 1928–2003 to estimate the predicted line in Figure 2. Confirming earlier evidence, the relation between beta and average return for the ten portfolios is much flatter than the Sharpe-Lintner CAPM predicts. The returns on the low beta portfolios are too high, and the returns on the high beta portfolios are too low. For example, the predicted return on the portfolio with the lowest beta is 8.3 percent per year; the actual return is 11.1 percent. The predicted return on the portfolio with the highest beta is 16.8 percent per year; the actual is 13.7 percent.

Although the observed premium per unit of beta is lower than the Sharpe-Lintner model predicts, the relation between average return and beta in Figure 2 is roughly linear. This is consistent with the Black version of the CAPM, which predicts only that the beta premium is positive. Even this less restrictive model, however, eventually succumbs to the data.

Testing Whether Market Betas Explain Expected Returns

The Sharpe-Lintner and Black versions of the CAPM share the prediction that the market portfolio is mean-variance-efficient. This implies that differences in expected return across securities and portfolios are entirely explained by differences in market beta; other variables should add nothing to the explanation of expected return. This prediction plays a prominent role in tests of the CAPM. In the early work, the weapon of choice is cross-section regressions.

In the framework of Fama and MacBeth (1973), one simply adds predetermined explanatory variables to the month-by-month cross-section regressions of returns on beta. If all differences in expected return are explained by beta, the average slopes on the additional variables should not be reliably different from zero. Clearly, the trick in the cross-section regression approach is to choose specific additional variables likely to expose any problems of the CAPM prediction that, because the market portfolio is efficient, market betas suffice to explain expected asset returns.

For example, in Fama and MacBeth (1973) the additional variables are squared market betas (to test the prediction that the relation between expected return and beta is linear) and residual variances from regressions of returns on the market return (to test the prediction that market beta is the only measure of risk needed to explain expected returns). These variables do not add to the explanation of average returns provided by beta. Thus, the results of Fama and MacBeth (1973) are consistent with the hypothesis that their market proxy—an equal-weight portfolio of NYSE stocks—is on the minimum variance frontier.

The hypothesis that market betas completely explain expected returns can also be tested using time-series regressions. In the time-series regression described above (the excess return on asset i regressed on the excess market return), the intercept is the difference between the asset's average excess return and the excess return predicted by the Sharpe-Lintner model, that is, beta times the average excess market return. If the model holds, there is no way to group assets into portfolios whose intercepts are reliably different from zero. For example, the intercepts for a portfolio of stocks with high ratios of earnings to price and a portfolio of stocks with low earning-price ratios should both be zero. Thus, to test the hypothesis that market betas suffice to explain expected returns, one estimates the time-series regression for a set of assets (or portfolios) and then jointly tests the vector of regression intercepts against zero. The trick in this approach is to choose the left-hand-side assets (or portfolios) in a way likely to expose any shortcoming of the CAPM prediction that market betas suffice to explain expected asset returns.

In early applications, researchers use a variety of tests to determine whether the intercepts in a set of time-series regressions are all zero. The tests have the same asymptotic properties, but there is controversy about which has the best small sample properties. Gibbons, Ross and Shanken (1989) settle the debate by providing an F-test on the intercepts that has exact small-sample properties. They also show that the test has a simple economic interpretation. In effect, the test constructs a candidate for the tangency portfolio T in Figure 1 by optimally combining the market proxy and the left-hand-side assets of the time-series regressions. The estimator then tests whether the efficient set provided by the combination of this tangency portfolio and the risk-free asset is reliably superior to the one obtained by combining the risk-free asset with the market proxy alone. In other words, the Gibbons, Ross and Shanken statistic tests whether the market proxy is the tangency portfolio in the set of portfolios that can be constructed by combining the market portfolio with the specific assets used as dependent variables in the time-series regressions.

Enlightened by this insight of Gibbons, Ross and Shanken (1989), one can see a similar interpretation of the cross-section regression test of whether market betas suffice to explain expected returns. In this case, the test is whether the additional explanatory variables in a cross-section regression identify patterns in the returns on the left-hand-side assets that are not explained by the assets' market betas. This amounts to testing whether the market proxy is on the minimum variance frontier that can be constructed using the market proxy and the left-hand-side assets included in the tests.

An important lesson from this discussion is that time-series and cross-section regressions do not, strictly speaking, test the CAPM. What is literally tested is whether a specific proxy for the market portfolio (typically a portfolio of U.S. common stocks) is efficient in the set of portfolios that can be constructed from it and the left-hand-side assets used in the test. One might conclude from this that the CAPM has never been tested, and prospects for testing it are not good because 1) the set of left-hand-side assets does not include all marketable assets, and 2) data for the true market portfolio of all assets are likely beyond reach (Roll, 1977; more on this later). But this criticism can be leveled at tests of any economic model when the tests are less than exhaustive or when they use proxies for the variables called for by the model.

The bottom line from the early cross-section regression tests of the CAPM, such as Fama and MacBeth (1973), and the early time-series regression tests, like Gibbons (1982) and Stambaugh (1982), is that standard market proxies seem to be on the minimum variance frontier. That is, the central predictions of the Black version of the CAPM, that market betas suffice to explain expected returns and that the risk premium for beta is positive, seem to hold. But the more specific prediction of the Sharpe-Lintner CAPM that the premium per unit of beta is the expected market return minus the risk-free interest rate is consistently rejected.

The success of the Black version of the CAPM in early tests produced a consensus that the model is a good description of expected returns. These early results, coupled with the model's simplicity and intuitive appeal, pushed the CAPM to the forefront of finance.

Recent Tests

Starting in the late 1970s, empirical work appears that challenges even the Black version of the CAPM. Specifically, evidence mounts that much of the variation in expected return is unrelated to market beta.

The first blow is Basu's (1977) evidence that when common stocks are sorted on earnings-price ratios, future returns on high E/P stocks are higher than predicted by the CAPM. Banz (1981) documents a size effect: when stocks are sorted on market capitalization (price times shares outstanding), average returns on small stocks are higher than predicted by the CAPM. Bhandari (1988) finds that high debt-equity ratios (book value of debt over the market value of equity, a measure of leverage) are associated with returns that are too high relative to their market betas.

Finally, Statman (1980) and Rosenberg, Reid and Lanstein (1985) document that stocks with high book-to-market equity ratios (B/M, the ratio of the book value of a common stock to its market value) have high average returns that are not captured by their betas.

There is a theme in the contradictions of the CAPM summarized above. Ratios involving stock prices have information about expected returns missed by market betas. On reflection, this is not surprising. A stock's price depends not only on the expected cash flows it will provide, but also on the expected returns that discount expected cash flows back to the present. Thus, in principle, the cross-section of prices has information about the cross-section of expected returns. (A high expected return implies a high discount rate and a low price.) The cross-section of stock prices is, however, arbitrarily affected by differences in scale (or units). But with a judicious choice of scaling variable X, the ratio X/P can reveal differences in the cross-section of expected stock returns. Such ratios are thus prime candidates to expose shortcomings of asset pricing models—in the case of the CAPM, shortcomings of the prediction that market betas suffice to explain expected returns (Ball, 1978). The contradictions of the CAPM summarized above suggest that earnings-price, debt-equity and book-to-market ratios indeed play this role.

Fama and French (1992) update and synthesize the evidence on the empirical failures of the CAPM. Using the cross-section regression approach, they confirm that size, earnings-price, debt-equity and book-to-market ratios add to the explanation of expected stock returns provided by market beta. Fama and French (1996) reach the same conclusion using the time-series regression approach applied to portfolios of stocks sorted on price ratios. They also find that different price ratios have much the same information about expected returns. This is not surprising given that price is the common driving force in the price ratios, and the numerators are just scaling variables used to extract the information in price about expected returns.

Fama and French (1992) also confirm the evidence (Reinganum, 1981; Stambaugh, 1982; Lakonishok and Shapiro, 1986) that the relation between average return and beta for common stocks is even flatter after the sample periods used in the early empirical work on the CAPM. The estimate of the beta premium is, however, clouded by statistical uncertainty (a large standard error). Kothari, Shanken and Sloan (1995) try to resuscitate the Sharpe-Lintner CAPM by arguing that the weak relation between average return and beta is just a chance result. But the strong evidence that other variables capture variation in expected return missed by beta makes this argument irrelevant. If betas do not suffice to explain expected returns, the market portfolio is not efficient, and the CAPM is dead in its tracks. Evidence on the size of the market premium can neither save the model nor further doom it.

The synthesis of the evidence on the empirical problems of the CAPM provided by Fama and French (1992) serves as a catalyst, marking the point when it is generally acknowledged that the CAPM has potentially fatal problems. Research then turns to explanations.

One possibility is that the CAPM's problems are spurious, the result of data dredging—publication-hungry researchers scouring the data and unearthing contradictions that occur in specific samples as a result of chance. A standard response to this concern is to test for similar findings in other samples. Chan, Hamao and Lakonishok (1991) find a strong relation between book-to-market equity (B/M) and average return for Japanese stocks. Capaul, Rowley and Sharpe (1993) observe a similar B/M effect in four European stock markets and in Japan. Fama and French (1998) find that the price ratios that produce problems for the CAPM in U.S. data show up in the same way in the stock returns of twelve non-U.S. major markets, and they are present in emerging market returns. This evidence suggests that the contradictions of the CAPM associated with price ratios are not sample specific.

Explanations: Irrational Pricing or Risk

Among those who conclude that the empirical failures of the CAPM are fatal, two stories emerge. On one side are the behavioralists. Their view is based on evidence that stocks with high ratios of book value to market price are typically firms that have fallen on bad times, while low B/M is associated with growth firms (Lakonishok, Shleifer and Vishny, 1994; Fama and French, 1995). The behavioralists argue that sorting firms on book-to-market ratios exposes investor overreaction to good and bad times. Investors overextrapolate past performance, resulting in stock prices that are too high for growth (low B/M) firms and too low for distressed (high B/M, so-called value) firms. When the overreaction is eventually corrected, the result is high returns for value stocks and low returns for growth stocks. Proponents of this view include DeBondt and Thaler (1987), Lakonishok, Shleifer and Vishny (1994) and Haugen (1995).

The second story for explaining the empirical contradictions of the CAPM is that they point to the need for a more complicated asset pricing model. The CAPM is based on many unrealistic assumptions. For example, the assumption that investors care only about the mean and variance of one-period portfolio returns is extreme. It is reasonable that investors also care about how their portfolio return covaries with labor income and future investment opportunities, so a portfolio's return variance misses important dimensions of risk. If so, market beta is not a complete description of an asset's risk, and we should not be surprised to find that differences in expected return are not completely explained by differences in beta. In this view, the search should turn to asset pricing models that do a better job explaining average returns.

Merton's (1973) intertemporal capital asset pricing model (ICAPM) is a natural extension of the CAPM. The ICAPM begins with a different assumption about investor objectives. In the CAPM, investors care only about the wealth their portfolio produces at the end of the current period. In the ICAPM, investors are concerned not only with their end-of-period payoff, but also with the opportunities they will have to consume or invest the payoff. Thus, when choosing a portfolio at time $t - 1$, ICAPM investors consider how their wealth at t might vary with future *state variables*, including labor income, the prices of consumption goods and the nature of portfolio opportunities at t, and expectations about the labor income, consumption and investment opportunities to be available after t.

Like CAPM investors, ICAPM investors prefer high expected return and low return variance. But ICAPM investors are also concerned with the covariances of portfolio returns with state variables. As a result, optimal portfolios are "multifactor efficient," which means they have the largest possible expected returns, given their return variances and the covariances of their returns with the relevant state variables.

Fama (1996) shows that the ICAPM generalizes the logic of the CAPM. That is, if there is risk-free borrowing and lending or if short sales of risky assets are allowed, market clearing prices imply that the market portfolio is multifactor efficient. Moreover, multifactor efficiency implies a relation between expected return and beta risks, but it requires additional betas, along with a market beta, to explain expected returns.

An ideal implementation of the ICAPM would specify the state variables that affect expected returns. Fama and French (1993) take a more indirect approach, perhaps more in the spirit of Ross's (1976) arbitrage pricing theory. They argue that though size and book-to-market equity are not themselves state variables, the higher average returns on small stocks and high book-to-market stocks reflect unidentified state variables that produce undiversifiable risks (covariances) in returns that are not captured by the market return and are priced separately from market betas. In support of this claim, they show that the returns on the stocks of small firms covary more with one another than with returns on the stocks of large firms, and returns on high book-to-market (value) stocks covary more with one another than with returns on low book-to-market (growth) stocks. Fama and French (1995) show that there are similar size and book-to-market patterns in the covariation of fundamentals like earnings and sales.

Based on this evidence, Fama and French (1993, 1996) propose a three-factor model for expected returns,

$$\text{(Three-Factor Model)} \quad E(R_{it}) - R_{ft} = \beta_{iM}[E(R_{Mt}) - R_{ft}] + \beta_{is}E(SMB_t) + \beta_{ih}E(HML_t).$$

In this equation, SMB_t (small minus big) is the difference between the returns on diversified portfolios of small and big stocks, HML_t (high minus low) is the difference between the returns on diversified portfolios of high and low B/M stocks, and the betas are slopes in the multiple regression of $R_{it} - R_{ft}$ on $R_{Mt} - R_{ft}$, SMB_t and HML_t.

For perspective, the average value of the market premium $R_{Mt} - R_{ft}$ for 1927–2003 is 8.3 percent per year, which is 3.5 standard errors from zero. The average values of SMB_t, and HML_t are 3.6 percent and 5.0 percent per year, and they are 2.1 and 3.1 standard errors from zero. All three premiums are volatile, with annual standard deviations of 21.0 percent ($R_{Mt} - R_{ft}$), 14.6 percent (SMB_t) and 14.2 percent (HML_t) per year. Although the average values of the premiums are large, high volatility implies substantial uncertainty about the true expected premiums.

One implication of the expected return equation of the three-factor model is that the intercept α_i in the time-series regression,

$$R_{it} - R_{ft} = \alpha_i + \beta_{iM}(R_{Mt} - R_{ft}) + \beta_{is}SMB_t + \beta_{ih}HML_t + \varepsilon_{it},$$

is zero for all assets i. Using this criterion, Fama and French (1993, 1996) find that the model captures much of the variation in average return for portfolios formed on size, book-to-market equity and other price ratios that cause problems for the CAPM. Fama and French (1998) show that an international version of the model performs better than an international CAPM in describing average returns on portfolios formed on scaled price variables for stocks in 13 major markets.

The three-factor model is now widely used in empirical research that requires a model of expected returns. Estimates of α_i from the time-series regression above are used to calibrate how rapidly stock prices respond to new information (for example, Loughran and Ritter, 1995; Mitchell and Stafford, 2000). They are also used to measure the special information of portfolio managers, for example, in Carhart's (1997) study of mutual fund performance. Among practitioners like Ibbotson Associates, the model is offered as an alternative to the CAPM for estimating the cost of equity capital.

From a theoretical perspective, the main shortcoming of the three-factor model is its empirical motivation. The small-minus-big (SMB) and high-minus-low (HML) explanatory returns are not motivated by predictions about state variables of concern to investors. Instead they are brute force constructs meant to capture the patterns uncovered by previous work on how average stock returns vary with size and the book-to-market equity ratio.

But this concern is not fatal. The ICAPM does not require that the additional portfolios used along with the market portfolio to explain expected returns "mimic" the relevant state variables. In both the ICAPM and the arbitrage pricing theory, it suffices that the additional portfolios are well diversified (in the terminology of Fama, 1996, they are multifactor minimum variance) and that they are sufficiently different from the market portfolio to capture covariation in returns and variation in expected returns missed by the market portfolio. Thus, adding diversified portfolios that capture covariation in returns and variation in average returns left unexplained by the market is in the spirit of both the ICAPM and the Ross's arbitrage pricing theory.

The behavioralists are not impressed by the evidence for a risk-based explanation of the failures of the CAPM. They typically concede that the three-factor model captures covariation in returns missed by the market return and that it picks up much of the size and value effects in average returns left unexplained by the CAPM. But their view is that the average return premium associated with the model's book-to-market factor—which does the heavy lifting in the improvements to the CAPM—is itself the result of investor overreaction that happens to be correlated across firms in a way that just looks like a risk story. In short, in the behavioral view, the market tries to set CAPM prices, and violations of the CAPM are due to mispricing.

The conflict between the behavioral irrational pricing story and the rational risk story for the empirical failures of the CAPM leaves us at a timeworn impasse. Fama (1970) emphasizes that the hypothesis that prices properly reflect available information must be tested in the context of a model of expected returns, like the CAPM. Intuitively, to test whether prices are rational, one must take a stand on what the market is trying to do in setting prices—that is, what is risk and what is the relation between expected return and risk? When tests reject the CAPM, one cannot say whether the problem is its assumption that prices are rational (the behavioral view) or violations of other assumptions that are also necessary to produce the CAPM (our position).

Fortunately, for some applications, the way one uses the three-factor model does not depend on one's view about whether its average return premiums are the rational result of underlying state variable risks, the result of irrational investor behavior or sample specific results of chance. For example, when measuring the response of stock prices to new information or when evaluating the performance of managed portfolios, one wants to account for known patterns in returns and average returns for the period examined, whatever their source. Similarly, when estimating the cost of equity capital, one might be unconcerned with whether expected return premiums are rational or irrational since they are in either case part of the opportunity cost of equity capital (Stein, 1996). But the cost of capital is forward looking, so if the premiums are sample specific they are irrelevant.

The three-factor model is hardly a panacea. Its most serious problem is the momentum effect of Jegadeesh and Titman (1993). Stocks that do well relative to the market over the last three to twelve months tend to continue to do well for the next few months, and stocks that do poorly continue to do poorly. This momentum effect is distinct from the value effect captured by book-to-market equity and other price ratios. Moreover, the momentum effect is left unexplained by the three-factor model, as well as by the CAPM. Following Carhart (1997), one response is to add a momentum factor (the difference between the returns on diversified portfolios of short-term winners and losers) to the three-factor model. This step is again legitimate in applications where the goal is to abstract from known patterns in average returns to uncover information-specific or manager-specific effects. But since the momentum effect is short-lived, it is largely irrelevant for estimates of the cost of equity capital.

Another strand of research points to problems in both the three-factor model and the CAPM. Frankel and Lee (1998), Dechow, Hutton and Sloan (1999), Piotroski (2000) and others show that in portfolios formed on price ratios like book-to-market equity, stocks with higher expected cash flows have higher average returns that are not captured by the three-factor model or the CAPM. The authors interpret their results as evidence that stock prices are irrational, in the sense that they do not reflect available information about expected profitability.

In truth, however, one can't tell whether the problem is bad pricing or a bad asset pricing model. A stock's price can always be expressed as the present value of expected future cash flows discounted at the expected return on the stock (Campbell and Shiller, 1989; Vuolteenaho, 2002). It follows that if two stocks have the same price, the one with higher expected cash flows must have a higher expected return. This holds true whether pricing is rational or irrational. Thus, when one observes a positive relation between expected cash flows and expected returns that is left unexplained by the CAPM or the three-factor model, one can't tell whether it is the result of irrational pricing or a misspecified asset pricing model.

The Market Proxy Problem

Roll (1977) argues that the CAPM has never been tested and probably never will be. The problem is that the market portfolio at the heart of the model is theoretically and empirically elusive. It is not theoretically clear which assets (for example, human capital) can legitimately be excluded from the market portfolio, and data availability substantially limits the assets that are included. As a result, tests of the CAPM are forced to use proxies for the market portfolio, in effect testing whether the proxies are on the minimum variance frontier. Roll argues that because the tests use proxies, not the true market portfolio, we learn nothing about the CAPM.

We are more pragmatic. The relation between expected return and market beta of the CAPM is just the minimum variance condition that holds in any efficient portfolio, applied to the market portfolio. Thus, if we can find a market proxy that is on the minimum variance frontier, it can be used to describe differences in expected returns, and we would be happy to use it for this purpose. The strong rejections of the CAPM described above, however, say that researchers have not uncovered a reasonable market proxy that is close to the minimum variance frontier. If researchers are constrained to reasonable proxies, we doubt they ever will.

Our pessimism is fueled by several empirical results. Stambaugh (1982) tests the CAPM using a range of market portfolios that include, in addition to U.S. common stocks, corporate and government bonds, preferred stocks, real estate and other consumer durables. He finds that tests of the CAPM are not sensitive to expanding the market proxy beyond common stocks, basically because the volatility of expanded market returns is dominated by the volatility of stock returns.

One need not be convinced by Stambaugh's (1982) results since his market proxies are limited to U.S. assets. If international capital markets are open and asset prices conform to an international version of the CAPM, the market portfolio should include international assets. Fama and French (1998) find, however, that betas for a global stock market portfolio cannot explain the high average returns observed around the world on stocks with high book-to-market or high earnings-price ratios.

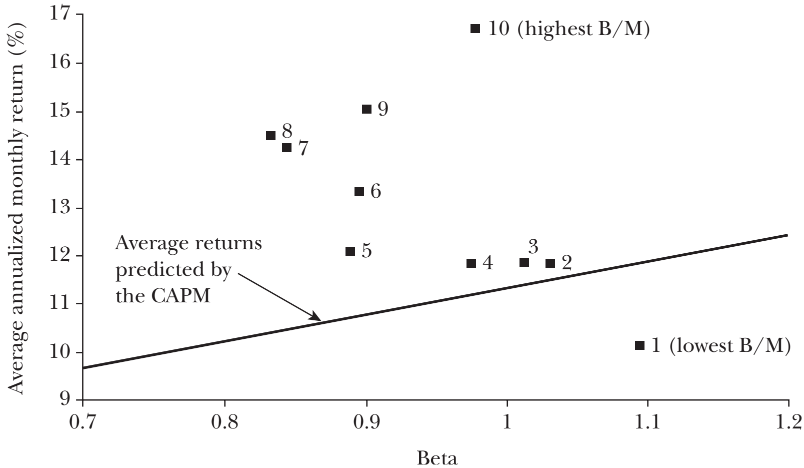
A major problem for the CAPM is that portfolios formed by sorting stocks on price ratios produce a wide range of average returns, but the average returns are not positively related to market betas (Lakonishok, Shleifer and Vishny, 1994; Fama and French, 1996, 1998). The problem is illustrated in Figure 3, which shows average returns and betas (calculated with respect to the CRSP value-weight portfolio of NYSE, AMEX and NASDAQ stocks) for July 1963 to December 2003 for ten portfolios of U.S. stocks formed annually on sorted values of the book-to-market equity ratio (B/M).[6]

Average returns on the B/M portfolios increase almost monotonically, from 10.1 percent per year for the lowest B/M group (portfolio 1) to an impressive 16.7 percent for the highest (portfolio 10). But the positive relation between beta and average return predicted by the CAPM is notably absent. For example, the portfolio with the lowest book-to-market ratio has the highest beta but the lowest average return. The estimated beta for the portfolio with the highest book-to-market ratio and the highest average return is only 0.98. With an average annualized value of the riskfree interest rate, R_f, of 5.8 percent and an average annualized market premium, $R_M - R_f$, of 11.3 percent, the Sharpe-Lintner CAPM predicts an average return of 11.8 percent for the lowest B/M portfolio and 11.2 percent for the highest, far from the observed values, 10.1 and 16.7 percent. For the Sharpe-Lintner model to "work" on these portfolios, their market betas must change dramatically, from 1.09 to 0.78 for the lowest B/M portfolio and from 0.98 to 1.98 for the highest. We judge it unlikely that alternative proxies for the market portfolio will produce betas and a market premium that can explain the average returns on these portfolios.

It is always possible that researchers will redeem the CAPM by finding a reasonable proxy for the market portfolio that is on the minimum variance frontier. We emphasize, however, that this possibility cannot be used to justify the way the CAPM is currently applied. The problem is that applications typically use the same

[6] Stock return data are from CRSP, and book equity data are from Compustat and the Moody's Industrials, Transportation, Utilities and Financials manuals. Stocks are allocated to ten portfolios at the end of June of each year t (1963 to 2003) using the ratio of book equity for the fiscal year ending in calendar year $t - 1$, divided by market equity at the end of December of $t - 1$. Book equity is the book value of stockholders' equity, plus balance sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock. Depending on availability, we use the redemption, liquidation or par value (in that order) to estimate the book value of preferred stock. Stockholders' equity is the value reported by Moody's or Compustat, if it is available. If not, we measure stockholders' equity as the book value of common equity plus the par value of preferred stock or the book value of assets minus total liabilities (in that order). The portfolios for year t include NYSE (1963–2003), AMEX (1963–2003) and NASDAQ (1972–2003) stocks with positive book equity in $t - 1$ and market equity (from CRSP) for December of $t - 1$ and June of t. The portfolios exclude securities CRSP does not classify as ordinary common equity. The breakpoints for year t use only securities that are on the NYSE in June of year t.

Figure 3
Average Annualized Monthly Return versus Beta for Value Weight Portfolios Formed on B/M, 1963–2003



market proxies, like the value-weight portfolio of U.S. stocks, that lead to rejections of the model in empirical tests. The contradictions of the CAPM observed when such proxies are used in tests of the model show up as bad estimates of expected returns in applications; for example, estimates of the cost of equity capital that are too low (relative to historical average returns) for small stocks and for stocks with high book-to-market equity ratios. In short, if a market proxy does not work in tests of the CAPM, it does not work in applications.

Conclusions

The version of the CAPM developed by Sharpe (1964) and Lintner (1965) has never been an empirical success. In the early empirical work, the Black (1972) version of the model, which can accommodate a flatter tradeoff of average return for market beta, has some success. But in the late 1970s, research begins to uncover variables like size, various price ratios and momentum that add to the explanation of average returns provided by beta. The problems are serious enough to invalidate most applications of the CAPM.

For example, finance textbooks often recommend using the Sharpe-Lintner CAPM risk-return relation to estimate the cost of equity capital. The prescription is to estimate a stock's market beta and combine it with the risk-free interest rate and the average market risk premium to produce an estimate of the cost of equity. The typical market portfolio in these exercises includes just U.S. common stocks. But empirical work, old and new, tells us that the relation between beta and average return is flatter than predicted by the Sharpe-Lintner version of the CAPM. As a result, CAPM estimates of the cost of equity for high beta stocks are too high (relative to historical average returns) and estimates for low beta stocks are too low (Friend and Blume, 1970). Similarly, if the high average returns on value stocks (with high book-to-market ratios) imply high expected returns, CAPM cost of equity estimates for such stocks are too low.[7]

The CAPM is also often used to measure the performance of mutual funds and other managed portfolios. The approach, dating to Jensen (1968), is to estimate the CAPM time-series regression for a portfolio and use the intercept (Jensen's alpha) to measure abnormal performance. The problem is that, because of the empirical failings of the CAPM, even passively managed stock portfolios produce abnormal returns if their investment strategies involve tilts toward CAPM problems (Elton, Gruber, Das and Hlavka, 1993). For example, funds that concentrate on low beta stocks, small stocks or value stocks will tend to produce positive abnormal returns relative to the predictions of the Sharpe-Lintner CAPM, even when the fund managers have no special talent for picking winners.

The CAPM, like Markowitz's (1952, 1959) portfolio model on which it is built, is nevertheless a theoretical tour de force. We continue to teach the CAPM as an introduction to the fundamental concepts of portfolio theory and asset pricing, to be built on by more complicated models like Merton's (1973) ICAPM. But we also warn students that despite its seductive simplicity, the CAPM's empirical problems probably invalidate its use in applications.

■ *We gratefully acknowledge the comments of John Cochrane, George Constantinides, Richard Leftwich, Andrei Shleifer, René Stulz and Timothy Taylor.*

[7] The problems are compounded by the large standard errors of estimates of the market premium and of betas for individual stocks, which probably suffice to make CAPM estimates of the cost of equity rather meaningless, even if the CAPM holds (Fama and French, 1997; Pastor and Stambaugh, 1999). For example, using the U.S. Treasury bill rate as the risk-free interest rate and the CRSP value-weight portfolio of publicly traded U.S. common stocks, the average value of the equity premium $R_{Mt} - R_{ft}$ for 1927–2003 is 8.3 percent per year, with a standard error of 2.4 percent. The two standard error range thus runs from 3.5 percent to 13.1 percent, which is sufficient to make most projects appear either profitable or unprofitable. This problem is, however, hardly special to the CAPM. For example, expected returns in all versions of Merton's (1973) ICAPM include a market beta and the expected market premium. Also, as noted earlier the expected values of the size and book-to-market premiums in the Fama-French three-factor model are also estimated with substantial error.

References

Ball, Ray. 1978. "Anomalies in Relationships Between Securities' Yields and Yield-Surrogates." *Journal of Financial Economics.* 6:2, pp. 103–26.

Banz, Rolf W. 1981. "The Relationship Between Return and Market Value of Common Stocks." *Journal of Financial Economics.* 9:1, pp. 3–18.

Basu, Sanjay. 1977. "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis." *Journal of Finance.* 12:3, pp. 129–56.

Bhandari, Laxmi Chand. 1988. "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence." *Journal of Finance.* 43:2, pp. 507–28.

Black, Fischer. 1972. "Capital Market Equilibrium with Restricted Borrowing." *Journal of Business.* 45:3, pp. 444–54.

Black, Fischer, Michael C. Jensen and Myron Scholes. 1972. "The Capital Asset Pricing Model: Some Empirical Tests," in *Studies in the Theory of Capital Markets.* Michael C. Jensen, ed. New York: Praeger, pp. 79–121.

Blume, Marshall. 1970. "Portfolio Theory: A Step Towards its Practical Application." *Journal of Business.* 43:2, pp. 152–74.

Blume, Marshall and Irwin Friend. 1973. "A New Look at the Capital Asset Pricing Model." *Journal of Finance.* 28:1, pp. 19–33.

Campbell, John Y. and Robert J. Shiller. 1989. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies.* 1:3, pp. 195–228.

Capaul, Carlo, Ian Rowley and William F. Sharpe. 1993. "International Value and Growth Stock Returns." *Financial Analysts Journal.* January/February, 49, pp. 27–36.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance.* 52:1, pp. 57–82.

Chan, Louis K.C., Yasushi Hamao and Josef Lakonishok. 1991. "Fundamentals and Stock Returns in Japan." *Journal of Finance.* 46:5, pp. 1739–789.

DeBondt, Werner F. M. and Richard H. Thaler. 1987. "Further Evidence on Investor Overreaction and Stock Market Seasonality." *Journal of Finance.* 42:3, pp. 557–81.

Dechow, Patricia M., Amy P. Hutton and Richard G. Sloan. 1999. "An Empirical Assessment of the Residual Income Valuation Model." *Journal of Accounting and Economics.* 26:1, pp. 1–34.

Douglas, George W. 1968. *Risk in the Equity Markets: An Empirical Appraisal of Market Efficiency.* Ann Arbor, Michigan: University Microfilms, Inc.

Elton, Edwin J., Martin J. Gruber, Sanjiv Das and Matt Hlavka. 1993. "Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios." *Review of Financial Studies.* 6:1, pp. 1–22.

Fama, Eugene F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance.* 25:2, pp. 383–417.

Fama, Eugene F. 1996. "Multifactor Portfolio Efficiency and Multifactor Asset Pricing." *Journal of Financial and Quantitative Analysis.* 31:4, pp. 441–65.

Fama, Eugene F. and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance.* 47:2, pp. 427–65.

Fama, Eugene F. and Kenneth R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics.* 33:1, pp. 3–56.

Fama, Eugene F. and Kenneth R. French. 1995. "Size and Book-to-Market Factors in Earnings and Returns." *Journal of Finance.* 50:1, pp. 131–55.

Fama, Eugene F. and Kenneth R. French. 1996. "Multifactor Explanations of Asset Pricing Anomalies." *Journal of Finance.* 51:1, pp. 55–84.

Fama, Eugene F. and Kenneth R. French. 1997. "Industry Costs of Equity." *Journal of Financial Economics.* 43:2 pp. 153–93.

Fama, Eugene F. and Kenneth R. French. 1998. "Value Versus Growth: The International Evidence." *Journal of Finance.* 53:6, pp. 1975–999.

Fama, Eugene F. and James D. MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy.* 81:3, pp. 607–36.

Frankel, Richard and Charles M.C. Lee. 1998. "Accounting Valuation, Market Expectation, and Cross-Sectional Stock Returns." *Journal of Accounting and Economics.* 25:3 pp. 283–319.

Friend, Irwin and Marshall Blume. 1970. "Measurement of Portfolio Performance under Uncertainty." *American Economic Review.* 60:4, pp. 607–36.

Gibbons, Michael R. 1982. "Multivariate Tests of Financial Models: A New Approach." *Journal of Financial Economics.* 10:1, pp. 3–27.

Gibbons, Michael R., Stephen A. Ross and Jay Shanken. 1989. "A Test of the Efficiency of a Given Portfolio." *Econometrica.* 57:5, pp. 1121–152.

Haugen, Robert. 1995. *The New Finance: The*

Case against Efficient Markets. Englewood Cliffs, N.J.: Prentice Hall.

Jegadeesh, Narasimhan and Sheridan Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance.* 48:1, pp. 65–91.

Jensen, Michael C. 1968. "The Performance of Mutual Funds in the Period 1945–1964." *Journal of Finance.* 23:2, pp. 389–416.

Kothari, S. P., Jay Shanken and Richard G. Sloan. 1995. "Another Look at the Cross-Section of Expected Stock Returns." *Journal of Finance.* 50:1, pp. 185–224.

Lakonishok, Josef and Alan C. Shapiro. 1986. Systemaitc Risk, Total Risk, and Size as Determinants of Stock Market Returns." *Journal of Banking and Finance.* 10:1, pp. 115–32.

Lakonishok, Josef, Andrei Shleifer and Robert W. Vishny. 1994. "Contrarian Investment, Extrapolation, and Risk." *Journal of Finance.* 49:5, pp. 1541–578.

Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics.* 47:1, pp. 13–37.

Loughran, Tim and Jay. R. Ritter. 1995. "The New Issues Puzzle." *Journal of Finance.* 50:1, pp. 23–51.

Markowitz, Harry. 1952. "Portfolio Selection." *Journal of Finance.* 7:1, pp. 77–99.

Markowitz, Harry. 1959. *Portfolio Selection: Efficient Diversification of Investments.* Cowles Foundation Monograph No. 16. New York: John Wiley & Sons, Inc.

Merton, Robert C. 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica.* 41:5, pp. 867–87.

Miller, Merton and Myron Scholes. 1972. "Rates of Return in Relation to Risk: A Reexamination of Some Recent Findings," in *Studies in the Theory of Capital Markets.* Michael C. Jensen, ed. New York: Praeger, pp. 47–78.

Mitchell, Mark L. and Erik Stafford. 2000. "Managerial Decisions and Long-Term Stock Price Performance." *Journal of Business.* 73:3, pp. 287–329.

Pastor, Lubos and Robert F. Stambaugh. 1999. "Costs of Equity Capital and Model Mispricing." *Journal of Finance.* 54:1, pp. 67–121.

Piotroski, Joseph D. 2000. "Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers." *Journal of Accounting Research.* 38:Supplement, pp. 1–51.

Reinganum, Marc R. 1981. "A New Empirical Perspective on the CAPM." *Journal of Financial and Quantitative Analysis.* 16:4, pp. 439–62.

Roll, Richard. 1977. "A Critique of the Asset Pricing Theory's Tests' Part I: On Past and Potential Testability of the Theory." *Journal of Financial Economics.* 4:2, pp. 129–76.

Rosenberg, Barr, Kenneth Reid and Ronald Lanstein. 1985. "Persuasive Evidence of Market Inefficiency." *Journal of Portfolio Management.* Spring, 11, pp. 9–17.

Ross, Stephen A. 1976. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory.* 13:3, pp. 341–60.

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance.* 19:3, pp. 425–42.

Stambaugh, Robert F. 1982. "On The Exclusion of Assets from Tests of the Two-Parameter Model: A Sensitivity Analysis." *Journal of Financial Economics.* 10:3, pp. 237–68.

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

Stein, Jeremy. 1996. "Rational Capital Budgeting in an Irrational World." *Journal of Business.* 69:4, pp. 429–55.

Tobin, James. 1958. "Liquidity Preference as Behavior Toward Risk." *Review of Economic Studies.* 25:2, pp. 65–86.

Vuolteenaho, Tuomo. 2002. "What Drives Firm Level Stock Returns?" *Journal of Finance.* 57:1, pp. 233–64.

Risk, Return, and Equilibrium: Empirical Tests

Eugene F. Fama and James D. MacBeth

University of Chicago

This paper tests the relationship between average return and risk for New York Stock Exchange common stocks. The theoretical basis of the tests is the "two-parameter" portfolio model and models of market equilibrium derived from the two-parameter portfolio model. We cannot reject the hypothesis of these models that the pricing of common stocks reflects the attempts of risk-averse investors to hold portfolios that are "efficient" in terms of expected value and dispersion of return. Moreover, the observed "fair game" properties of the coefficients and residuals of the risk-return regressions are consistent with an "efficient capital market"—that is, a market where prices of securities fully reflect available information.

I. Theoretical Background

In the two-parameter portfolio model of Tobin (1958), Markowitz (1959), and Fama (1965*b*), the capital market is assumed to be perfect in the sense that investors are price takers and there are neither transactions costs nor information costs. Distributions of one-period percentage returns on all assets and portfolios are assumed to be normal or to conform to some other two-parameter member of the symmetric stable class. Investors are assumed to be risk averse and to behave as if they choose among portfolios on the basis of maximum expected utility. A perfect capital market, investor risk aversion, and two-parameter return distributions imply the important "efficient set theorem": The optimal portfolio for any investor must be efficient in the sense that no other portfolio with the same or higher expected return has lower dispersion of return.[1]

Received August 24, 1971. Final version received for publication September 2, 1972.
Research supported by a grant from the National Science Foundation. The comments of Professors F. Black, L. Fisher, N. Gonedes, M. Jensen, M. Miller, R. Officer, H. Roberts, R. Roll, and M. Scholes are gratefully acknowledged. A special note of thanks is due to Black, Jensen, and Officer.

[1] Although the choice of dispersion parameter is arbitrary, the standard deviation

In the portfolio model the investor looks at individual assets only in terms of their contributions to the expected value and dispersion, or risk, of his portfolio return. With normal return distributions the risk of portfolio p is measured by the standard deviation, $\sigma(\tilde{R}_p)$, of its return, $\tilde{R}_p$,[2] and the risk of an asset for an investor who holds p is the contribution of the asset to $\sigma(\tilde{R}_p)$. If x_{ip} is the proportion of portfolio funds invested in asset i, $\sigma_{ij} = \text{cov}(\tilde{R}_i, \tilde{R}_j)$ is the covariance between the returns on assets i and j, and N is the number of assets, then

$$\sigma(\tilde{R}_p) = \sum_{i=1}^{N} x_{ip} \left[\frac{\sum_{j=1}^{N} x_{jp}\sigma_{ij}}{\sigma(\tilde{R}_p)} \right] = \sum_{i=1}^{N} x_{ip} \frac{\text{cov}(\tilde{R}_i, \tilde{R}_p)}{\sigma(\tilde{R}_p)}.$$

Thus, the contribution of asset i to $\sigma(\tilde{R}_p)$—that is, the risk of asset i in the portfolio p—is proportional to

$$\sum_{j=1}^{N} x_{jp}\sigma_{ij}/\sigma(\tilde{R}_p) = \text{cov}(\tilde{R}_i, \tilde{R}_p)/\sigma(\tilde{R}_p).$$

Note that since the weights x_{jp} vary from portfolio to portfolio, the risk of an asset is different for different portfolios.

For an individual investor the relationship between the risk of an asset and its expected return is implied by the fact that the investor's optimal portfolio is efficient. Thus, if he chooses the portfolio m, the fact that m is efficient means that the weights x_{im}, $i = 1, 2, \ldots, N$, maximize expected portfolio return

$$E(\tilde{R}_m) = \sum_{i=1}^{N} x_{im} E(\tilde{R}_i),$$

subject to the constraints

is common when return distributions are assumed to be normal, whereas an inter-fractile range is usually suggested when returns are generated from some other symmetric stable distribution.

It is well known that the mean–standard deviation version of the two-parameter portfolio model can be derived from the assumption that investors have quadratic utility functions. But the problems with this approach are also well known. In any case, the empirical evidence of Fama (1965*a*), Blume (1970), Roll (1970), K. Miller (1971), and Officer (1971) provides support for the "distribution" approach to the model. For a discussion of the issues and a detailed treatment of the two-parameter model, see Fama and Miller (1972, chaps. 6–8).

We also concentrate on the special case of the two-parameter model obtained with the assumption of normally distributed returns. As shown in Fama (1971) or Fama and Miller (1972, chap. 7), the important testable implications of the general symmetric stable model are the same as those of the normal model.

[2] Tildes (~) are used to denote random variables. And the one-period percentage return is most often referred to as just as the return.

$$\sigma(\tilde{R}_p) = \sigma(\tilde{R}_m) \quad \text{and} \quad \sum_{i=1}^{N} x_{im} = 1.$$

Lagrangian methods can then be used to show that the weights x_{jm} must be chosen in such a way that for any asset i in m

$$E(\tilde{R}_i) - E(\tilde{R}_m) = S_m \left[\frac{\sum_{j=1}^{N} x_{jm}\sigma_{ij}}{\sigma(\tilde{R}_m)} - \sigma(\tilde{R}_m) \right], \qquad (1)$$

where S_m is the rate of change of $E(\tilde{R}_p)$ with respect to a change in $\sigma(\tilde{R}_p)$ at the point on the efficient set corresponding to portfolio m. If there are nonnegativity constraints on the weights (that is, if short selling is prohibited), then (1) only holds for assets i such that $x_{im} > 0$.

Although equation (1) is just a condition on the weights x_{jm} that is required for portfolio efficiency, it can be interpreted as the relationship between the risk of asset i in portfolio m and the expected return on the asset. The equation says that the difference between the expected return on the asset and the expected return on the portfolio is proportional to the difference between the risk of the asset and the risk of the portfolio. The proportionality factor is S_m, the slope of the efficient set at the point corresponding to the portfolio m. And the risk of the asset is its contribution to total portfolio risk, $\sigma(\tilde{R}_m)$.

II. Testable Implications

Suppose now that we posit a market of risk-averse investors who make portfolio decisions period by period according to the two-parameter model.[3] We are concerned with determining what this implies for observable properties of security and portfolio returns. We consider two categories of implications. First, there are conditions on expected returns that are implied by the fact that in a two-parameter world investors hold efficient portfolios. Second, there are conditions on the behavior of returns through time that are implied by the assumption of the two-parameter model that the capital market is perfect or frictionless in the sense that there are neither transactions costs nor information costs.

A. *Expected Returns*

The implications of the two-parameter model for expected returns derive from the efficiency condition or expected return-risk relationship of equation (1). First, it is convenient to rewrite (1) as

[3] A multiperiod version of the two-parameter model is in Fama (1970*a*) or Fama and Miller (1972, chap. 8).

$$E(\tilde{R}_i) = [E(\tilde{R}_m) - S_m\, \sigma(\tilde{R}_m)] + S_m\, \sigma(\tilde{R}_m)\beta_i, \tag{2}$$

where

$$\beta_i \equiv \frac{\text{cov}(\tilde{R}_i, \tilde{R}_m)}{\sigma^2(\tilde{R}_m)} = \frac{\sum_{j=1}^{N} x_{jm}\sigma_{ij}}{\sigma^2(\tilde{R}_m)} = \frac{\text{cov}(\tilde{R}_i, \tilde{R}_m)/\sigma(\tilde{R}_m)}{\sigma(\tilde{R}_m)}. \tag{3}$$

The parameter β_i can be interpreted as the risk of asset i in the portfolio m, measured relative to $\sigma(\tilde{R}_m)$, the total risk of m. The intercept in (2),

$$E(\tilde{R}_0) \equiv E(\tilde{R}_m) - S_m\, \sigma(\tilde{R}_m), \tag{4}$$

is the expected return on a security whose return is uncorrelated with $\tilde{R}_m$—that is, a zero-β security. Since $\beta = 0$ implies that a security contributes nothing to $\sigma(\tilde{R}_m)$, it is appropriate to say that it is riskless in this portfolio. It is well to note from (3), however, that since $x_{im}\, \sigma_{ii} = x_{im}\, \sigma^2(\tilde{R}_i)$ is just one of the N terms in β_i, $\beta_i = 0$ does not imply that security i has zero variance of return.

From (4), it follows that

$$S_m = \frac{E(\tilde{R}_m) - E(\tilde{R}_0)}{\sigma(\tilde{R}_m)}, \tag{5}$$

so that (2) can be rewritten

$$E(\tilde{R}_i) = E(\tilde{R}_0) + [E(\tilde{R}_m) - E(\tilde{R}_0)]\beta_i. \tag{6}$$

In words, the expected return on security i is $E(\tilde{R}_0)$, the expected return on a security that is riskless in the portfolio m, plus a risk premium that is β_i times the difference between $E(\tilde{R}_m)$ and $E(\tilde{R}_0)$.

Equation (6) has three testable implications: (C1) The relationship between the expected return on a security and its risk in any efficient portfolio m is linear. (C2) β_i is a complete measure of the risk of security i in the efficient portfolio m; no other measure of the risk of i appears in (6). (C3) In a market of risk-averse investors, higher risk should be associated with higher expected return; that is, $E(\tilde{R}_m) - E(\tilde{R}_0) > 0$.

The importance of condition C3 is obvious. The importance of C1 and C2 should become clear as the discussion proceeds. At this point suffice it to say that if C1 and C2 do not hold, market returns do not reflect the attempts of investors to hold efficient portfolios: Some assets are systematically underpriced or overpriced relative to what is implied by the expected return-risk or efficiency equation (6).

B. *Market Equilibrium and the Efficiency of the Market Portfolio*

To test conditions C1–C3 we must identify some efficient portfolio m. This in turn requires specification of the characteristic of market equilibrium when investors make portfolio decisions according to the two-parameter model.

Assume again that the capital market is perfect. In addition, suppose that from the information available without cost all investors derive the same and correct assessment of the distribution of the future value of any asset or portfolio—an assumption usually called "homogeneous expectations." Finally, assume that short selling of all assets is allowed. Then Black (1972) has shown that in a market equilibrium, the so-called market portfolio, defined by the weights

$$x_{im} = \frac{\text{total market value of all units of asset } i}{\text{total market value of all assets}},$$

is always efficient.

Since it contains all assets in positive amounts, the market portfolio is a convenient reference point for testing the expected return-risk conditions C1–C3 of the two-parameter model. And the homogeneous-expectations assumption implies a correspondence between ex ante assessments of return distributions and distributions of ex post returns that is also required for meaningful tests of these three hypotheses.

C. *A Stochastic Model for Returns*

Equation (6) is in terms of expected returns. But its implications must be tested with data on period-by-period security and portfolio returns. We wish to choose a model of period-by-period returns that allows us to use observed average returns to test the expected-return conditions C1–C3, but one that is nevertheless as general as possible. We suggest the following stochastic generalization of (6):

$$\tilde{R}_{it} = \tilde{\gamma}_{0t} + \tilde{\gamma}_{1t}\beta_i + \tilde{\gamma}_{2t}\beta_i^2 + \tilde{\gamma}_{3t}s_i + \tilde{\eta}_{it}. \tag{7}$$

The subscript t refers to period t, so that $\tilde{R}_{it}$ is the one-period percentage return on security i from $t-1$ to t. Equation (7) allows $\tilde{\gamma}_{0t}$ and $\tilde{\gamma}_{1t}$ to vary stochastically from period to period. The hypothesis of condition C3 is that the expected value of the risk premium $\tilde{\gamma}_{1t}$, which is the slope $[E(\tilde{R}_{mt}) - E(\tilde{R}_{0t})]$ in (6), is positive—that is, $E(\tilde{\gamma}_{1t}) = E(\tilde{R}_{mt}) - E(\tilde{R}_{0t}) > 0$.

The variable β_i^2 is included in (7) to test linearity. The hypothesis of condition C1 is $E(\tilde{\gamma}_{2t}) = 0$, although $\tilde{\gamma}_{2t}$ is also allowed to vary stochastically from period to period. Similar statements apply to the term involving s_i in (7), which is meant to be some measure of the risk of security i that is not deterministically related to β_i. The hypothesis of condition C2 is $E(\tilde{\gamma}_{3t}) = 0$, but $\tilde{\gamma}_{3t}$ can vary stochastically through time.

The disturbance $\tilde{\eta}_{it}$ is assumed to have zero mean and to be independent of all other variables in (7). If all portfolio return distributions are to be normal (or symmetric stable), then the variables $\tilde{\eta}_{it}$, $\tilde{\gamma}_{0t}$, $\tilde{\gamma}_{1t}$, $\tilde{\gamma}_{2t}$ and $\tilde{\gamma}_{3t}$ must have a multivariate normal (or symmetric stable) distribution.

D. Capital Market Efficiency: The Behavior of Returns through Time

C1–C3 are conditions on expected returns and risk that are implied by the two-parameter model. But the model, and especially the underlying assumption of a perfect market, implies a capital market that is efficient in the sense that prices at every point in time fully reflect available information. This use of the word efficient is, of course, not to be confused with portfolio efficiency. The terminology, if a bit unfortunate, is at least standard.

Market efficiency in combination with condition C1 requires that scrutiny of the time series of the stochastic nonlinearity coefficient $\tilde{\gamma}_{2t}$ does not lead to nonzero estimates of expected future values of $\tilde{\gamma}_{2t}$. Formally, $\tilde{\gamma}_{2t}$ must be a fair game. In practical terms, although nonlinearities are observed ex post, because $\tilde{\gamma}_{2t}$ is a fair game, it is always appropriate for the investor to act ex ante under the presumption that the two-parameter model, as summarized by (6), is valid. That is, in his portfolio decisions he always assumes that there is a linear relationship between the risk of a security and its expected return. Likewise, market efficiency in the two-parameter model requires that the non-β risk coefficient $\tilde{\gamma}_{3t}$ and the time series of return disturbances $\tilde{\eta}_{it}$ are fair games. And the fair-game hypothesis also applies to the time series of $\tilde{\gamma}_{1t} - [E(\tilde{R}_{mt}) - E(\tilde{R}_{0t})]$, the difference between the risk premium for period t and its expected value.

In the terminology of Fama (1970*b*), these are "weak-form" propositions about capital market efficiency for a market where expected returns are generated by the two-parameter model. The propositions are weak since they are only concerned with whether prices fully reflect any information in the time series of past returns. "Strong-form" tests would be concerned with the speed-of-adjustment of prices to all available information.

E. Market Equilibrium with Riskless Borrowing and Lending

We have as yet presented no hypothesis about $\tilde{\gamma}_{0t}$ in (7). In the general two-parameter model, given $E(\tilde{\gamma}_{2t}) = E(\tilde{\gamma}_{3t}) = E(\tilde{\eta}_{it}) = 0$, then, from (6), $E(\tilde{\gamma}_{0t})$ is just $E(\tilde{R}_{0t})$, the expected return on any zero-β security. And market efficiency requires that $\tilde{\gamma}_{0t} - E(\tilde{R}_{0t})$ be a fair game.

But if we add to the model as presented thus far the assumption that there is unrestricted riskless borrowing and lending at the known rate R_{ft}, then one has the market setting of the original two-parameter "capital asset pricing model" of Sharpe (1964) and Lintner (1965). In this world, since $\beta_f = 0$, $E(\tilde{\gamma}_{0t}) = R_{ft}$. And market efficiency requires that $\tilde{\gamma}_{0t} - R_{ft}$ be a fair game.

It is well to emphasize that to refute the proposition that $E(\tilde{\gamma}_{0t}) = R_{ft}$ is only to refute a specific two-parameter model of market equilibrium. Our view is that tests of conditions C1–C3 are more fundamental. We regard C1–C3 as the general expected return implications of the two-parameter model in the sense that they are the implications of the fact that in the two-parameter portfolio model investors hold efficient portfolios, and they are consistent with any two-parameter model of market equilibrium in which the market portfolio is efficient.

F. The Hypotheses

To summarize, given the stochastic generalization of (2) and (6) that is provided by (7), the testable implications of the two-parameter model for expected returns are:

C1 (linearity)—$E(\tilde{\gamma}_{2t}) = 0$.
C2 (no systematic effects of non-β risk)—$E(\tilde{\gamma}_{3t}) = 0$.
C3 (positive expected return-risk tradeoff)—$E(\tilde{\gamma}_{1t}) = E(\tilde{R}_{mt}) - E(\tilde{R}_{0t}) > 0$.
Sharpe-Lintner (S-L) Hypothesis—$E(\tilde{\gamma}_{0t}) = R_{ft}$.

Finally, capital market efficiency in a two-parameter world requires
ME (market efficiency)—the stochastic coefficients $\tilde{\gamma}_{2t}$, $\tilde{\gamma}_{3t}$, $\tilde{\gamma}_{1t} - [E(\tilde{R}_{mt}) - E(\tilde{R}_{0t})]$, $\tilde{\gamma}_{0t} - E(\tilde{R}_{0t})$, and the disturbances $\tilde{\eta}_{it}$ are fair games.[4]

III. Previous Work[5]

The earliest tests of the two-parameter model were done by Douglas (1969), whose results seem to refute condition C2. In annual and quarterly return data, there seem to be measures of risk, in addition to β, that contribute systematically to observed average returns. These results, if valid, are inconsistent with the hypothesis that investors attempt to hold efficient portfolios. Assuming that the market portfolio is efficient, premiums are paid for risks that do not contribute to the risk of an efficient portfolio.

Miller and Scholes (1972) take issue both with Douglas's statistical techniques and with his use of annual and quarterly data. Using different methods and simulations, they show that Douglas's negative results could be expected even if condition C2 holds. Condition C2 is tested below with extensive monthly data, and this avoids almost all of the problems discussed by Miller and Scholes.

[4] If $\tilde{\gamma}_{2t}$ and $\tilde{\gamma}_{3t}$ are fair games, then $E(\tilde{\gamma}_{2t}) = E(\tilde{\gamma}_{3t}) = 0$. Thus, C1 and C2 are implied by ME. Keeping the expected return conditions separate, however, better emphasizes the economic basis of the various hypotheses.

[5] A comprehensive survey of empirical and theoretical work on the two-parameter model is in Jensen (1972).

Much of the available empirical work on the two-parameter model is concerned with testing the S-L hypothesis that $E(\tilde{\gamma}_{0t}) = R_{ft}$. The tests of Friend and Blume (1970) and those of Black, Jensen, and Scholes (1972) indicate that, at least in the period since 1940, on average $\tilde{\gamma}_{0t}$ is systematically greater than R_{ft}. The results below support this conclusion.

In the empirical literature to date, the importance of the linearity condition C1 has been largely overlooked. Assuming that the market portfolio m is efficient, if $E(\tilde{\gamma}_{2t})$ in (7) is positive, the prices of high-β securities are on average too low—their expected returns are too high—relative to those of low-β securities, while the reverse holds if $E(\tilde{\gamma}_{2t})$ is negative. In short, if the process of price formation in the capital market reflects the attempts of investors to hold efficient portfolios, then the linear relationship of (6) between expected return and risk must hold.

Finally, the previous empirical work on the two-parameter model has not been concerned with tests of market efficiency.

IV. Methodology

The data for this study are monthly percentage returns (including dividends and capital gains, with the appropriate adjustments for capital changes such as splits and stock dividends) for all common stocks traded on the New York Stock Exchange during the period January 1926 through June 1968. The data are from the Center for Research in Security Prices of the University of Chicago.

A. General Approach

Testing the two-parameter model immediately presents an unavoidable "errors-in-the-variables" problem: The efficiency condition or expected return-risk equation (6) is in terms of true values of the relative risk measure β_i, but in empirical tests estimates, $\hat{\beta}_i$, must be used. In this paper

$$\hat{\beta}_i = \frac{\widehat{\operatorname{cov}}(\tilde{R}_i, \tilde{R}_m)}{\hat{\sigma}^2(\tilde{R}_m)},$$

where $\widehat{\operatorname{cov}}(\tilde{R}_i, \tilde{R}_m)$ and $\hat{\sigma}^2(\tilde{R}_m)$ are estimates of $\operatorname{cov}(\tilde{R}_i, \tilde{R}_m)$ and $\sigma^2(\tilde{R}_m)$ obtained from monthly returns, and where the proxy chosen for $\tilde{R}_{mt}$ is "Fisher's Arithmetic Index," an equally weighted average of the returns on all stocks listed on the New York Stock Exchange in month t. The properties of this index are analyzed in Fisher (1966).

Blume (1970) shows that for any portfolio p, defined by the weights x_{ip}, $i = 1, 2, \ldots, N$,

$$\hat{\beta}_p = \frac{\widehat{\operatorname{cov}}(\tilde{R}_p, \tilde{R}_m)}{\hat{\sigma}^2(\tilde{R}_m)} = \sum_{i=1}^{N} x_{ip} \frac{\widehat{\operatorname{cov}}(\tilde{R}_i, \tilde{R}_m)}{\hat{\sigma}^2(\tilde{R}_m)} = \sum_{i=1}^{N} x_{ip} \hat{\beta}_i.$$

If the errors in the $\hat{\beta}_i$ are substantially less than perfectly positively correlated, the $\hat{\beta}$'s of portfolios can be much more precise estimates of true β's than the $\hat{\beta}$'s for individual securities.

To reduce the loss of information in the risk-return tests caused by using portfolios rather than individual securities, a wide range of values of portfolio $\hat{\beta}_p$'s is obtained by forming portfolios on the basis of ranked values of $\hat{\beta}_i$ for individual securities. But such a procedure, naïvely executed could result in a serious regression phenomenon. In a cross section of $\hat{\beta}_i$, high observed $\hat{\beta}_i$ tend to be above the corresponding true β_i and low observed $\hat{\beta}_i$ tend to be below the true β_i. Forming portfolios on the basis of ranked $\hat{\beta}_i$ thus causes bunching of positive and negative sampling errors within portfolios. The result is that a large portfolio $\hat{\beta}_p$ would tend to overstate the true β_p, while a low $\hat{\beta}_p$ would tend to be an underestimate.

The regression phenomenon can be avoided to a large extent by forming portfolios from ranked $\hat{\beta}_i$ computed from data for one time period but then using a subsequent period to obtain the $\hat{\beta}_p$ for these portfolios that are used to test the two-parameter model. With fresh data, within a portfolio errors in the individual security $\hat{\beta}_i$ are to a large extent random across securities, so that in a portfolio $\hat{\beta}_p$ the effects of the regression phenomenon are, it is hoped, minimized.[6]

B. *Details*

The specifics of the approach are as follows. Let N be the total number of securities to be allocated to portfolios and let $\text{int}(N/20)$ be the largest integer equal to or less than $N/20$. Using the first 4 years (1926–29) of monthly return data, 20 portfolios are formed on the basis of ranked $\hat{\beta}_i$ for individual securities. The middle 18 portfolios each has $\text{int}(N/20)$ securities. If N is even, the first and last portfolios each has $\text{int}(N/20) + \frac{1}{2}[N - 20\,\text{int}(N/20)]$ securities. The last (highest $\hat{\beta}$) portfolio gets an additional security if N is odd.

The following 5 years (1930–34) of data are then used to recompute the $\hat{\beta}_i$, and these are averaged across securities within portfolios to obtain 20 initial portfolio $\hat{\beta}_{pt}$ for the risk-return tests. The subscript t is added to indicate that each month t of the following four years (1935–38) these $\hat{\beta}_{pt}$ are recomputed as simple averages of individual security $\hat{\beta}_i$, thus adjusting the portfolio $\hat{\beta}_{pt}$ month by month to allow for delisting of securities. The component $\hat{\beta}_i$ for securities are themselves updated yearly—that

[6] The errors-in-the-variables problem and the technique of using portfolios to solve it were first pointed out by Blume (1970). The portfolio approach is also used by Friend and Blume (1970) and Black, Jensen, and Scholes (1972). The regression phenomenon that arises in risk-return tests was first recognized by Blume (1970) and then by Black, Jensen, and Scholes (1972), who offer a solution to the problem that is similar in spirit to ours.

is, they are recomputed from monthly returns for 1930 through 1935, 1936, or 1937.

As a measure of the non-β risk of security i we use $s(\hat{\epsilon}_i)$, the standard deviation of the least-squares residuals $\hat{\epsilon}_{it}$ from the so-called market model

$$\tilde{R}_{it} = a_i + \beta^i \tilde{R}_{mt} + \tilde{\epsilon}_{it}. \tag{8}$$

The standard deviation $s(\hat{\epsilon}_i)$ is a measure of non-β risk in the following sense. One view of risk, antithetic to that of portfolio theory, says that the risk of a security is measured by the total dispersion of its return distribution. Given a market dominated by risk averters, this model would predict that a security's expected return is related to its total return dispersion rather than just to the contribution of the security to the dispersion in the return on an efficient portfolio.[7] If $B_i \equiv \text{cov}\,(\tilde{R}_i, \tilde{R}_m)/\sigma^2(\tilde{R}_m)$, then in (8) $\text{cov}(\tilde{\epsilon}_i, \tilde{R}_m) = 0$, and

$$\sigma^2(\tilde{R}_i) = \beta_i^2\sigma^2(\tilde{R}_m) + \sigma^2(\tilde{\epsilon}_i) + 2\beta_i\,\text{cov}(\tilde{R}_m, \tilde{\epsilon}_i). \tag{9}$$

Thus, from (9), one can say that $s(\hat{\epsilon}_i)$ is an estimate of that part of the dispersion of the distribution of the return on security i that is not directly related to β_i.

The month-by-month returns on the 20 portfolios, with equal weighting of individual securities each month, are also computed for the 4-year period 1935–38. For each month t of this period, the following cross-sectional regression—the empirical analog of equation (7)—is run:

$$R_{pt} = \hat{\gamma}_{0t} + \hat{\gamma}_{1t}\,\hat{\beta}_{p,t-1} + \hat{\gamma}_{2t}\,\hat{\beta}^2{}_{p,t-1} + \hat{\gamma}_{3t}\bar{s}_{p,t-1}(\hat{\epsilon}_i) + \hat{\eta}_{pt}, \tag{10}$$

$$p = 1, 2, \ldots, 20.$$

The independent variable $\hat{\beta}_{p,t-1}$ is the average of the $\hat{\beta}_i$ for securities in portfolio p discussed above; $\hat{\beta}^2{}_{p,t-1}$ is the average of the squared values of these $\hat{\beta}_i$ (and is thus somewhat mislabeled); and $\bar{s}_{p,t-1}(\hat{\epsilon}_i)$ is likewise the average of $s(\hat{\epsilon}_i)$ for securities in portfolio p. The $s(\hat{\epsilon}_i)$ are computed from data for the same period as the component $\hat{\beta}_i$ of $\hat{\beta}_{p,t-1}$, and like these $\hat{\beta}_i$, they are updated annually.

The regression equation (10) is (7) averaged across the securities in a portfolio, with estimates $\hat{\beta}_{p,t-1}$, $\hat{\beta}^2{}_{p,t-1}$, and $\bar{s}_{p,t-1}(\hat{\epsilon}_i)$ used as explanatory variables, and with least-squares estimates of the stochastic coefficients $\hat{\gamma}_{0t}$, $\hat{\gamma}_{1t}$, $\hat{\gamma}_{2t}$, and $\hat{\gamma}_{3t}$. The results from (10)—the time series of month-by-month values of the regression coefficients $\hat{\gamma}_{0t}$, $\hat{\gamma}_{1t}$, $\hat{\gamma}_{2t}$, and $\hat{\gamma}_{3t}$ for the 4-year period 1935–38—are the inputs for our tests of the two-parameter model for this period. To get results for other periods, the steps described

[7] For those accustomed to the portfolio viewpoint, this alternative model may seem so naïve that it should be classified as a straw man. But it is the model of risk and return implied by the "liquidity preference" and "market segmentation" theories of the term structure of interest rates and by the Keynesian "normal backwardation" theory of commodity futures markets. For a discussion of the issues with respect to these markets, see Roll (1970) and K. Miller (1971).

above are repeated. That is, 7 years of data are used to form portfolios; the next 5 years are used to compute initial values of the independent variables in (10); and then the risk-return regressions of (10) are fit month by month for the following 4-year period.

The nine different portfolio formation periods (all except the first 7 years in length), initial 5-year estimation periods, and testing periods (all but the last 4 years in length) are shown in table 1. The choice of 4-year testing periods is a balance of computation costs against the desire to reform portfolios frequently. The choice of 7-year portfolio formation periods and 5–8-year periods for estimating the independent variables $\hat{\beta}_{p,t-1}$ and $\bar{s}_{p,t-1}(\hat{\epsilon}_i)$ in the risk-return regressions reflects a desire to balance the statistical power obtained with a large sample from a stationary process against potential problems caused by any nonconstancy of the β_i. The choices here are in line with the results of Gonedes (1973). His results also led us to require that to be included in a portfolio a security available in the first month of a testing period must also have data for all 5 years of the preceding estimation period and for at least 4 years of the portfolio formation period. The total number of securities available in the first month of each testing period and the number of securities meeting the data requirement are shown in table 1.

C. *Some Observations on the Approach*

Table 2 shows the values of the 20 portfolios $\hat{\beta}_{p,t-1}$ and their standard errors $s(\hat{\beta}_{p,t-1})$ for four of the nine 5-year estimation periods. Also shown are: $r(R_p, R_m)^2$, the coefficient of determination between R_{pt} and R_{mt}; $s(R_p)$, the sample standard deviation of R_p; and $s(\hat{\epsilon}_p)$, the standard deviation of the portfolio residuals from the market model of (8), not to be confused with $\bar{s}_{p,t-1}(\hat{\epsilon}_i)$, the average for individual securities, which is also shown. The $\hat{\beta}_{p,t-1}$ and $\bar{s}_{p,t-1}(\hat{\epsilon}_i)$ are the independent variables in the risk return regressions of (10) for the first month of the 4-year testing periods following the four estimation periods shown.

Under the assumptions that for a given security the disturbances $\tilde{\epsilon}_{jt}$ in (8) are serially independent, independent of $\tilde{R}_{mt}$, and identically distributed through time, the standard error of $\hat{\beta}_i$ is

$$\sigma(\hat{\beta}_i) = \frac{\sigma(\tilde{\epsilon}_i)}{\sqrt{n}\,\sigma(\tilde{R}_m)},$$

where n is the number of months used to compute $\hat{\beta}_i$. Likewise,

$$\sigma(\tilde{\beta}_{p,t-1}) = \frac{\sigma(\tilde{\epsilon}_p)}{\sqrt{n}\,\sigma(\tilde{R}_m)}.$$

Thus, the fact that in table 2, $s(\hat{\epsilon}_p)$ is generally on the order of one-third to one-seventh $\bar{s}_{p,t-1}(\hat{\epsilon}_i)$ implies that $s(\hat{\beta}_{p,t-1})$ is one-third to one-seventh

TABLE 1

PORTFOLIO FORMATION, ESTIMATION, AND TESTING PERIODS

	Periods				
	1	2	3	4	5
Portfolio formation period ...	1926–29	1927–33	1931–37	1935–41	1939–45
Initial estimation period	1930–34	1934–38	1938–42	1942–46	1946–50
Testing period	1935–38	1939–42	1943–46	1947–50	1951–54
No. of securities available	710	779	804	908	1,011
No. of securities meeting data requirement	435	576	607	704	751

$s(\hat{\beta}_i)$. Estimates of β for portfolios are indeed more precise than those for individual securities.

Nevertheless, it is interesting to note that if the disturbances $\tilde{\epsilon}_{jt}$ in (8) were independent from security to security, the relative increase in the precision of the $\hat{\beta}$ obtained by using portfolios rather than individual securities would be about the same for all portfolios. We argue in the Appendix, however, that the results from (10) imply that the $\tilde{\epsilon}_{it}$ in (8) are interdependent, and the interdependence is strongest among high-β securities and among low-β securities. This is evident in table 2: The ratios $s(\hat{\epsilon}_p)/\bar{s}_{p,t-1}(\hat{\epsilon}_i)$ are always highest at the extremes of the $\hat{\beta}_{p,t-1}$ range and lowest for $\hat{\beta}_{p,t-1}$ close to 1.0. But it is important to emphasize that since these ratios are generally less than .33, interdependence among the $\tilde{\epsilon}_{it}$ of different securities does not destroy the value of using portfolios to reduce the dispersion of the errors in estimated β's.

Finally, all the tests of the two-parameter model are predictive in the sense that the explanatory variables $\hat{\beta}_{p,t-1}$ and $\bar{s}_{p,t-1}(\hat{\epsilon}_i)$ in (10) are computed from data for a period prior to the month of the returns, the R_{pt}, on which the regression is run. Although we are interested in testing the two-parameter model as a positive theory—that is, examining the extent to which it is helpful in describing actual return data—the model was initially developed by Markowitz (1959) as a normative theory—that is, as a model to help people make better decisions. As a normative theory the model only has content if there is some relationship between future returns and estimates of risk that can be made on the basis of current information.

Now that the predictive nature of the tests has been emphasized, to simplify the notation, the explanatory variables in (10) are henceforth referred to as $\hat{\beta}_p$, $\hat{\beta}_p^2$, and $\bar{s}_p(\hat{\epsilon}_i)$.

V. Results

The major tests of the implications of the two-parameter model are in table 3. Results are presented for 10 periods: the overall period 1935–

TABLE 1 (*Continued*)

	Periods			
	6	7	8	9
Portfolio formation period ...	1943–49	1947–53	1951–57	1955–61
Initial estimation period	1950–54	1954–58	1958–62	1962–66
Testing period	1955–58	1959–62	1963–66	1967–68
No. of securities available	1,053	1,065	1,162	1,261
No. of securities meeting data requirement	802	856	858	845

6/68; three long subperiods, 1935–45, 1946–55, and 1956–6/68; and six subperiods which, except for the first and last, cover 5 years each. This choice of subperiods reflects the desire to keep separate the pre– and post–World War II periods. Results are presented for four different versions of the risk-return regression equation (10): Panel D is based on (10) itself, but in panels A–C, one or more of the variables in (10) is suppressed. For each period and model, the table shows: $\bar{\hat{\gamma}}_j$, the average of the month-by-month regression coefficient estimates, $\hat{\gamma}_{jt}$; $s(\hat{\gamma}_j)$, the standard deviation of the monthly estimates; and $\bar{r}^2$ and $s(r^2)$, the mean and standard deviation of the month-by-month coefficients of determination, r_t^2, which are adjusted for degrees of freedom. The table also shows the first-order serial correlations of the various monthly $\hat{\gamma}_{jt}$ computed either about the sample mean of $\hat{\gamma}_{jt}$ [in which case the serial correlations are labeled $\rho_M(\hat{\gamma}_i)$] or about an assumed mean of zero [in which case they are labeled $\rho_0(\hat{\gamma}_j)$]. Finally, t-statistics for testing the hypothesis that $\bar{\hat{\gamma}}_j = 0$ are presented. These t-statistics are

$$t(\bar{\hat{\gamma}}_j) = \frac{\bar{\hat{\gamma}}_j}{s(\hat{\gamma}_j)/\sqrt{n}},$$

where n is the number of months in the period, which is also the number of estimates $\hat{\gamma}_{jt}$ used to compute $\bar{\hat{\gamma}}_j$ and $s(\hat{\gamma}_j)$.

In interpreting these t-statistics one should keep in mind the evidence of Fama (1965*a*) and Blume (1970) which suggests that distributions of common stock returns are "thick-tailed" relative to the normal distribution and probably conform better to nonnormal symmetric stable distributions than to the normal. From Fama and Babiak (1968), this evidence means that when one interprets large t-statistics under the assumption that the underlying variables are normal, the probability or significance levels obtained are likely to be overestimates. But it is important to note that, with the exception of condition C3 (positive expected return-risk tradeoff), upward-biased probability levels lead to biases toward rejection of the hypotheses of the two-parameter model. Thus, if these hypotheses cannot

110138-OPC-POD-60-115

TABLE 2

SAMPLE STATISTICS FOR FOUR SELECTED ESTIMATION PERIODS

Statistic	1	2	3	4	5	6	7	8	9	10
	Portfolios for Estimation Period 1934–38									
$\hat{\beta}_{p,t-1}$	.322	.508	.651	.674	.695	.792	.921	.942	.970	1.005
$s(\hat{\beta}_{p,t-1})$	.027	.027	.025	.023	.028	.026	.032	.029	.034	.027
$r(R_p, R_m)^2$	.709	.861	.921	.936	.912	.941	.932	.946	.933	.958
$s(R_p)$	.040	.058	.072	.074	.077	.087	.101	.103	.106	.109
$s(\hat{\epsilon}_p)$	.022	.022	.020	.019	.023	.021	.026	.024	.028	.022
$\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.085	.075	.083	.078	.090	.095	.109	.106	.111	.097
$s(\hat{\epsilon}_p)/\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.259	.293	.241	.244	.256	.221	.238	.226	.252	.227
	Portfolios for Estimation Period 1942–46									
$\hat{\beta}_{p,t-1}$	.467	.537	.593	.628	.707	.721	.770	.792	.805	.894
$s(\hat{\beta}_{p,t-1})$	.045	.041	.044	.037	.027	.032	.035	.035	.028	.040
$r(R_p, R_m)^2$	.645	.745	.753	.829	.919	.898	.889	.898	.934	.896
$s(R_p)$	.035	.037	.041	.041	.044	.046	.049	.050	.050	.057
$s(\hat{\epsilon}_p)$	.021	.019	.020	.017	.013	.015	.016	.016	.013	.018
$\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.055	.055	.063	.058	.058	.063	.064	.064	.062	.069
$s(\hat{\epsilon}_p)/\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.382	.345	.317	.293	.224	.238	.250	.250	.210	.261
	Portfolios for Estimation Period 1950–54									
$\hat{\beta}_{p,t-1}$	.418	.590	.694	.751	.777	.784	.929	.950	.996	1.014
$s(\hat{\beta}_{p,t-1})$	.042	.047	.045	.037	.038	.035	.050	.038	.035	.029
$r(R_p, R_m)^2$	.629	.723	.798	.872	.878	.895	.856	.913	.933	.954
$s(R_p)$	.019	.025	.028	.029	.030	.030	.036	.036	.037	.038
$s(\hat{\epsilon}_p)$	.012	.013	.013	.010	.010	.010	.014	.011	.010	.008
$\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.040	.044	.046	.048	.051	.051	.052	.053	.054	.057
$s(\hat{\epsilon}_p)/\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.300	.295	.283	.208	.196	.196	.269	.208	.185	.140
	Portfolios for Estimation Period 1958–62									
$\hat{\beta}_{p,t-1}$	.626	.635	.719	.801	.817	.860	.920	.950	.975	.995
$s(\hat{\beta}_{p,t-1})$	.043	.048	.039	.046	.047	.033	.037	.038	.032	.037
$r(R_p, R_m)^2$	.783	.745	.851	.835	.838	.920	.913	.915	.939	.925
$s(R_p)$	.030	.031	.033	.037	.038	.038	.041	.042	.043	.044
$s(\hat{\epsilon}_p)$	.014	.016	.013	.015	.015	.011	.012	.012	.011	.012
$\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.049	.052	.056	.059	.064	.061	.070	.069	.068	.064
$s(\hat{\epsilon}_p)/\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.286	.308	.232	.254	.234	.180	.171	.174	.162	.188

be rejected when t-statistics are interpreted under the assumption of normality, the hypotheses are on even firmer ground when one takes into account the thick tails of empirical return distributions.

Further justification for using t-statistics to test hypotheses on monthly common stock returns is in the work of Officer (1971). Under the assumption that distributions of monthly returns are symmetric stable, he estimates that in the post–World War II period the characteristic exponent

TABLE 2 (*Continued*)

Statistic	11	12	13	14	15	16	17	18	19	20
	Portfolios for Estimation Period 1934–38									
$\hat{\beta}_{p,t-1}$	1.046	1.122	1.181	1.192	1.196	1.295	1.335	1.396	1.445	1.458
$s(\hat{\beta}_{p,t-1})$	.028	.031	.035	.028	.029	.032	.032	.053	.039	.053
$r(R_p, R_m)^2$	.959	.956	.951	.969	.966	.966	.967	.922	.958	.927
$s(R_p)$	.113	.122	.128	.128	.129	.140	.144	.154	.156	.160
$s(\hat{\epsilon}_p)$	.023	.026	.029	.023	.024	.026	.026	.043	.032	.043
$\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.094	.124	.120	.122	.132	.125	.129	.158	.145	.170
$s(\hat{\epsilon}_p)/\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.245	.210	.242	.188	.182	.208	.202	.272	.221	.253
	Portfolios for Estimation Period 1942–46									
$\hat{\beta}_{p,t-1}$	.949	.952	1.010	1.038	1.254	1.312	1.316	1.473	1.631	1.661
$s(\hat{\beta}_{p,t-1})$	.031	.036	.040	.030	.034	.039	.041	.084	.083	.077
$r(R_p, R_m)^2$	.942	.923	.917	.954	.958	.951	.945	.839	.867	.887
$s(R_p)$	.059	.060	.063	.064	.077	.081	.081	.097	.105	.106
$s(\hat{\epsilon}_p)$	.014	.016	.018	.014	.016	.018	.019	.039	.038	.036
$\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.073	.074	.085	.077	.096	.083	.086	.134	.117	.122
$s(\hat{\epsilon}_p)/\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.192	.216	.212	.182	.167	.217	.221	.291	.325	.295
	Portfolios for Estimation Period 1950–54									
$\hat{\beta}_{p,t-1}$	1.117	1.123	1.131	1.134	1.186	1.235	1.295	1.324	1.478	1.527
$s(\hat{\beta}_{p,t-1})$	.039	.027	.044	.033	.037	.049	.045	.046	.058	.086
$r(R_p, R_m)^2$	.934	.968	.919	.952	.944	.915	.933	.934	.917	.841
$s(R_p)$	.042	.041	.043	.042	.044	.047	.049	.050	.056	.060
$s(\hat{\epsilon}_p)$	.011	.007	.012	.009	.010	.014	.013	.013	.016	.024
$\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.066	.057	.066	.060	.064	.064	.065	.068	.076	.088
$s(\hat{\epsilon}_p)/\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.167	.123	.182	.150	.156	.219	.200	.192	.210	.273
	Portfolios for Estimation Period 1958–62									
$\hat{\beta}_{p,t-1}$	1.013	1.019	1.037	1.048	1.069	1.081	1.092	1.098	1.269	1.388
$s(\hat{\beta}_{p,t-1})$	.038	.031	.036	.033	.036	.038	.045	.045	.048	.065
$r(R_p, R_m)^2$	.922	.948	.934	.945	.936	.931	.907	.910	.922	.886
$s(R_p)$	.045	.045	.046	.046	.047	.048	.049	.049	.056	.063
$s(\hat{\epsilon}_p)$	.013	.010	.012	.011	.012	.013	.015	.015	.016	.021
$\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.069	.066	.067	.062	.070	.072	.076	.068	.070	.078
$s(\hat{\epsilon}_p)/\bar{s}_{p,t-1}(\hat{\epsilon}_i)$	.188	.152	.179	.177	.171	.180	.197	.220	.228	.269

for these distributions is about 1.8 (as compared with a value of 2.0 for a normal distribution). From Fama and Roll (1968), for values of the characteristic exponent so close to 2.0 stable nonnormal distributions differ noticeably from the normal only in their extreme tails—that is, beyond the .05 and .95 fractiles. Thus, as long as one is not concerned with precise estimates of probability levels (always a somewhat meaningless activity), interpreting t-statistics in the usual way does not lead to serious errors.

TABLE 3

SUMMARY RESULTS FOR THE REGRESSION

$$R_p = \hat{\gamma}_{0t} + \hat{\gamma}_{1t}\hat{\beta}_p + \hat{\gamma}_{2t}\hat{\beta}^2_p + \hat{\gamma}_{3t}\bar{s}_p(\hat{\epsilon}_i) + \hat{\eta}_{pt}$$

Period	Statistic																			
	$\bar{\hat{\gamma}}_0$	$\bar{\hat{\gamma}}_1$	$\bar{\hat{\gamma}}_2$	$\bar{\hat{\gamma}}_3$	$\overline{\hat{\gamma}_0 - R_f}$	$s(\hat{\gamma}_0)$	$s(\hat{\gamma}_1)$	$s(\hat{\gamma}_2)$	$s(\hat{\gamma}_3)$	$\rho_0(\hat{\gamma}_0 - R_f)$	$\rho_M(\hat{\gamma}_1)$	$\rho_0(\hat{\gamma}_2)$	$\rho_0(\hat{\gamma}_3)$	$t(\bar{\hat{\gamma}}_0)$	$t(\bar{\hat{\gamma}}_1)$	$t(\bar{\hat{\gamma}}_2)$	$t(\bar{\hat{\gamma}}_3)$	$t(\overline{\hat{\gamma}_0 - R_f})$	$\bar{r}^2$	$s(r^2)$
Panel A:																				
1935–6/68 ..	.0061	.0085	...	...	.0048	.038	.066	...	...	.15	.02	...	...	3.24	2.57	...	...	2.55	.29	.30
1935–45	.0039	.0163	...	...	.0037	.052	.098	...	...	.10	-.03	...	...	.86	1.92	...	...	.82	.29	.29
1946–55	.0087	.0027	...	...	.0078	.026	.041	...	...	.18	.07	...	...	3.71	.70	...	...	3.31	.31	.32
1956–6/68 ..	.0060	.0062	...	...	.0034	.030	.044	...	...	.27	.15	...	...	2.45	1.73	...	...	1.39	.28	.29
1935–40	.0024	.0109	...	...	.0023	.064	.116	...	...	.07	-.09	...	...	.32	.79	...	...	.31	.23	.30
1941–45	.0056	.0229	...	...	.0054	.034	.069	...	...	.23	.15	...	...	1.27	2.55	...	...	1.22	.37	.28
1946–50	.0050	.0029	...	...	.0044	.031	.047	...	...	.20	.04	...	...	1.27	.48	...	...	1.10	.39	.33
1951–55	.0123	.0024	...	...	.0111	.019	.035	...	...	.20	.08	...	...	5.06	.53	...	...	4.56	.24	.29
1956–60	.0148	-.0059	...	...	.0128	.020	.034	...	...	.37	.18	...	...	5.68	-1.37	...	...	4.89	.22	.31
1961–6/68 ..	.0001	.0143	...	...	-.0029	.034	.048	...	...	.22	.09	...	...	.03	2.81	...	...	-.80	.32	.27
Panel B:																				
1935–6/68 ..	.0049	.0105	-.0008	...	.0036	.052	.118	.056	...	.03	-.11	-.11	...	1.92	1.79	-.29	...	1.42	.32	.31
1935–45	.0074	.0079	.0040	...	.0073	.061	.139	.074	...	-.10	-.31	-.21	...	1.39	.65	.61	...	1.36	.32	.30
1946–55	-.0002	.0217	-.0087	...	-.0012	.036	.095	.034	...	.04	.00	.00	...	-.07	2.51	-2.83	...	-.38	.36	.32
1956–6/68 ..	.0069	.0040	.0013	...	.0043	.054	.116	.053	...	.17	.07	.03	...	1.56	.42	.29	...	.97	.30	.30
1935–40	.0013	.0141	-.0017	...	.0012	.069	.160	.075	...	-.13	-.36	-.35	...	.16	.75	-.19	...	.14	.24	.30
1941–45	.0148	.0004	.0108	...	.0146	.050	.111	.073	...	-.04	-.19	-.04	...	2.28	.03	1.15	...	2.24	.39	.29
1946–50	-.0008	.0152	-.0051	...	-.0015	.037	.104	.032	...	.14	.04	.00	...	-.18	1.14	-1.24	...	-.32	.44	.32
1951–55	.0004	.0281	-.0122	...	-.0008	.030	.085	.035	...	-.17	-.14	-.01	...	.10	2.55	-2.72	...	-.20	.28	.29
1956–60	.0128	-.0015	-.0020	...	.0108	.030	.072	.029	...	.35	.11	.26	...	3.38	-.16	-.54	...	2.84	.25	.31
1961–6/68 ..	.0029	.0077	.0034	...	-.0000	.066	.138	.064	...	.14	.06	-.01	...	.42	.53	.51	...	-.01	.34	.29

TABLE 3 (*Continued*)

Period	$\bar{\hat{\gamma}}_0$	$\bar{\hat{\gamma}}_1$	$\bar{\hat{\gamma}}_2$	$\bar{\hat{\gamma}}_3$	$\overline{\hat{\gamma}_0 - R_f}$	$s(\hat{\gamma}_0)$	$s(\hat{\gamma}_1)$	$s(\hat{\gamma}_2)$	$s(\hat{\gamma}_3)$	$\rho_0(\hat{\gamma}_0 - R_f)$	$\rho_M(\hat{\gamma}_1)$	$\rho_0(\hat{\gamma}_2)$	$\rho_0(\hat{\gamma}_3)$	$t(\bar{\hat{\gamma}}_0)$	$t(\bar{\hat{\gamma}}_1)$	$t(\bar{\hat{\gamma}}_2)$	$t(\bar{\hat{\gamma}}_3)$	$t(\overline{\hat{\gamma}_0 - R_f})$	$\bar{r}^2$	$s(r^2)$
Panel C:																				
1935–6/68 ..	.0054	.0072	...	.0198	.0041	.052	.065	...	.868	.04	—.12	...	—.04	2.10	2.20	...	.46	1.59	.32	.31
1935–45	.0017	.0104	...	.0841	.0015	.073	.083	...	.921	—.00	—.26	...	—.08	.26	1.41	...	1.05	.24	.32	.31
1946–55	.0110	.0075	...	—.1052	.0100	.032	.056	...	.609	.08	.02	...	—.20	3.78	1.47	...	—1.89	3.46	.34	.32
1956–6/68 ..	.0042	.0041	...	.0633	.0016	.040	.052	...	.984	.12	.08	...	.03	1.28	.96	...	.79	.50	.30	.29
1935–40	.0036	.0119	...	—.0170	.0035	.082	.105	...	.744	—.03	—.26	...	—.18	.37	.97	...	—.19	.36	.25	.30
1941–45	—.0006	.0085	...	.2053	—.0009	.061	.052	...	1.091	.07	—.29	...	—.02	—.08	1.25	...	1.46	—.11	.41	.30
1946–50	.0069	.0081	...	—.0920	.0062	.034	.066	...	.504	.14	.06	...	—.02	1.56	.95	...	—1.41	1.40	.42	.33
1951–55	.0150	.0069	...	—.1185	.0138	.029	.043	...	.702	.06	—.18	...	—.32	4.05	1.24	...	—1.31	3.72	.27	.29
1956–60	.0127	—.0081	...	.0728	.0107	.037	.045	...	1.164	.15	.15	...	.21	2.68	—1.40	...	.48	2.26	.26	.30
1961–6/68 ..	—.0014	.0122	...	.0570	—.0044	.042	.055	...	.850	.10	.00	...	—.19	—.32	2.12	...	.64	—.98	.33	.27
Panel D:																				
1935–6/68 ..	.0020	.0114	—.0026	.0516	.0008	.075	.123	.060	.929	—.09	—.09	—.12	—.10	.55	1.85	—.86	1.11	.20	.34	.31
1935–45	.0011	.0118	—.0009	.0817	.0010	.103	.146	.079	1.003	—.20	—.23	—.24	—.15	.13	.94	—.14	.94	.11	.34	.31
1946–55	.0017	.0209	—.0076	—.0378	.0008	.042	.096	.038	.619	—.10	—.00	—.01	—.20	.44	2.39	—2.16	—.67	.20	.36	.32
1956–6/68 ..	.0031	.0034	—.0000	.0966	.0005	.065	.122	.055	1.061	.12	.03	.01	—.05	.59	.34	—.00	1.11	.10	.32	.29
1935–40	.0009	.0156	—.0029	.0025	.0008	.112	.171	.085	.826	—.16	—.23	—.26	—.12	.07	.78	—.29	.03	.06	.26	.30
1941–45	.0015	.0073	.0014	.1767	.0012	.092	.109	.072	1.181	—.28	—.21	—.22	—.18	.12	.52	.15	1.16	.10	.43	.31
1946–50	.0011	.0141	—.0040	—.0313	.0004	.047	.106	.042	.590	—.10	.03	—.01	—.12	.18	1.03	—.73	—.41	.07	.44	.33
1951–55	.0023	.0277	—.0112	—.0443	.0011	.037	.085	.034	.651	—.11	—.13	—.01	—.28	.48	2.53	—2.54	—.53	.23	.29	.30
1956–60	.0103	—.0047	—.0020	.0979	.0083	.049	.078	.032	1.286	—.16	.19	—.01	.02	1.63	—.47	—.49	.59	1.31	.28	.30
1961–6/68 ..	—.0017	.0088	.0013	.0957	—.0046	.073	.144	.066	.887	.20	.00	.01	—.15	—.21	.58	.19	1.02	—.60	.35	.29

Inferences based on approximate normality are on even safer ground if one assumes, again in line with the results of Officer (1971), that although they are well approximated by stable nonnormal distributions with $\alpha \cong 1.8$, distributions of monthly returns in fact have finite variances and converge—but very slowly—toward the normal as one takes sums or averages of individual returns. Then the distributions of the means of month-by-month regression coefficients from the risk-return model are likely to be close to normal since each mean is based on coefficients for many months.

A. Tests of the Major Hypotheses of the Two-Parameter Model

Consider first condition C2 of the two-parameter model, which says that no measure of risk, in addition to β, systematically affects expected returns. This hypothesis is not rejected by the results in panels C and D of table 3. The values of $t(\bar{\hat{\gamma}}_3)$ are small, and the signs of the $t(\bar{\hat{\gamma}}_3)$ are randomly positive and negative.

Likewise, the results in panels B and D of table 3 do not reject condition C1 of the two-parameter model, which says that the relationship between expected return and β is linear. In panel B, the value of $t(\bar{\hat{\gamma}}_2)$ for the overall period 1935–6/68 is only $-.29$. In the 5-year subperiods, $t(\bar{\hat{\gamma}}_2)$ for 1951–55 is approximately -2.7, but for subperiods that do not cover 1951–55, the values of $t(\bar{\hat{\gamma}}_2)$ are much closer to zero.

So far, then, the two-parameter model seems to be standing up well to the data. All is for naught, however, if the critical condition C3 is rejected. That is, we are not happy with the model unless there is on average a positive tradeoff between risk and return. This seems to be the case. For the overall period 1935–6/68, $t(\bar{\hat{\gamma}}_1)$ is large for all models. Except for the period 1956–60, the values of $t(\bar{\hat{\gamma}}_1)$ are also systematically positive in the subperiods, but not so systematically large.

The small t-statistics for subperiods reflect the substantial month-to-month variability of the parameters of the risk-return regressions. For example, in the one-variable regressions summarized in panel A, for the period 1935–40, $\bar{\hat{\gamma}}_1 = .0109$. In other words, for this period the average incremental return per unit of β was almost 1.1 percent per month, so that on average, bearing risk had substantial rewards. Nevertheless, because of the variability of $\hat{\gamma}_{1t}$—in this period $s(\hat{\gamma}_1)$ is 11.6 percent per month (!)—$t(\bar{\hat{\gamma}}_1)$ is only .79. It takes the statistical power of the large sample for the overall period before values of $\bar{\hat{\gamma}}_1$ that are large in practical terms also yield large t-values.

But at least with the sample of the overall period $t(\bar{\hat{\gamma}}_1)$ achieves values supportive of the conclusion that on average there is a statistically observable positive relationship between return and risk. This is not the case with respect to $t(\bar{\hat{\gamma}}_2)$ and $t(\bar{\hat{\gamma}}_3)$. Even, or indeed especially, for the overall period, these t-statistics are close to zero.

The behavior through time of $\hat{\gamma}_{1t}$, $\hat{\gamma}_{2t}$, and $\hat{\gamma}_{3t}$ is also consistent with hypothesis ME that the capital market is efficient. The serial correlations $\rho_M(\hat{\gamma}_1)$, $\rho_0(\hat{\gamma}_2)$, and $\rho_0(\hat{\gamma}_3)$, are always low in terms of explanatory power and generally low in terms of statistical significance. The proportion of the variance of $\tilde{\gamma}_{jt}$ explained by first-order serial correlation is estimated by $\rho(\hat{\gamma}_j)^2$ which in all cases is small. As for statistical significance, under the hypothesis that the true serial correlation is zero, the standard deviation of the sample coefficient can be approximated by $\sigma(\hat{\rho}) = 1/\sqrt{n}$. For the overall period, $\sigma(\hat{\rho})$ is approximately .05, while for the 10- and 5-year subperiods $\sigma(\hat{\rho})$ is approximately .09 and .13, respectively. Thus, the values of $\rho_M(\hat{\gamma}_1)$, $\rho_0(\hat{\gamma}_2)$, and $\rho_0(\hat{\gamma}_3)$ in table 3 are generally statistically close to zero. The exceptions involve primarily periods that include the 1935–40 subperiod, and the results for these periods are not independent.[8]

To conserve space, the serial correlations of the portfolio residuals, $\hat{\eta}_{pt}$, are not shown. In these serial correlations, negative values predominate. But like the serial correlations of the $\hat{\gamma}$'s, those of the $\hat{\eta}$'s are close to zero. Higher-order serial correlations of the $\hat{\gamma}$'s and $\hat{\eta}$'s have been computed, and these also are never systematically large.

In short, one cannot reject the hypothesis that the pricing of securities is in line with the implications of the two-parameter model for expected returns. And given a two-parameter pricing model, the behavior of returns through time is consistent with an efficient capital market.

B. *The Behavior of the Market*

Some perspective on the behavior of the market during different periods and on the interpretation of the coefficients $\hat{\gamma}_{0t}$ and $\hat{\gamma}_{1t}$ in the risk-return regressions can be obtained from table 4. For the various periods of table 3, table 4 shows the sample means (and with some exceptions), the standard

[8] The serial correlations of $\hat{\gamma}_2$ and $\hat{\gamma}_3$ about means that are assumed to be zero provide a test of the fair game property of an efficient market, given that expected returns are generated by the two-parameter model—that is, given $E(\tilde{\gamma}_{2t}) = E(\tilde{\gamma}_{3t}) = 0$. Likewise, $\rho_0(\hat{\gamma}_{0t} - R_{ft})$ provides a test of market efficiency with respect to the behavior of $\hat{\gamma}_{0t}$ through time, given the validity of the Sharpe-Lintner hypothesis (about which we have as yet said nothing). But, at least for $\hat{\gamma}_{2t}$ and $\hat{\gamma}_{3t}$, computing the serial correlations about sample means produces essentially the same results.

To test the market efficiency hypothesis on $\tilde{\gamma}_{1t} - [E(\tilde{R}_{mt}) - E(\tilde{R}_{0t})]$, the sample mean of the $\hat{\gamma}_{1t}$ is used to estimate $E(\tilde{R}_{mt}) - E(\tilde{R}_{0t})$, thus implicitly assuming that the expected risk premium is constant. That this is a reasonable approximation [in the sense that the $\rho_M(\hat{\gamma}_1)$ are small], probably reflects the fact that variation in $E(\tilde{R}_{mt}) - E(\tilde{R}_{0t})$ is trivial relative to the month-by-month variation in $\hat{\gamma}_{1t}$.

Finally, it is well to note that in terms of the implications of the serial correlations for making good portfolio decisions—and thus for judging whether market efficiency is a workable representation of reality—the fact that the serial correlations are low in terms of explanatory power is more important than whether or not they are low in terms of statistical significance.

TABLE 4

THE BEHAVIOR OF THE MARKET

Period	Statistic*								
	$\overline{R_m}$	$\overline{R_m - R_f}$	$\overline{\hat{\gamma}_1}$	$\overline{\hat{\gamma}_0}$	$\overline{R_f}$	$\overline{\dfrac{R_m - R_f}{s(R_m)}}$	$\overline{\dfrac{\hat{\gamma}_1}{s(R_m)}}$	$s(R_m)$	$s(R_m)$
1935–6/68	.0143	.0130	.0085	.0061	.0013	.2136	.1388	.061	.066
1935–45	.0197	.0195	.0163	.0039	.0002	.2207	.1844	.089	.098
1946–55	.0112	.0103	.0027	.0087	.0009	.2378	.0614	.043	.041
1956–6/68	.0121	.0095	.0062	.0060	.0026	.2387	.1560	.040	.044
1935–40	.0132	.0132	.0109	.0024	.0001	.1221	.1009	.108	.116
1941–45	.0274	.0272	.0229	.0056	.0002	.4715	.3963	.058	.069
1946–50	.0077	.0070	.0029	.0050	.0007	.1351	.0564	.052	.047
1951–55	.0148	.0136	.0024	.0123	.0012	.4174	.0735	.033	.035
1956–60	.0090	.0070	—.0059	.0148	.0020	.2080	—.1755	.034	.034
1961–6/68	.0141	.0111	.0143	.0001	.0030	.2567	.3294	.043	.048

* Since $s(R_f)$ is so small relative to $s(R_m)$, $s(R_m - R_f)$, which is not shown, is essentially the same as $s(R_m)$. The standard deviations of $(R_m - R_f)/s(R_m)$ and $\hat{\gamma}_1/s(R_m)$, also not shown, can be obtained directly from $s(R_m - R_f)$, $s(\hat{\gamma}_1)$ and $s(R_m)$. Finally, the t-statistics for $\overline{(R_m - R_f)/s(R_m)}$ and $\overline{\hat{\gamma}_1/s(R_m)}$ are identical with those for $\overline{R_m - R_f}$ and $\overline{\hat{\gamma}_1}$.

deviations, t-statistics for sample means, and first-order serial correlations for the month-by-month values of the following variables and coefficients: the market return R_{mt}; the riskless rate of interest R_{ft}, taken to be the yield on 1-month Treasury bills; $R_{mt} - R_{ft}$; $(R_{mt} - R_{ft})/s(R_m)$; $\hat{\gamma}_{0t}$ and $\hat{\gamma}_{1t}$, repeated from panel A of table 3; and $\hat{\gamma}_{1t}/s(R_m)$. The t-statistics on sample means are computed in the same way as those in table 3.

If the two-parameter model is valid, then in equation (7), $E(\tilde{\gamma}_{0t}) = E(\tilde{R}_{0t})$, where $E(\tilde{R}_{0t})$ is the expected return on any zero-β security or portfolio. Likewise, the expected risk premium per unit of β is $E(\tilde{R}_{mt}) - E(\tilde{R}_{0t}) = E(\tilde{\gamma}_{1t})$. In fact, for the one-variable regressions of panel A, table 3, that is,

$$R_{pt} = \hat{\gamma}_{0t} + \hat{\gamma}_{1t}\,\hat{\beta}_p + \hat{\eta}_{pt}, \tag{11}$$

we have, period by period,

$$\hat{\gamma}_{1t} = R_{mt} - \hat{\gamma}_{0t}. \tag{12}$$

This condition is obtained by averaging (11) over p and making use of the least-squares constraint

$$\sum_p \hat{\eta}_{pt} = 0.^{9}$$

Moreover, the least-squares estimate $\hat{\gamma}_{0t}$ can always be interpreted as the return for month t on a zero-$\hat{\beta}$ portfolio, where the weights given to each

[9] There is some degree of approximation in (12). The averages over p of R_{pt} and $\hat{\beta}_p$ are R_{mt} and 1.0, respectively, only if every security in the market is in some portfolio. With our methodology (see table 1) this is never true. But the degree of approximation turns out to be small: The average of the R_{pt} is always close to R_{mt} and the average $\hat{\beta}_p$ is always close to 1.0.

TABLE 4 (*Continued*)

				STATISTIC*						
$s(\hat{\gamma}_0)$	$s(R_f)$	$t(\bar{R}_m)$	$t(\overline{R_m-R_f})$	$t(\bar{\hat{\gamma}}_1)$	$t(\bar{\hat{\gamma}}_0)$	$\rho_M(R_m)$	$\rho_M(R_m-R_f)$	$\rho_M(\hat{\gamma}_1)$	$\rho_M(\hat{\gamma}_0)$	$\rho_M(R_f)$
.038	.0012	4.71	4.28	2.57	3.24	—.01	—.01	.02	.14	.98
.052	.0001	2.56	2.54	1.92	.86	—.07	—.07	—.03	.10	.88
.026	.0004	2.84	2.60	.70	3.71	.09	.09	.07	.10	.94
.030	.0009	3.72	2.92	1.73	2.45	.14	.14	.15	.25	.92
.064	.0001	1.04	1.04	.79	.32	—.13	—.13	—.09	.07	.72
.034	.0001	3.68	3.65	2.55	1.27	.14	.14	.15	.21	.83
.031	.0003	1.15	1.05	.48	1.27	.09	.09	.04	.18	.97
.019	.0004	3.51	3.22	.53	5.06	—.02	—.01	.08	—.07	.89
.020	.0007	2.07	1.60	—1.37	5.68	.12	.13	.18	.13	.80
.034	.0008	3.08	2.44	2.81	.03	.13	.13	.09	.21	.93

of the 20 portfolios to form this zero-$\hat{\beta}$ portfolio are the least-squares weights that are applied to the R_{pt} in computing $\hat{\gamma}_{0t}$.[10]

In the Sharpe-Lintner two-parameter model of market equilibrium $E(\tilde{\gamma}_{0t}) = E(\tilde{R}_{0t}) = R_{ft}$ and $E(\tilde{\gamma}_{1t}) = E(\tilde{R}_{mt}) - E(\tilde{R}_{0t}) = E(\tilde{R}_{mt}) - R_{ft}$. In the period 1935–40 and in the most recent period 1961–6/68, $\bar{\hat{\gamma}}_{1t}$ is close to $\overline{R_m - R_f}$ and the t-statistics for the two averages are similar. In other periods, and especially in the period 1951–60, $\bar{\hat{\gamma}}_1$ is substantially less than $\overline{R_m - R_f}$. This is a consequence of the fact that for these periods $\bar{\hat{\gamma}}_0$ is noticeably greater than $\bar{R}_f$. In economic terms, the tradeoff of average return for risk between common stocks and short-term bonds has been more consistently large through time than the tradeoff of average return for risk among common stocks. Testing whether the differences between $\overline{R_m - R_f}$ and $\bar{\hat{\gamma}}_1$ are statistically large, however, is equivalent to testing the S-L hypothesis $E(\tilde{\gamma}_{0t}) = R_{ft}$, which we prefer to take up after examining further the stochastic process generating monthly returns.

Finally, although the differences between values of $\overline{R_m - R_f}$ for different periods or between values of $\bar{\hat{\gamma}}_1$ are never statistically large, there is a hint in table 4 that average-risk premiums declined from the pre– to the post–World War II periods. These are average risk premiums per unit of $\hat{\beta}$, however, which are not of prime interest to the investor. In making his portfolio decision, the investor is more concerned with the tradeoff of expected portfolio return for dispersion of return—that is, the slope of the efficient set of portfolios. In the Sharpe-Lintner model this slope is

[10] That $\hat{\gamma}_{0t}$ is the return on a zero-$\hat{\beta}$ portfolio can be shown to follow from the unbiasedness of the least-squares coefficients in the cross-sectional risk-return regressions. If one makes the Gauss-Markov assumptions that the underlying disturbances $\tilde{\eta}_{pt}$ of (11) have zero means, are uncorrelated across p, and have the same variance for all p, then it follows almost directly from the Gauss-Markov Theorem that the least-squares estimate $\hat{\gamma}_{0t}$ is also the return for month t on the minimum variance zero-$\hat{\beta}$ portfolio that can be constructed from the 20 portfolio $\hat{\beta}_p$.

always $[E(\tilde{R}_{mt}) - R_{ft}]/\sigma(\tilde{R}_{mt})$, and in the more general model of Black (1972), it is $[E(\tilde{R}_{mt}) - E(\tilde{R}_{0t})]/\sigma(\tilde{R}_{mt})$ at the point on the efficient set corresponding to the market portfolio m. In table 4, especially for the three long subperiods, dividing $\overline{R_m - R_f}$ and $\bar{\hat{\gamma}}_1$, by $s(R_m)$ seems to yield estimated risk premiums that are more constant through time. This results from the fact that any declines in $\bar{\hat{\gamma}}_1$ or $\overline{R_m - R_f}$ are matched by a quite noticeable downward shift in $s(R_m)$ from the early to the later periods (cf. Blume [1970] or Officer [1971]).

C. *Errors and True Variation in the Coefficients* $\hat{\gamma}_{jt}$

Each cross-sectional regression coefficient $\hat{\gamma}_{jt}$ in (10) has two components: the true $\tilde{\gamma}_{jt}$ and the estimation error, $\tilde{\phi}_{jt} = \hat{\gamma}_{jt} - \tilde{\gamma}_{jt}$. A natural question is: To what extent is the variation in $\hat{\gamma}_{jt}$ through time due to variation in $\tilde{\gamma}_{jt}$ and to what extent is it due to $\tilde{\phi}_{jt}$? In addition to providing important information about the precision of the coefficient estimates used to test the two-parameter model, the answer to this question can be used to test hypotheses about the stochastic process generating returns. For example, although we cannot reject the hypothesis that $E(\tilde{\gamma}_{2t}) = 0$, does including the term involving $\hat{\beta}_p^2$ in (10) help in explaining the month-by-month behavior of returns? That is, can we reject the hypothesis that for all t, $\tilde{\gamma}_{2t} = 0$? Likewise, can we reject the hypothesis that month-by-month $\tilde{\gamma}_{3t} = 0$? And is the variation through time in $\hat{\gamma}_{0t}$ due entirely to $\tilde{\phi}_{0t}$ and to variation in R_{ft}?

The answers to these questions are in table 5. For the models and time periods of table 3, table 5 shows for each $\hat{\gamma}_j$: $s^2(\hat{\gamma}_j)$, the sample variance of the month-by-month $\hat{\gamma}_{jt}$; $\overline{s^2(\tilde{\phi}_j)}$, the average of the month-by-month values of $s^2(\tilde{\phi}_{jt})$, where $s(\tilde{\phi}_{jt})$ is the standard error of $\hat{\gamma}_{jt}$ from the cross-sectional risk-return regression of (10) for month t; $s^2(\tilde{\gamma}_j) \equiv s^2(\hat{\gamma}_j) - \overline{s^2(\tilde{\phi}_j)}$; and the F-statistic $F = s^2(\hat{\gamma}_j)/\overline{s^2(\tilde{\phi}_j)}$, which is relevant for testing the hypothesis, $s^2(\hat{\gamma}_j) = \overline{s^2(\tilde{\phi}_j)}$. The numerator of F has $n - 1$ df, where n is the number of months in the sample period; and the denominator has $n(20 - K)$ df, where K is the number of coefficients $\hat{\gamma}_j$ in the model.[11]

[11] The standard error of $\hat{\gamma}_{jt}$, $s(\tilde{\phi}_{jt})$, is proportional to the standard error of the risk-return residuals, $\hat{\eta}_{pt}$, for month t, which has $20 - K$ df. And n values of $s^2(\tilde{\phi}_{jt})$ are averaged to get $\overline{s^2(\tilde{\phi}_j)}$, so that the latter has $n(20 - K)$ df. Note that if the underlying return disturbances $\tilde{\eta}_{pt}$ of (10) are independent across p and have identical normal distributions for all p, then $\hat{\gamma}_{jt}$ is the sample mean of a normal distribution and $s^2(\tilde{\phi}_{jt})$ is proportional to the sample variance of the same normal distribution. If the process is also assumed to be stationary through time, it then follows that $s^2(\hat{\gamma}_{jt})$ and $s^2(\tilde{\phi}_{jt})$ are independent, as required by the F-test. Finally, in the F-statistics of table 5, the values of n are 60 or larger, so that, since K is from 2 to 4, $n(20 - K) \geq 960$. From Mood and Graybill (1963), some upper percentage points of the F-distribution are:

One clear-cut result in table 5 is that there is a substantial decline in the reliability of the coefficients $\hat{\gamma}_{0t}$ and $\hat{\gamma}_{1t}$—that is, a substantial increase in $s^2(\overline{\hat{\phi}_0})$ and $s^2(\overline{\hat{\phi}_1})$—when $\hat{\beta}_p^2$ and/or $\bar{s}_p(\hat{\epsilon}_j)$ are included in the risk-return regressions. The variable $\hat{\beta}_p^2$ is obviously collinear with $\hat{\beta}_p$, and, as can be seen from table 2, $\bar{s}_p(\hat{\epsilon}_i)$ likewise increases with $\hat{\beta}_p$. From panels B and C of table 5, the collinearity with $\hat{\beta}_p$ is stronger for $\hat{\beta}_p^2$ than for $\bar{s}_p(\hat{\epsilon}_j)$.

In spite of the loss in precision that arises from multicollinearity, however, the F-statistics for $\hat{\gamma}_2$ (the coefficient of $\hat{\beta}_p^2$) and $\hat{\gamma}_3$ [the coefficient of $\bar{s}_p(\hat{\epsilon}_j)$] are generally large for the models of panels B and C of table 5, and for the model of panel D which includes both variables. From the F-statistics in panel D, it seems that, except for the period 1935–45, the variation through time of $\tilde{\gamma}_{2t}$ is statistically more noticeable than that of $\tilde{\gamma}_{3t}$, but there are periods (1941–45, 1956–60) when the values of F for both $\tilde{\gamma}_{2t}$ and $\tilde{\gamma}_{3t}$ are large.

The F-statistics for $\hat{\gamma}_{1t} = \tilde{\gamma}_{1t} + \hat{\phi}_{1t}$ also indicate that $\tilde{\gamma}_{1t}$ has substantial variation through time. This is not surprising, however, since $\hat{\gamma}_{1t}$ is always directly related to $\tilde{R}_{mt}$. For example, from equation (12), for the one-variable model of panel A, $\hat{\gamma}_{1t} = \tilde{R}_{mt} - \hat{\gamma}_{0t}$.

Finally, the F-statistics for $\hat{\gamma}_{0t} = \tilde{\gamma}_{0t} + \hat{\phi}_{0t}$ are also in general large. And the month-by-month variation in $\tilde{\gamma}_{0t}$ cannot be accounted for by variation in R_{ft}. The variance of R_{ft} is so small relative to $s^2(\hat{\gamma}_{0t})$, $s^2(\tilde{\gamma}_{0t})$, and $s^2(\hat{\phi}_{0t})$ that doing the F-tests in terms of $\hat{\gamma}_{0t} - R_{ft}$ produces results almost identical with those for $\hat{\gamma}_{0t}$.

Rejection of the hypothesis that $\tilde{\gamma}_{0t} - R_{ft} = 0$ does not imply rejection of the S-L hypothesis—to be tested next—that $E(\tilde{\gamma}_{0t}) = R_{ft}$. Likewise, to find that month-by-month $\tilde{\gamma}_{2t} \neq 0$ and $\tilde{\gamma}_{3t} \neq 0$ does not imply rejection of hypotheses C1 and C2 of the two-parameter model. These hypotheses, which we are unable to reject on the basis of the results in table 3, say that $E(\tilde{\gamma}_{2t}) = 0$ and $E(\tilde{\gamma}_{3t}) = 0$.

What we have found in table 5 is that there are variables in addition to $\hat{\beta}_p$ that systematically affect period-by-period returns. Some of these omitted variables are apparently related to $\hat{\beta}_p^2$ and $\bar{s}_p(\tilde{\epsilon}_i)$. But the latter are almost surely proxies, since there is no economic rationale for their presence in our stochastic risk-return model.

---

n	$F_{.90}$	$F_{.95}$	$F_{.975}$	$F_{.99}$	$F_{.995}$
60 (120)	1.35	1.47	1.58	1.73	1.83
60 (∞)	1.29	1.39	1.48	1.60	1.69
120 (120)	1.26	1.35	1.43	1.53	1.61
120 (∞)	1.19	1.25	1.31	1.38	1.43

TABLE 5

COMPONENTS OF THE VARIANCES OF THE $\hat{\gamma}_{jt}$

PERIOD	$s^2(\tilde{\gamma}_0)$	$s^2(\hat{\gamma}_0)$	$\overline{s^2(\tilde{\phi}_0)}$	F	$s^2(\tilde{\gamma}_1)$	$s^2(\hat{\gamma}_1)$	$\overline{s^2(\tilde{\phi}_1)}$	F
Panel A:								
1935–6/68 ...	.00105	.00142	.00037	3.84	.00401	.00436	.00035	12.46
1935–45	.00182	.00273	.00091	3.00	.00863	.00950	.00087	10.92
1946–55	.00057	.00066	.00009	7.33	.00163	.00171	.00008	21.38
1956–6/68 ...	.00077	.00090	.00013	6.92	.00181	.00193	.00012	16.08
1935–40	.00265	.00404	.00139	2.91	.01212	.01347	.00135	9.98
1941–45	.00086	.00118	.00032	3.69	.00452	.00481	.00029	16.59
1946–50	.00086	.00094	.00008	11.75	.00216	.00224	.00008	28.00
1951–55	.00027	.00036	.00009	4.00	.00113	.00121	.00008	15.12
1956–60	.00032	.00041	.00009	4.56	.00104	.00112	.00008	21.50
1961–6/68 ...	.00100	.00114	.00014	8.14	.00217	.00231	.00014	16.50
Panel B:								
1935–6/68 ...	.00092	.00267	.00175	1.52	.00564	.01403	.00839	1.67
1935–45	.00057	.00377	.00320	1.18	.00372	.01941	.01569	1.24
1946–55	.00053	.00112	.00059	1.90	.00651	.00897	.00245	3.66
1956–6/68 ...	.00155	.00294	.00139	2.12	.00667	.01338	.00671	1.99
1935–40	.00018	.00476	.00458	1.04	.00374	.02555	.02181	1.17
1941–45	.00101	.00254	.00153	1.66	.00389	.01225	.00836	1.46
1946–50	.00084	.00136	.00052	2.62	.00862	.01071	.00209	5.12
1951–55	.00024	.00090	.00066	1.36	.00447	.00729	.00282	2.58
1956–60	.00037	.00087	.00050	1.74	.00289	.00517	.00228	2.27
1961–6/68 ...	.00232	.00431	.00199	2.16	.00928	.01894	.00966	1.96
Panel C:								
1935–6/68 ...	.00192	.00266	.00075	3.55	.00285	.00428	.00142	3.01
1935–45	.00394	.00533	.00139	3.83	.00433	.00717	.00283	2.52
1946–55	.00083	.00101	.00018	5.61	.00261	.00310	.00050	6.20
1956–6/68 ...	.00100	.00164	.00063	2.60	.00178	.00270	.00092	2.93
1935–40	.00473	.00669	.00196	3.41	.00732	.01094	.00362	3.02
1941–45	.00307	.00377	.00070	5.38	.00085	.00274	.00189	1.45
1946–50	.00103	.00117	.00014	8.36	.00386	.00439	.00053	8.28
1951–55	.00061	.00083	.00022	3.77	.00140	.00188	.00047	4.00
1956–60	.00079	.00134	.00055	2.44	.00106	.00204	.00098	2.08
1961–6/68 ...	.00109	.00177	.00068	2.60	.00212	.00300	.00088	3.41
Panel D:								
1935–6/68 ...	.00150	.00566	.00406	1.39	.00608	.01521	.00913	1.66
1935–45	.00233	.01065	.00832	1.28	.00402	.02118	.01716	1.23
1946–55	.00013	.00176	.00163	1.08	.00647	.00916	.00269	3.41
1956–6/68 ...	.00194	.00420	.00226	1.86	.00763	.01485	.00722	2.06
1935–40	.00157	.01263	.01106	1.14	.00457	.02910	.02453	1.19
1941–45	.00340	.00843	.00503	1.68	.00365	.01196	.00832	1.44
1946–50	.00023	.00220	.00197	1.12	.00858	.01119	.00261	4.29
1951–55	.00006	.00136	.00130	1.05	.00442	.00719	.00277	2.60
1956–60	.00092	.00239	.00147	1.62	.00328	.00602	.00274	2.20
1961–6/68 ...	.00260	.00539	.00279	1.93	.01060	.02081	.01021	2.04

D. *Tests of the S-L Hypothesis*

In the Sharpe-Lintner two-parameter model of market equilibrium one has, in addition to conditions C1–C3, the hypothesis that $E(\tilde{\gamma}_{0t}) = R_{ft}$. The work of Friend and Blume (1970) and Black, Jensen, and Scholes (1972) suggests that the S-L hypothesis is not upheld by the data. At least in the post–World War II period, estimates of $E(\tilde{\gamma}_{0t})$ seem to be significantly greater than R_{ft}.

Each of the four models of table 3 can be used to test the S-L hypothe-

TABLE 5 (*Continued*)

Period	$s^2(\tilde{\gamma}_2)$	$s^2(\hat{\gamma}_2)$	$\overline{s^2(\tilde{\phi}_2)}$	F	$s^2(\tilde{\gamma}_3)$	$s^2(\hat{\gamma}_3)$	$\overline{s^2(\tilde{\phi}_3)}$	F
Panel A:								
1935–6/68	...	...	...	...	...	...	...	...
1935–45	...	...	...	...	...	...	...	...
1946–55	...	...	...	...	...	...	...	...
1956–6/68	...	...	...	...	...	...	...	...
1935–40	...	...	...	...	...	...	...	...
1941–45	...	...	...	...	...	...	...	...
1946–50	...	...	...	...	...	...	...	...
1951–55	...	...	...	...	...	...	...	...
1956–60	...	...	...	...	...	...	...	...
1961–6/68	...	...	...	...	...	...	...	...
Panel B:								
1935–6/68	.00121	.00318	.00197	1.61	...	...	...	...
1935–45	.00171	.00548	.00377	1.45	...	...	...	...
1946–55	.00063	.00112	.00049	2.29	...	...	...	...
1956–6/68	.00122	.00278	.00156	1.78	...	...	...	...
1935–40	.00041	.00566	.00524	1.08	...	...	...	...
1941–45	.00327	.00527	.00201	2.62	...	...	...	...
1946–50	.00066	.00103	.00037	2.78	...	...	...	...
1951–55	.00058	.00120	.00062	1.94	...	...	...	...
1956–60	.00033	.00083	.00050	1.66	...	...	...	...
1961–6/68	.00182	.00410	.00227	1.81	...	...	...	...
Panel C:								
1935–6/68	...	...	...	...	.341	.753	.412	1.83
1935–45	...	...	...	...	.535	.847	.313	2.71
1946–55	...	...	...	...	.165	.370	.206	1.80
1956–6/68	...	...	...	...	.304	.968	.664	1.46
1935–40	...	...	...	...	.270	.553	.282	1.96
1941–45	...	...	...	...	.840	1.189	.349	3.41
1946–50	...	...	...	...	.118	.254	.136	1.87
1951–55	...	...	...	...	.217	.493	.276	1.79
1956–60	...	...	...	...	.622	1.355	.734	1.85
1961–6/68	...	...	...	...	.105	.722	.617	1.17
Panel D:								
1935–6/68	.00061	.00362	.00301	1.21	.276	.864	.588	1.47
1935–45	...	.00624	.00644	.97	.392	1.001	.613	1.63
1946–55	.00061	.00148	.00087	1.70	.028	.383	.355	1.08
1956–6/68	.00134	.00304	.00169	1.80	.374	1.125	.751	1.50
1935–40	...	.00723	.00886	.82	.120	.682	.562	1.21
1941–45	.00162	.00515	.00353	1.46	.720	1.395	.675	2.07
1946–50	.00083	.00180	.00096	1.87	.023	.348	.325	1.07
1951–55	.00039	.00116	.00077	1.51	.038	.424	.386	1.10
1956–60	.00037	.00103	.00066	1.56	.712	1.654	.941	1.76
1961–6/68	.00202	.00440	.00238	1.85	.163	.787	.624	1.26

sis.[12] The most efficient tests, however, are provided by the one-variable

[12] The least-squares intercepts $\hat{\gamma}_{0t}$ in the four cross-sectional risk-return regressions can always be interpreted as returns for month t on zero-$\hat{\beta}$ portfolios (n. 10). For the three-variable model of panel D, table 3, the unbiasedness of the least-squares coefficients can be shown to imply that in computing $\hat{\gamma}_{0t}$, negative and positive weights are assigned to the 20 portfolios in such a way that the resulting portfolio has not only zero-$\hat{\beta}$ but also zero averages of the 20 $\hat{\beta}_p^2$ and of the 20 $\bar{s}_p(\hat{\epsilon}_i)$. Analogous statements apply to the two-variable models of panels B and C.

Black, Jensen, and Scholes test the S-L hypothesis with a time series of monthly returns on a "minimum variance zero-$\hat{\beta}$ portfolio" which they derive directly. It turns model of panel A, since the values of $s(\hat{\gamma}_0)$ for this model [which are nearly identical with the values of $s(\hat{\gamma}_0 - R_f)$] are substantially smaller than those for other models. Except for the most recent period 1961–6/68, the values of $\overline{\hat{\gamma}_0 - R_f}$ in panel A are all positive and generally greater than 0.4 percent per month. The value of $t(\overline{\hat{\gamma}_0 - R_f})$ for the overall period 1935–6/68 is 2.55, and the t-statistics for the subperiods 1946–55, 1951–55, and 1956–60 are likewise large. Thus, the results in panel A, table 3, support the negative conclusions of Friend and Blume (1970) and Black, Jensen, and Scholes (1972) with respect to the S-L hypothesis.

The S-L hypothesis seems to do somewhat better in the two-variable quadratic model of panel B, table 3 and especially in the three-variable model of panel D. The values of $t(\overline{\hat{\gamma}_0 - R_f})$ are substantially closer to zero for these models than for the model of panel A. This is due to values of $\overline{\hat{\gamma}_0 - R_f}$ that are closer to zero, but it also reflects the fact that $s(\hat{\gamma}_0)$ is substantially higher for the models of panels B and D than for the model of panel A.

But the effects of $\hat{\beta}_p^2$ and $\bar{s}_p(\hat{\epsilon}_i)$ on tests of the S-L hypothesis are in fact not at all so clear-cut. Consider the model

$$\tilde{R}_{it} = \tilde{\gamma}'_{0t} + \tilde{\gamma}'_{1t}\beta_i + \tilde{\gamma}_{2t}(1 - \beta_i)^2 + \tilde{\gamma}_{3t}s_i + \tilde{\eta}_{it}. \tag{13}$$

Equations (7) and (13) are equivalent representations of the stochastic process generating returns, with $\tilde{\gamma}_{1t} = \tilde{\gamma}'_{1t} - 2\tilde{\gamma}_{2t}$ and $\tilde{\gamma}_{0t} = \tilde{\gamma}'_{0t} + \tilde{\gamma}_{2t}$. Moreover, if the steps used to obtain the regression equation (10) from the stochastic model (7) are applied to (13), we get the regression equation,

$$R_{pt} = \hat{\gamma}'_{0t} + \hat{\gamma}'_{1t}\hat{\beta}_p + \hat{\gamma}_{2t}(1 - \hat{\beta}_p)^2 + \hat{\gamma}_{3t}\bar{s}_p(\hat{\epsilon}_i) + \hat{\eta}_{pt}, \tag{14}$$

where, just as $\hat{\beta}_p^2$ in (10) is the average of $\hat{\beta}_i^2$ for securities i in portfolio p, $(1 - \hat{\beta}_p)^2$ is the average of $(1 - \hat{\beta}_i)^2$. The values of the estimates $\hat{\gamma}_{2t}$ and $\hat{\gamma}_{3t}$ are identical in (10) and (14); in addition, $\hat{\gamma}_{1t} = \hat{\gamma}'_{1t} - 2\hat{\gamma}_{2t}$ and $\hat{\gamma}_{0t} = \hat{\gamma}'_{0t} + \hat{\gamma}_{2t}$. But although the regression equations (10) and (14) are statistically indistinguishable, tests of the hypothesis $E(\tilde{\gamma}_{0t}) =$

out, however, that this portfolio is constructed under what amounts to the assumptions of the Gauss-Markov Theorem on the underlying disturbances of the one-variable risk-return regression (11). With these assumptions the least-squares estimate $\hat{\gamma}_{0t}$, obtained from the cross-sectional risk-return regression of (11) for month t, is precisely the return for month t on the minimum variance zero-$\hat{\beta}$ portfolio that can be constructed from the 20 portfolio $\hat{\beta}_p$. Thus, the tests of the S-L hypothesis in panel A of table 3 are conceptually the same as those of Black, Jensen, and Scholes.

If one makes the assumptions of the Gauss-Markov Theorem on the underlying disturbances of the models of panels B–D of table 3, the regression intercepts for these models can likewise be interpreted as returns on minimum-variance zero-$\hat{\beta}$ portfolios. These portfolios then differ in terms of whether or not they also constrain the averages of the 20 $\hat{\beta}_p^2$ and of the 20 $\bar{s}_p(\hat{\epsilon}_i)$ to be zero. Given the collinearity of $\hat{\beta}_p$, $\hat{\beta}_p^2$, and $\bar{s}_p(\hat{\epsilon}_i)$, however, the assumptions of the Gauss-Markov Theorem cannot apply to all four of the models.

R_{ft} from (10) do not yield the same results as tests of the hypothesis $E(\tilde{\gamma}'_{0t}) = R_{ft}$ from (14). In panel D of table 3, $\overline{\hat{\gamma}_0 - R_f}$ is never statistically very different from zero, whereas in tests (not shown) from (14), the results are similar to those of panel A, table 3. That is, $\overline{\hat{\gamma}'_0 - R_f}$ is systematically positive for all periods but 1961–6/68 and statistically very different from zero for the overall period 1935–6/68 and for the 1946–55, 1951–55, and 1956–60 subperiods.

Thus, tests of the S-L hypothesis from our three-variable models are ambiguous. Perhaps the ambiguity could be resolved and more efficient tests of the hypothesis could be obtained if the omitted variables for which $\bar{s}_p(\hat{\epsilon}_i)$, $\hat{\beta}_p^2$, or $(1 - \hat{\beta}_p)^2$ are almost surely proxies were identified. As indicated above, however, at the moment the most efficient tests of the S-L hypothesis are provided by the one-variable model of panel A, table 3, and the results for that model support the negative conclusions of others.

Given that the S-L hypothesis is not supported by the data, tests of the market efficiency hypothesis that $\tilde{\gamma}_{0t} - E(\tilde{R}_{0t})$ is a fair game are difficult since we no longer have a specific hypothesis about $E(\tilde{R}_{0t})$. And using the mean of the $\hat{\gamma}_{0t}$ as an estimate of $E(\tilde{R}_{0t})$ does not work as well in this case as it does for the market efficiency tests on γ_{1t}. One should note, however, that although the serial correlations $\rho_M(\hat{\gamma}_0)$ in table 4 are often large relative to estimates of their standard errors, they are small in terms of the proportion of the time series variance of $\hat{\gamma}_{0t}$ that they explain, and the latter is the more important criterion for judging whether market efficiency is a workable representation of reality (see n. 8).

VI. Conclusions

In sum our results support the important testable implications of the two-parameter model. Given that the market portfolio is efficient—or, more specifically, given that our proxy for the market portfolio is at least approximately efficient—we cannot reject the hypothesis that average returns on New York Stock Exchange common stocks reflect the attempts of risk-averse investors to hold efficient portfolios. Specifically, on average there seems to be a positive tradeoff between return and risk, with risk measured from the portfolio viewpoint. In addition, although there are "stochastic nonlinearities" from period to period, we cannot reject the hypothesis that on average their effects are zero and unpredictably different from zero from one period to the next. Thus, we cannot reject the hypothesis that in making a portfolio decision, an investor should assume that the relationship between a security's portfolio risk and its expected return is linear, as implied by the two-parameter model. We also cannot reject the hypothesis of the two-parameter model that no measure of risk, in addition to portfolio risk, systematically affects average returns. Finally, the observed fair game properties of the coefficients and residuals of the risk-return regressions are consistent with an efficient capital market—that is, a market where prices of securities fully reflect available information.

Appendix

Some Related Issues

A1. Market Models and Tests of Market Efficiency

The time series of regression coefficients from (10) are, of course, the inputs for the tests of the two-parameter model. But these coefficients can also be useful in tests of capital market efficiency—that is, tests of the speed of price adjustment to different types of new information. Since the work of Fama et al. (1969), such tests have commonly been based on the "one-factor market model":

$$R_{it} = \hat{a}_i + \hat{\beta}_i R_{mt} + \hat{\epsilon}_{it}. \tag{15}$$

In this regression equation, the term involving R_{mt} is assumed to capture the effects of market-wide factors. The effects on returns of events specific to company i, like a stock split or a change in earnings, are then studied through the residuals $\hat{\epsilon}_{it}$.

But given that there is period-to-period variation in $\hat{\gamma}_{0t}$, $\hat{\gamma}_{2t}$, and $\hat{\gamma}_{3t}$ in (10) that is above and beyond pure sampling error, then these coefficients can be interpreted as market factors, (in addition to R_{mt}) that influence the returns on all securities. To see this, substitute (12) into (11) to obtain the "two-factor market model":

$$R_{pt} = \hat{\gamma}_{0t}(1 - \hat{\beta}_p) + \hat{\beta}_p R_{mt} + \hat{\eta}_{pt}. \qquad 16$$

In like fashion, from equation (10) itself we easily obtain the "four-factor market model":

$$R_{pt} = \hat{\gamma}_{0t}(1 - \hat{\beta}_p) + \hat{\beta}_p R_{mt} + \hat{\gamma}_{2t}(\hat{\beta}^2{}_p - \hat{\beta}_p \bar{\hat{\beta}}{}^2) + \hat{\gamma}_{3t} [\bar{s}_p(\hat{\epsilon}_i) - \hat{\beta}_p \bar{\bar{s}}(\hat{\epsilon}_i)] + \hat{\eta}_{pt}, \tag{17}$$

where $\bar{\hat{\beta}}{}^2$ and $\bar{\bar{s}}(\hat{\epsilon}_i)$ are the averages over p of the $\hat{\beta}_p{}^2$ and the $\bar{s}_p(\hat{\epsilon}_i)$.

Comparing equations (15–17) it is clear that the residuals $\hat{\epsilon}_{it}$ from the one-factor market model contain variation in the market factors $\hat{\gamma}_{0t}$, $\hat{\gamma}_{2t}$, and $\hat{\gamma}_{3t}$. Thus, if one is interested in the effect on a security's return of an event specific to the given company, this effect can probably be studied more precisely from the residuals of the two- or even the four-factor market models of (16) and (17) than from the one-factor model of (15). This has in fact already been done in a study of changes in accounting techniques by Ball (1972), in a study of insider trading by Jaffe (1972), and in a study of mergers by Mandelker (1972).

Ball, Jaffe, and Mandelker use the two-factor rather than the four-factor market model, and there is probably some basis for this. First, one can see from table 5 that because of the collinearity of $\hat{\beta}_p$, $\hat{\beta}_p{}^2$, and $\bar{s}_p(\hat{\epsilon}_i)$, the coefficient estimates $\hat{\gamma}_{0t}$ and $\hat{\gamma}_{1t}$ have much smaller standard errors in the two-factor model. Second, we have computed residual variances for each of our 20 portfolios for various time periods from the time series of $\hat{\epsilon}_{pt}$ and $\hat{\eta}_{pt}$ from (15), (16), and (17). The decline in residual variance that is obtained in going from (15) to (16) is as predicted: That is, the decline is noticeable over more or less the entire range of $\hat{\beta}_p$ and it is proportional to $(1 - \hat{\beta}_p)^2$. On the other hand, in going from the two- to the four-factor model, reductions in residual variance are generally noticeable only in the portfolios with the lowest and highest $\hat{\beta}_p$, and the reductions for these two portfolios are generally small. Moreover, including $\bar{s}_p(\hat{\epsilon}_i)$ as an explanatory variable in addition to $\hat{\beta}_p$ and $\hat{\beta}_p{}^2$ never results in a noticeable reduction in residual variances.

A2. Multifactor Models and Errors in the $\hat{\beta}$

If the return-generating process is a multifactor market model, then the usual estimates of β_i from the one-factor model of (15) are not most efficient. For example, if the return-generating process is the population analog of (16), more efficient estimates of β_i could in principle be obtained from a constrained regression applied to

$$\tilde{R}_{it} - \tilde{\gamma}_{0t} = \beta_i(\tilde{R}_{mt} - \tilde{\gamma}_{0t}) + \tilde{\eta}_{it}.$$

But this approach requires the time series of the true $\tilde{\gamma}_{0t}$. All we have are estimates $\hat{\gamma}_{0t}$, themselves obtained from estimates of $\hat{\beta}_p$ from the one-factor model of (15).

It can also be shown that with a multifactor return-generating process the errors in the $\hat{\beta}$ computed from the one-factor market model of (8) and (15) are correlated across securities and portfolios. This results from the fact that if the true process is a multifactor model, the disturbances of the one-factor model are correlated across securities and portfolios. Moreover, the interdependence of the errors in the $\hat{\beta}$ is higher the farther the true β's are from 1.0. This was already noted in the discussion of table 2 where we found that the relative reduction in the standard errors of the $\hat{\beta}$'s obtained by using portfolios rather than individual securities is lower the farther $\hat{\beta}_p$ is from 1.0.

Interdependence of the errors in the $\hat{\beta}_p$ also complicates the formal analysis of the effects of errors-in-the-variables on properties of the estimated coefficients (the $\hat{\gamma}_{jt}$) in the risk-return regressions of (10). This topic is considered in detail in an appendix to an earlier version of this paper that can be made available to the reader on request.

References

Ball, R. "Changes in Accounting Techniques and Stock Prices." Ph.D. dissertation, University of Chicago, 1972.

Black, F. "Capital Market Equilibrium with Restricted Borrowing." *J. Bus.* 45 (July 1972): 444–55.

Black, F.; Jensen, M.; Scholes, M. "The Capital Asset Pricing Model: Some Empirical Results." In *Studies in the Theory of Capital Markets,* edited by Michael Jensen. New York: Praeger, 1972.

Blume, M. E. "Portfolio Theory: A Step toward Its Practical Application." *J. Bus.* 43 (April 1970): 152–73.

Douglas, G. W. "Risk in the Equity Markets: An Empirical Appraisal of Market Efficiency." *Yale Econ. Essays* 9 (Spring 1969): 3–45.

Fama, E. F. "The Behavior of Stock Market Prices." *J. Bus.* 38 (January 1965): 34–105. (*a*)

———. "Portfolio Analysis in a Stable Paretian Market." *Management Sci.* 11 (January 1965): 404–19. (*b*)

———. "Multiperiod Consumption-Investment Decisions." *A.E.R.* 60 (March 1970): 163–74. (*a*)

———. "Efficient Capital Markets: A Review of Theory and Empirical Work." *J. Finance* 25 (May 1970): 383–417. (*b*)

———. "Risk, Return and Equilibrium." *J.P.E.* 79 (January/February 1971): 30–55.

Fama, E. F., and Babiak, H. "Dividend Policy: An Empirical Analysis." *J. American Statis. Assoc.* 48 (December 1968): 1132–61.

Fama, E. F.; Fisher, L.; Jensen, M.; Roll, R. "The Adjustment of Stock Prices to New Information." *Internat. Econ. Rev.* 10 (February 1969): 1–21.

Fama, E. F., and Miller, M. *The Theory of Finance.* New York: Holt, Rinehart & Winston, 1972.

Fama, E. F., and Roll, R. "Some Properties of Symmetric Stable Distributions." *J. American Statis. Assoc.* 48 (September 1968): 817–36.

Fisher, L. "Some New Stock Market Indexes." *J. Bus.* 39 (January 1966): 191–225.

Friend, I., and Blume, M. "Measurement of Portfolio Performance under Uncertainty." *A.E.R.* 60 (September 1970): 561–75.

Gonedes, N. J. "Evidence on the Information Content of Accounting Numbers: Accounting-based and Market-based Estimates of Systematic Risk." *J. Financial and Quantitative Analysis* (1973): in press.

Jaffe, J. "Security Regulation, Special Information, and Insider Trading." Ph.D. dissertation, University of Chicago, 1972.

Jensen, M. "The Foundations and Current State of Capital Market Theory." *Bell J. Econ. and Management Sci.* 3 (Autumn 1972): 357–98.

Lintner, J. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Rev. Econ. and Statis.* 47 (February 1965): 13–37.

Mandelker, G. "Returns to Stockholders from Mergers." Ph.D. proposal, University of Chicago, 1972.

Markowitz, H. *Portfolio Selection: Efficient Diversification of Investments.* New York: Wiley, 1959.

Miller, K. D. "Futures Trading and Investor Returns: An Investigation of Commodity Market Risk Premiums." Ph.D. dissertation, University of Chicago, 1971.

Miller, M., and Scholes, M. "Rates of Return in Relation to Risk: A Re-Examination of Some Recent Findings." In *Studies in the Theory of Capital Markets,* edited by Michael Jensen. New York: Praeger, 1972.

Mood, A. M., and Graybill, F. A. *Introduction to the Theory of Statistics.* New York: McGraw-Hill, 1963.

Officer, R. R. "A Time Series Examination of the Market Factor of the New York Stock Exchange." Ph.D. dissertation, University of Chicago, 1971.

Roll, R. *The Behavior of Interest Rates.* New York: Basic, 1970.

Sharpe, W. F. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *J. Finance* 19 (September 1964): 425–42.

Tobin, J. "Liquidity Preference as Behavior towards Risk." *Rev. Econ. Studies* 25 (February 1958): 65–86.

Ibbotson® SBBI®

2010 Valuation Yearbook

Market Results for

Stocks, Bonds, Bills, and Inflation

1926–2009

MORNINGSTAR®

approximately when quality financial data became available. They also made a conscious effort to include the period of extreme market volatility from the late twenties and early thirties; 1926 was chosen because it includes one full business cycle of data before the market crash of 1929. These are the most basic reasons why our equity risk premium calculation window starts in 1926.

Implicit in using history to forecast the future is the assumption that investors' expectations for future outcomes conform to past results. This method assumes that the price of taking on risk changes only slowly, if at all, over time. This "future equals the past" assumption is most applicable to a random time-series variable. A time-series variable is random if its value in one period is independent of its value in other periods.

Does the Equity Risk Premium Revert to Its Mean Over Time?

Some have argued that the estimate of the equity risk premium is upwardly biased since the stock market is currently priced high. In other words, since there have been several years with extraordinarily high market returns and realized equity risk premia, the expectation is that returns and realized equity risk premia will be lower in the future, bringing the average back to a normalized level. This argument relies on several studies that have tried to determine whether reversion to the mean exists in stock market prices and the equity risk premium.[1] Several academics contradict each other on this topic; moreover, the evidence supporting this argument is neither conclusive nor compelling enough to make such a strong assumption.

Our own empirical evidence suggests that the yearly difference between the stock market total return and the U.S. Treasury bond income return in any particular year is random. Graph 5-2, presented earlier, illustrates the randomness of the realized equity risk premium.

A statistical measure of the randomness of a return series is its serial correlation. Serial correlation (or autocorrelation) is defined as the degree to which the return of a given series is related from period to period. A serial correlation near positive one indicates that returns are predictable from one period to the next period and are positively related. That is, the returns of one period are a good predictor of the returns in the next period. Conversely, a serial correlation near negative one indicates that the returns in one period are inversely related to those of the next period. A serial correlation near zero indicates that the returns are random or unpredictable from one period to the next. Table 5-3 contains the serial correlation of the market total returns, the realized long-horizon equity risk premium, and inflation.

Table 5-3: Interpretation of Annual Serial Correlations

Series	Serial Correlation	Interpretation
Large Company Stock Total Returns	0.02	Random
Equity Risk Premium	0.02	Random
Inflation Rates	0.64	Trend

Data from 1926–2009

The significance of this evidence is that the realized equity risk premium next year will not be dependent on the realized equity risk premium from this year. That is, there is no discernable pattern in the realized equity risk premium—it is virtually impossible to forecast next year's realized risk premium based on the premium of the previous year. For example, if this year's difference between the riskless rate and the return on the stock market is higher than last year's, that does not imply that next year's will be higher than this year's. It is as likely to be higher as it is lower. The best estimate of the expected value of a variable that has behaved randomly in the past is the average (or arithmetic mean) of its past values.

Table 5-4 also indicates that the equity risk premium varies considerably by decade. The complete decades ranged from a high of 17.9 percent in the 1950s to a low of -3.7 percent in the 2000s. This look at historical equity risk premium reveals no observable pattern.

Table 5-4: Long-Horizon Equity Risk Premium by Decade (%)

1920s*	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
17.6	2.3	8.0	17.9	4.2	0.3	7.9	12.1	-3.7

Data from 1926–2009.
*Based on the period 1926–1929.

Table 2-1: Total Returns, Income Returns, and Capital Appreciation of the Basic Asset Classes: Summary Statistics of Annual Returns

Series	Geometric Mean (%)	Arithmetic Mean (%)	Standard Deviation (%)	Serial Correl-ation
Large Company Stocks				
Total Returns	9.8	11.8	20.5	0.02
Income	4.1	4.1	1.6	0.90
Capital Appreciation	5.5	7.4	19.8	0.01
Ibbotson Small Company Stocks				
Total Returns	11.9	16.6	32.8	0.06
Mid-Cap Stocks*				
Total Returns	10.9	13.7	25.0	-0.04
Income	3.9	4.0	1.7	0.90
Capital Appreciation	6.7	9.5	24.3	-0.05
Low-Cap Stocks*				
Total Returns	11.3	15.2	29.4	0.02
Income	3.6	3.6	2.0	0.89
Capital Appreciation	7.5	11.4	28.7	0.01
Micro-Cap Stocks*				
Total Returns	12.1	18.2	39.2	0.07
Income	2.5	2.5	1.7	0.91
Capital Appreciation	9.5	15.6	38.6	0.06
Long-Term Corporate Bonds				
Total Returns	5.9	6.2	8.3	0.08
Long-Term Government Bonds				
Total Returns	5.4	5.8	9.6	-0.12
Income	5.1	5.2	2.7	0.96
Capital Appreciation	0.1	0.4	8.4	-0.26
Intermediate-Term Government Bonds				
Total Returns	5.3	5.5	5.7	0.13
Income	4.7	4.7	2.9	0.96
Capital Appreciation	0.5	0.6	4.5	-0.18
Treasury Bills				
Total Returns	3.7	3.7	3.1	0.91
Inflation	3.0	3.1	4.2	0.64

Data from 1926–2009. Total return is equal to the sum of three component returns: income return, capital appreciation return, and reinvestment return.

*Source: Morningstar and CRSP. Calculated (or Derived) based on data from CRSP US Stock Database and CRSP US Indices Database ©2010 Center for Research in Security Prices (CRSP®), The University of Chicago Booth School of Business. Used with permission.

Investor growth expectations: Analysts vs. history

Analysts' growth forecasts dominate past trends in predicting stock prices.

James H. Vander Weide and Willard T. Carleton

For the purposes of implementing the Discounted Cash Flow (DCF) cost of equity model, the analyst must know which growth estimate is embodied in the firm's stock price. A study by Cragg and Malkiel (1982) suggests that the stock valuation process embodies analysts' forecasts rather than historically based growth figures such as the ten-year historical growth in dividends per share or the five-year growth in book value per share. The Cragg and Malkiel study is based on data for the 1960s, however, a decade that was considerably more stable than the recent past.

As the issue of which growth rate to use in implementing the DCF model is so important to applications of the model, we decided to investigate whether the Cragg and Malkiel conclusions continue to hold in more recent periods. This paper describes the results of our study.

STATISTICAL MODEL

The DCF model suggests that the firm's stock price is equal to the present value of the stream of dividends that investors expect to receive from owning the firm's shares. Under the assumption that investors expect dividends to grow at a constant rate, g, in perpetuity, the stock price is given by the following simple expression:

$$P_S = \frac{D\,(1 + g)}{k - g} \tag{1}$$

where:

P_S = current price per share of the firm's stock;

D = current annual dividend per share;

g = expected constant dividend growth rate; and

k = required return on the firm's stock.

Dividing both sides of Equation (1) by the firm's current earnings, E, we obtain:

$$\frac{P_S}{E} = \frac{D}{E} \cdot \frac{(1 + g)}{k - g} \tag{2}$$

Thus, the firm's price/earnings (P/E) ratio is a non-linear function of the firm's dividend payout ratio (D/E), the expected growth in dividends (g), and the required rate of return.

To investigate what growth expectation is embodied in the firm's current stock price, it is more convenient to work with a linear approximation to Equation (2). Thus, we will assume that:

$$P/E = a_0(D/E) + a_1 g + a_2 k. \tag{3}$$

(Cragg and Malkiel found this assumption to be reasonable throughout their investigation.)

Furthermore, we will assume that the required

JAMES H. VANDER WEIDE is Research Professor at the Fuqua School of Business at Duke University in Durham (NC 27706). WILLARD T. CARLETON is Karl Eller Professor of Finance at the University of Arizona in Tucson (AZ 85721).
Financial support for this project was provided by BellSouth and Pacific Telesis. The authors wish to thank Paul Blalock at BellSouth, Mohan Gyani at Pacific Telesis, Bill Keck at Southern Bell, and John Carlson, their programmer, for help with this project.

rate of return, k, in Equation (3) depends on the values of the risk variables B, Cov, Rsq, and Sa, where B is the firm's Value Line beta; Cov is the firm's pretax interest coverage ratio; Rsq is a measure of the stability of the firm's five-year historical EPS; and Sa is the standard deviation of the consensus analysts' five-year EPS growth forecast for the firm. Finally, as the linear form of the P/E equation is only an approximation to the true P/E equation, and B, Cov, Rsq, and Sa are only proxies for k, we will add an error term, e, that represents the degree of approximation to the true relationship.

With these assumptions, the final form of our P/E equation is as follows:

$$P/E = a_0(D/E) + a_1 g + a_2 B + a_3 Cov + a_4 Rsq + a_5 Sa + e. \quad (4)$$

The purpose of our study is to use more recent data to determine which of the popular approaches for estimating future growth in the Discounted Cash Flow model is embodied in the market price of the firm's shares.

We estimated Equation (4) to determine which estimate of future growth, g, when combined with the payout ratio, D/E, and risk variables B, Cov, Rsq, and Sa, provides the best predictor of the firm's P/E ratio. To paraphrase Cragg and Malkiel, we would expect that growth estimates found in the best-fitting equation more closely approximate the expectation used by investors than those found in poorer-fitting equations.

DESCRIPTION OF DATA

Our data sets include both historically based measures of future growth and the consensus analysts' forecasts of five-year earnings growth supplied by the Institutional Brokers Estimate System of Lynch, Jones & Ryan (IBES). The data also include the firm's dividend payout ratio and various measures of the firm's risk. We include the latter items in the regression, along with earnings growth, to account for other variables that may affect the firm's stock price.

The data include:

Earnings Per Share. Because our goal is to determine which earnings variable is embodied in the firm's market price, we need to define this variable with care. Financial analysts who study a firm's financial results in detail generally prefer to "normalize" the firm's reported earnings for the effect of extraordinary items, such as write-offs of discontinued operations, or mergers and acquisitions. They also attempt, to the extent possible, to state earnings for different firms using a common set of accounting conventions.

We have defined "earnings" as the consensus analyst estimate (as reported by IBES) of the firm's earnings for the forthcoming year.[1] This definition approximates the normalized earnings that investors most likely have in mind when they make stock purchase and sell decisions. It implicitly incorporates the analysts' adjustments for differences in accounting treatment among firms and the effects of the business cycle on each firm's results of operations. Although we thought at first that this earnings estimate might be highly correlated with the analysts' five-year earnings growth forecasts, that was not the case. Thus, we avoided a potential spurious correlation problem.

Price/Earnings Ratio. Corresponding to our definition of "earnings," the price/earnings ratio (P/E) is calculated as the closing stock price for the year divided by the consensus analyst earnings forecast for the forthcoming fiscal year.

Dividends. Dividends per share represent the common dividends declared per share during the calendar year, after adjustment for all stock splits and stock dividends). The firm's dividend payout ratio is then defined as common dividends per share divided by the consensus analyst estimate of the earnings per share for the forthcoming calendar year (D/E). Although this definition has the deficiency that it is obviously biased downward — it divides this year's dividend by next year's earnings — it has the advantage that it implicitly uses a "normalized" figure for earnings. We believe that this advantage outweighs the deficiency, especially when one considers the flaws of the apparent alternatives. Furthermore, we have verified that the results are insensitive to reasonable alternative definitions (see footnote 1).

Growth. In comparing historically based and consensus analysts' forecasts, we calculated forty-one different historical growth measures. These included the following: 1) the past growth rate in EPS as determined by a log-linear least squares regression for the latest year,[2] two years, three years, . . ., and ten years; 2) the past growth rate in DPS for the latest year, two years, three years, . . ., and ten years; 3) the past growth rate in book value per share (computed as the ratio of common equity to the outstanding common equity shares) for the latest year, two years, three years, . . ., and ten years; 4) the past growth rate in cash flow per share (computed as the ratio of pretax income, depreciation, and deferred taxes to the outstanding common equity shares) for the latest year, two years, three years, . . ., and ten years; and 5) plowback growth (computed as the firm's retention ratio for the current year times the firm's latest annual return on common equity).

We also used the five-year forecast of earnings per share growth compiled by IBES and reported in mid-January of each year. This number represents the consensus (i.e., mean) forecast produced by analysts from the research departments of leading Wall Street and regional brokerage firms over the preceding three months. IBES selects the contributing brokers "because of the superior quality of their research, professional reputation, and client demand" (IBES *Monthly Summary Book*).

Risk Variables. Although many risk factors could potentially affect the firm's stock price, most of these factors are highly correlated with one another. As shown above in Equation (4), we decided to restrict our attention to four risk measures that have intuitive appeal and are followed by many financial analysts: 1) B, the firm's beta as published by Value Line; 2) Cov, the firm's pretax interest coverage ratio (obtained from Standard & Poor's Compustat); 3) Rsq, the stability of the firm's five-year historical EPS (measured by the R^2 from a log-linear least squares regression); and 4) Sa, the standard deviation of the consensus analysts' five-year EPS growth forecast (mean forecast) as computed by IBES.

After careful analysis of the data used in our study, we felt that we could obtain more meaningful results by imposing six restrictions on the companies included in our study:

1. Because of the need to calculate ten-year historical growth rates, and because we studied three different time periods, 1981, 1982, and 1983, our study requires data for the thirteen-year period 1971-1983. We included only companies with at least a thirteen-year operating history in our study.
2. As our historical growth rate calculations were based on log-linear regressions, and the logarithm of a negative number is not defined, we excluded all companies that experienced negative EPS during any of the years 1971-1983.
3. For similar reasons, we also eliminated companies that did not pay a dividend during any one of the years 1971-1983.
4. To insure comparability of time periods covered by each consensus earnings figure in the P/E ratios, we eliminated all companies that did not have a December 31 fiscal year-end.
5. To eliminate distortions caused by highly unusual events that distort current earnings but not expected future earnings, and thus the firm's price/earnings ratio, we eliminated any firm with a price/earnings ratio greater than 50.
6. As the evaluation of analysts' forecasts is a major part of this study, we eliminated all firms that IBES did not follow.

Our final sample consisted of approximately sixty-five utility firms.[3]

RESULTS

To keep the number of calculations in our study to a reasonable level, we performed the study in two stages. In Stage 1, all forty-one historically oriented approaches for estimating future growth were correlated with each firm's P/E ratio. In Stage 2, the historical growth rate with the highest correlation to the P/E ratio was compared to the consensus analyst growth rate in the multiple regression model described by Equation (4) above. We performed our regressions for each of three recent time periods, because we felt the results of our study might vary over time.

First-Stage Correlation Study

Table 1 gives the results of our first-stage correlation study for each group of companies in each of the years 1981, 1982, and 1983. The values in this table measure the correlation between the historically oriented growth rates for the various time periods and the firm's end-of-year P/E ratio.

The four variables for which historical growth rates were calculated are shown in the left-hand column: EPS indicates historical earnings per share growth, DPS indicates historical dividend per share growth, BVPS indicates historical book value per share growth, and CFPS indicates historical cash flow per share growth. The term "plowback" refers to the product of the firm's retention ratio in the currennt year and its return on book equity for that year. In all, we calculated forty-one historically oriented growth rates for each group of firms in each study period.

The goal of the first-stage correlation analysis was to determine which historically oriented growth rate is most highly correlated with each group's year-end P/E ratio. Eight-year growth in CFPS has the highest correlation with P/E in 1981 and 1982, and ten-year growth in CFPS has the highest correlation with year-end P/E in 1983. In all cases, the plowback estimate of future growth performed poorly, indicating that — contrary to generally held views — plowback is not a factor in investor expectations of future growth.

Second-Stage Regression Study

In the second stage of our regression study, we ran the regression in Equation (4) using two different measures of future growth, g: 1) the best historically oriented growth rate (g_h) from the first-stage correlation study, and 2) the consensus analysts' forecast (g_a) of five-year EPS growth. The regression results, which are shown in Table 2, support at least

TABLE 1

Correlation Coefficients of All Historically Based Growth Estimates by Group and by Year with P/E

Historical Growth Rate Period in Years

Current Year	1	2	3	4	5	6	7	8	9	10
1981										
EPS	−0.02	0.07	0.03	0.01	0.03	0.12	0.08	0.09	0.09	0.09
DPS	0.05	0.18	0.14	0.15	0.14	0.15	0.19	0.23	0.23	0.23
BVPS	0.01	0.11	0.13	0.13	0.16	0.18	0.15	0.15	0.15	0.15
CFPS	−0.05	0.04	0.13	0.22	0.28	0.31	0.30	0.31	−0.57	−0.54
Plowback	0.19									
1982										
EPS	−0.10	−0.13	−0.06	−0.02	−0.02	−0.01	−0.03	−0.03	0.00	0.00
DPS	−0.19	−0.10	0.03	0.05	0.07	0.08	0.09	0.11	0.13	0.13
BVPS	0.07	0.08	0.11	0.11	0.09	0.10	0.11	0.11	0.09	0.09
CFPS	−0.02	−0.08	0.00	0.10	0.16	0.19	0.23	0.25	0.24	0.07
Plowback	0.04									
1983										
EPS	−0.06	−0.25	−0.25	−0.24	−0.16	−0.11	−0.05	0.00	0.02	0.02
DPS	0.03	−0.10	−0.03	0.08	0.15	0.21	0.21	0.21	0.22	0.24
BVPS	0.03	0.10	0.04	0.09	0.15	0.16	0.19	0.21	0.22	0.21
CFPS	−0.08	0.01	0.02	0.08	0.20	0.29	0.35	0.38	0.40	0.42
Plowback	−0.08									

two general conclusions regarding the pricing of equity securities.

First, we found overwhelming evidence that the consensus analysts' forecast of future growth is superior to historically oriented growth measures in predicting the firm's stock price. In every case, the R^2 in the regression containing the consensus analysts' forecast is higher than the R^2 in the regression containing the historical growth measure. The regression coefficients in the equation containing the consensus analysts' forecast also are considerably more significant than they are in the alternative regression. These results are consistent with those found by Cragg and Malkiel for data covering the period 1961-1968. Our results also are consistent with the hypothesis that investors use analysts' forecasts, rather than historically oriented growth calculations, in making stock buy–and–sell decisions.

TABLE 2

Regression Results
Model I

Part A: *Historical*

$P/E = a_0 + a_1 D/E + a_2 g_h + a_3 B + a_4 Cov + a_5 Rsq + a_6 Sa$

Year	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	$\hat{a}_6$	R^2	F Ratio
1981	−6.42*	10.31*	7.67*	3.24	0.54*	1.42*	57.43	0.83	46.49
	(5.50)	(14.79)	(2.20)	(2.86)	(2.50)	(2.85)	(4.07)		
1982	−2.90*	9.32*	8.49*	2.85	0.45*	−0.42	3.63	0.86	65.53
	(2.75)	(18.52)	(4.18)	(2.83)	(2.60)	(0.05)	(0.26)		
1983	−5.96*	10.20*	19.78*	4.85	0.44*	0.33	32.49	0.82	45.26
	(3.70)	(12.20)	(4.83)	(2.95)	(1.89)	(0.50)	(1.29)		

Part B: *Analysis*

$P/E = a_0 + a_1 D/E + a_2 g_a + a_3 B + a_4 Cov + a_5 Rsq + a_6 Sa$

Year	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	$\hat{a}_6$	R^2	F Ratio
1981	−4.97*	10.62*	54.85*	−0.61	0.33*	0.63*	4.34	0.91	103.10
	(6.23)	(21.57)	(8.56)	(0.68)	(2.28)	(1.74)	(0.37)		
1982	−2.16*	9.47*	50.71*	−1.07	0.36*	−0.31	119.05*	0.90	97.62
	(2.59)	(22.46)	(9.31)	(1.14)	(2.53)	(1.09)	(1.60)		
1983	−8.47*	11.96*	79.05*	2.16	0.56*	0.20	−34.43	0.87	69.81
	(7.07)	(16.48)	(7.84)	(1.55)	(3.08)	(0.38)	(1.44)		

Notes:
* Coefficient is significant at the 5% level (using a one-tailed test) and has the correct sign. T-statistic in parentheses.

Second, there is some evidence that investors tend to view risk in traditional terms. The interest coverage variable is statistically significant in all but one of our samples, and the stability of the operating income variable is statistically significant in six of the twelve samples we studied. On the other hand, the beta is never statistically significant, and the standard deviation of the analysts' five-year growth forecasts is statistically significant in only two of our twelve samples. This evidence is far from conclusive, however, because, as we demonstrate later, a significant degree of cross-correlation among our four risk variables makes any general inference about risk extremely hazardous.

Possible Misspecification of Risk

The stock valuation theory says nothing about which risk variables are most important to investors. Therefore, we need to consider the possibility that the risk variables of our study are only proxies for the "true" risk variables used by investors. The inclusion of proxy variables may increase the variance of the parameters of most concern, which in this case are the coefficients of the growth variables.[4]

To allow for the possibility that the use of risk proxies has caused us to draw incorrect conclusions concerning the relative importance of analysts' growth forecasts and historical growth extrapolations, we have also estimated Equation (4) with the risk variables excluded. The results of these regressions are shown in Table 3.

Again, there is overwhelming evidence that the consensus analysts' growth forecast is superior to the historically oriented growth measures in predicting the firm's stock price. The R^2 and t-statistics are higher in every case.

CONCLUSION

The relationship between growth expectations and share prices is important in several major areas of finance. The data base of analysts' growth forecasts collected by Lynch, Jones & Ryan provides a unique opportunity to test the hypothesis that investors rely more heavily on analysts' growth forecasts than on historical growth extrapolations in making security buy–and–sell decisions. With the help of this data base, our studies affirm the superiority of analysts' forecasts over simple historical growth extrapolations in the stock price formation process. Indirectly, this finding lends support to the use of valuation models whose input includes expected growth rates.

TABLE 3

Regression Results
Model II

Part A: *Historical*

$P/E = a_0 + a_1 D/E + a_2 g_h$

Year	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	R^2	F Ratio
1981	−1.05	9.59	21.20	0.73	82.95
	(1.61)	(12.13)	(7.05)		
1982	0.54	8.92	12.18	0.83	167.97
	(1.38)	(17.73)	(6.95)		
1983	−0.75	8.92	12.18	0.77	107.82
	(1.13)	(12.38)	(7.94)		

Part B: *Analysis*

$P/E + a_0 + a_1 D/E + a_2 g_a$

Year	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	R^2	F Ratio
1981	3.96	10.07	60.53	0.90	274.16
	(8.31)	(8.31)	(20.91)	(15.79)	
1982	−1.75	9.19	44.92	0.88	246.36
	(4.00)	(4.00)	(21.35)	(11.06)	
1983	−4.97	10.95	82.02	0.83	168.28
	(6.93)	(6.93)	(15.93)	(11.02)	

Notes:
* Coefficient is significant at the 5% level (using a one-tailed test) and has the correct sign. T-statistic in parentheses.

[1] We also tried several other definitions of "earnings," including the firm's most recent primary earnings per share prior to any extraordinary items or discontinued operations. As our results were insensitive to reasonable alternative definitions of "earnings " we report only the results for the IBES consensus.

[2] For the latest year, we actually employed a point-to-point growth calculation because there were only two available observations.

[3] We use the word "approximately," because the set of available firms varied each year. In any case, the number varied only from zero to three firms on either side of the figures cited here.

[4] See Maddala (1977).

REFERENCES

Bower, R. S., and D. H. Bower. "Risk and the Valuation of Common Stock." *Journal of Political Economy,* May-June 1969, pp. 349-362.

Cragg, J. G., and Malkiel, B. G. "The Consensus and Accuracy of Some Predictions of the Growth of Corporate Earnings." *Journal of Finance,* March 1968, pp. 67-84.

Cragg, J. G., and Malkiel, B. G. *Expectations and the Structure of Share Prices.* Chicago: University of Chicago Press, 1982.

Elton, E. J., M. J. Gruber, and Mustava N. Gultekin. "Expectations and Share Prices." *Management Science,* September 1981, pp. 975-987.

Federal Communications Commission. *Notice of Proposed Rulemaking.* CC Docket No. 84-800, August 13, 1984.

IBES *Monthly Summary Book.* New York: Lynch, Jones & Ryan, various issues.

Maddala, G. E. *Econometrics.* New York: McGraw-Hill Book Company, 1977.

Malkiel, B. G. "The Valuation of Public Utility Equities." *Bell Journal of Economics and Management Science,* Spring 1970, pp. 143-160.

Peterson, D., and P. Peterson. "The Effect of Changing Expectations upon Stock Returns." *Journal of Financial and Quantitative Analysis,* September 1982, pp. 799-813.

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

INVESTOR GROWTH EXPECTATIONS

Summer 2004

A study done by Vander Weide and Carleton in 1988[1] suggests that consensus analysts' forecast of future growth is superior to historically oriented growth measures in stock valuation process for domestic companies. We worked with one of the original authors of the study, Dr. James H. Vander Weide, and closely followed his suggestions and methodology to investigate whether the results still hold in more recent times (2001- 2003).

We used the following equation to determine which estimate of future growth (g) best predicts the firm's P/E ratio when combined with the dividend payout ratio, D/E, and risk variables, B, Cov, Stb, and Sa.

$$P/E = a_0(D/E) + a_1 g(\text{Growth}) + a_2 B(\text{Beta}) + a_3 Cov(\text{Interest Coverage Ratio}) + a_4 Stb(\text{Stability}) + a_5 Sa(\text{Std Dev}) + e$$

Data Description

Earnings Per Share: IBES consensus analyst estimate of the firm's earnings for the unreported year.

Price/Earnings Ratio: Closing stock price for the year divided by the consensus analyst earnings per share for the forthcoming year.

Dividends: Ratio of common dividends per share to the consensus analyst earnings forecast for the forthcoming fiscal year (D/E).

Historical Growth measures

EPS Growth Rate: Determined by a log-linear least squares regression for the latest year, two years, three years, …, and ten years.

Dividend per Share Growth Rate: Determined by a log-linear least squares regression for the latest year, two years, three years, …, and ten years.

Book Value per Share Growth Rate: Common equity divided by the common shares outstanding. Determined by a log-linear least squares regression for the latest year, two years, three years, …, and ten years.

Cash Flow per Share Growth Rate: Ratio of gross cash flow to common shares outstanding. Determined by a log-linear least squares regression for the latest year, two years, three years, …, and ten years.

Plowback Growth: Firm's retention ratio for the current year times the firm's latest annual return on equity.

3yr Plowback Growth: Firm's three-year average retention ratio times the firm's three-year average return on equity.

Consensus Analysts' Forecasts

Five-Year Earnings Per Share Growth: Mean analysts' forecast compiled by IBES.

[1] Vander Weide, J. H., and W. T. Carleton. "Investor Growth Expectations: Analysts vs. History." *The Journal of Portfolio Management*, Spring 1988, pp. 78-82.

Risk Variables

B: Beta, the firm's beta versus NYSE from Value Line.

Cov: The firm's pretax interest coverage ratio from Compustat.

Stb: Five-year historical earnings per share stability. Average absolute percentage difference between actual reported EPS and a 5yr historical EPS growth trend line from IBES.

Sa: The standard deviation of earnings per share estimate for the fiscal year from IBES.

We set five restrictions on the companies included in the study in order to be consistent with the original study and to obtain more meaningful results.

- Excluded all firms that IBES did not follow.
- Eliminated companies with:
  - Negative EPS during any of the years 1991-2003.
  - No dividend during any one of the years 1991-2003.
  - P/E ratio greater than 60 in years 2001-2003.
  - Less than five years of operating history.

The final universe consisted of 411 US firms, fifty-nine of which are utility companies.

Results

The study was performed in two stages.

Stage 1

In order to determine which historically oriented growth measure is most highly correlated with each firm's end-of-year P/E ratio, we computed spearman (rank) correlations between all forty-two historically oriented future growth measures and P/E.

The result of the stage 1 study is displayed in Table 1. Three-year plowback ratio has the highest correlation with P/E in 2001 and 2002, and five-year EPS growth rate has the highest correlation with P/E in 2003.

Table 1

Stage1 Results for Utility and Non-Utility Companies Combined

Correlations between Historically Based Growth Estimates by Year with P/E

Current Year		y1	y2	y3	y4	y5	y6	y7	y8	y9	y10
2001	EPS	0.232	0.210	0.145	0.122	0.059	0.034	-0.007	-0.076	-0.117	-0.154
	DPS	-0.243	-0.297	-0.296	-0.293	-0.313	-0.316	-0.336	-0.334	-0.329	-0.333
	BVPS	0.059	-0.017	-0.098	-0.138	-0.150	-0.182	-0.219	-0.259	-0.271	-0.273
	CFPS	0.092	0.092	0.087	0.042	-0.063	-0.102	-0.141	-0.193	-0.237	-0.262
	plowback	0.203									
	plowback3	0.308									
2002	EPS	-0.007	0.147	0.076	0.080	0.083	0.050	0.030	-0.018	-0.060	-0.089
	DPS	-0.126	-0.202	-0.251	-0.224	-0.215	-0.239	-0.232	-0.233	-0.211	-0.198
	BVPS	-0.036	-0.036	-0.078	-0.115	-0.114	-0.127	-0.152	-0.162	-0.175	-0.171
	CFPS	0.056	0.045	0.017	0.021	0.030	-0.024	-0.050	-0.080	-0.125	-0.162
	plowback	0.093									
	plowback3	0.180									
2003	EPS	0.073	0.084	0.214	0.231	0.244	0.228	0.182	0.158	0.104	0.049
	DPS	0.120	0.054	-0.001	-0.078	-0.090	-0.126	-0.152	-0.165	-0.183	-0.185
	BVPS	0.097	0.076	0.067	0.036	-0.045	-0.062	-0.063	-0.083	-0.105	-0.131
	CFPS	0.146	0.196	0.243	0.239	0.206	0.178	0.107	0.089	0.039	-0.022
	plowback	-0.017									
	plowback3	0.038									

We also independently examined utility and non-utility firms. Table 2 shows the result for the fifty-nine utility firms. Two-year growth in EPS has the highest correlation with P/E in 2001, four-year EPS has the highest correlation in 2002, and six-year EPS has the highest correlation in 2003.

Table 3 exhibits the result for the remaining non-utility firms. EPS one-year growth, two-year growth, and five-year growth has the highest correlation with P/E in 2001, 2002, and 2003, respectively.

Table 2

Stage1 Results for Utility Companies

Correlations between Historically Based Growth Estimates by Year with P/E

Current Year		y1	y2	y3	y4	y5	y6	y7	y8	y9	y10
	EPS	0.305	0.330	0.305	0.319	0.238	0.157	0.129	0.107	0.079	0.048
	DPS	-0.215	-0.321	-0.302	-0.294	-0.316	-0.281	-0.332	-0.414	-0.435	-0.429
2001	BVPS	0.164	0.137	0.147	-0.027	-0.072	-0.135	-0.117	-0.104	-0.106	-0.140
	CFPS	0.194	0.135	0.020	-0.018	-0.122	-0.157	-0.135	-0.134	-0.103	-0.219
	plowback	-0.143									
	plowback3	-0.027									
	EPS	-0.065	0.044	0.069	0.119	0.071	0.004	-0.038	-0.069	-0.061	-0.070
	DPS	-0.333	-0.327	-0.278	-0.313	-0.280	-0.321	-0.277	-0.226	-0.203	-0.210
2002	BVPS	-0.325	-0.239	-0.182	-0.177	-0.230	-0.237	-0.250	-0.247	-0.235	-0.235
	CFPS	-0.205	-0.132	-0.172	-0.166	-0.216	-0.289	-0.285	-0.265	-0.227	-0.218
	plowback	-0.151									
	plowback3	-0.133									
	EPS	0.010	0.136	0.186	0.263	0.365	0.367	0.344	0.343	0.309	0.302
	DPS	0.151	-0.029	-0.014	-0.022	-0.054	-0.117	-0.142	-0.137	-0.105	-0.092
2003	BVPS	0.212	0.060	0.047	0.019	0.003	0.040	0.022	0.005	0.003	-0.002
	CFPS	0.222	-0.046	0.173	0.115	0.165	0.100	0.017	0.077	0.057	0.077
	plowback	-0.365									
	plowback3	-0.403									

Table 3

Stage1 Results for Non-Utility Companies

Correlations between Historically Based Growth Estimates by Year with P/E

Current Year		y1	y2	y3	y4	y5	y6	y7	y8	y9	y10
	EPS	0.1843	0.1660	0.1293	0.1218	0.0873	0.0829	0.0618	0.0106	-0.0194	-0.0412
	DPS	-0.2036	-0.2211	-0.2042	-0.1935	-0.2098	-0.2066	-0.2186	-0.2155	-0.2046	-0.1975
2001	BVPS	0.0757	0.0084	-0.0791	-0.0997	-0.0916	-0.1146	-0.1388	-0.1783	-0.1866	-0.1823
	CFPS	0.0864	0.0710	0.0956	0.0704	-0.0033	-0.0162	-0.0366	-0.0747	-0.1186	-0.1325
	plowback	0.0781									
	plowback3	0.1781									
	EPS	0.0762	0.1767	0.0755	0.0817	0.0936	0.0757	0.0708	0.0316	-0.0011	-0.0254
	DPS	-0.0804	-0.1693	-0.2103	-0.1672	-0.1519	-0.1720	-0.1645	-0.1636	-0.1394	-0.1226
2002	BVPS	0.0527	0.0236	-0.0363	-0.0777	-0.0710	-0.0753	-0.0953	-0.1019	-0.1118	-0.1061
	CFPS	0.0905	0.0488	0.0143	0.0237	0.0563	0.0246	0.0097	-0.0079	-0.0458	-0.0821
	plowback	0.0634									
	plowback3	0.1306									
	EPS	0.1254	0.1783	0.2788	0.2689	0.2791	0.2622	0.2219	0.2039	0.1559	0.1090
	DPS	0.1810	0.1290	0.0655	-0.0128	-0.0101	-0.0400	-0.0630	-0.0772	-0.0930	-0.0952
2003	BVPS	0.1555	0.1740	0.1534	0.1056	0.0127	-0.0069	-0.0054	-0.0218	-0.0416	-0.0636
	CFPS	0.1479	0.2200	0.2512	0.2429	0.2004	0.1839	0.1349	0.1286	0.0892	0.0388
	plowback	-0.1109									
	plowback3	-0.0402									

Stage 2
We compared the multiple regression model of historical growth rate with the highest correlation to the P/E ratio from stage 1 to the five-year earnings per share growth forecast.

$$P/E = a_0(D/E) + a_1 g + a_2 B + a_3 Cov + a_4 Stb + a_5 Sa + e$$

The regression results are displayed in table 4. The results show that the consensus analysts' forecast of future growth better approximates the firm's P/E ratio, which is consistent with the results found by Vander Weide and Carleton. In both regressions, R^2 in the regression with the consensus analysts' forecast is higher than the R^2 in the regression with the historical growth.

Table 4
Stage2 Results for Utility and Non-Utility Companies Combined
Multiple Regression Results
P/E = a0 + a1 **D/E** + a2 **g** + a3 **B** + a4 **Cov** + a5 **Stb** + a6 **Sa**

Historical

	a0	a1	a2	a3	a4	a5	a6	Rsq	F Ratio
2001	10.43	8.46	10.79	6.79	0.02	-0.03	-18.83	0.20	13.90
	4.73	5.53	2.93	3.54	3.05	-3.06	-3.32		
2002	12.36	7.60	6.66	1.01	0.00	0.01	-32.48	0.15	9.46
	7.21	6.18	2.61	0.66	1.57	1.48	-4.04		
2003	13.34	5.96	9.87	5.27	0.01	-0.01	-20.46	0.24	17.61
	7.29	4.04	2.95	3.39	3.62	-1.31	-4.25		

Analysts' Forecasts

	a0	a1	a2	a3	a4	a5	a6	Rsq	F Ratio
2001	-1.26	16.14	144.75	-0.64	0.01	-0.03	-10.76	0.47	48.00
	-0.62	11.63	13.22	-0.38	3.07	-4.04	-2.29		
2002	3.37	13.37	106.07	-3.60	0.00	0.01	-21.85	0.35	29.73
	1.93	10.97	10.59	-2.57	1.25	1.50	-3.06		
2003	4.77	12.76	61.93	4.38	0.01	0.00	-19.41	0.33	26.38
	2.65	9.48	7.25	3.01	2.45	-0.81	-4.33		

*T-stats below the coefficients in smaller font

For utility companies shown in table 5, consensus analysts' forecast of future growth is superior to historically oriented growth in 2002 and 2003. R^2 is lower in the regression with the consensus analysts' forecast in 2001. For non-utility companies, we found that consensus analysts' forecast of future growth is superior to the alternative in all three years (table 6).

Table 5
Stage2 Results for Utility Companies

Multiple Regression Results

P/E = a0 + a1 **D/E** + a2 **g** + a3 **B** + a4 **Cov** + a5 **Stb** + a6 **Sa**

Historical

	a0	a1	a2	a3	a4	a5	a6	Rsq	F Ratio
2001	7.90	11.07	-11.19	-3.00	0.29	0.00	-9.37	0.44	6.38
	2.16	4.80	-5.71	-0.86	0.88	0.64	-1.51		
2002	13.87	7.00	-3.80	-6.89	0.56	0.00	-29.89	0.38	5.11
	4.02	3.54	-0.66	-2.01	1.48	0.42	-2.70		
2003	11.29	7.74	-1.65	-1.40	0.32	0.00	-5.69	0.25	2.68
	3.22	3.30	-0.23	-0.43	1.05	-0.73	-0.75		

Analysts' Forecasts

	a0	a1	a2	a3	a4	a5	a6	Rsq	F Ratio
2001	9.61	9.20	66.61	-7.92	0.50	-0.01	-12.83	0.27	2.95
	2.31	3.45	3.66	-1.86	1.31	-1.33	-1.76		
2002	12.43	7.86	50.74	-9.61	0.50	0.00	-24.94	0.48	7.56
	3.89	5.29	3.10	-2.94	1.50	0.17	-2.41		
2003	5.81	11.06	101.12	-1.69	-0.19	0.00	-4.75	0.50	7.81
	1.89	6.32	4.80	-0.58	-0.74	-0.22	-0.74		

*T-stats below the coefficients in smaller font

Table 6
Stage2 Results for Non-Utility Companies

Multiple Regression Results

P/E = a0 + a1 **D/E** + a2 **g** + a3 **B** + a4 **Cov** + a5 **Stb** + a6 **Sa**

Historical

	a0	a1	a2	a3	a4	a5	a6	Rsq	F Ratio
2001	15.90	8.39	2.82	3.53	0.02	-0.03	-21.05	0.21	12.45
	6.57	4.13	1.96	1.68	2.97	-2.14	-3.40		
2002	17.76	8.46	6.02	-3.06	0.00	0.02	-36.97	0.27	16.78
	9.39	5.19	3.28	-1.88	1.37	2.52	-4.31		
2003	14.24	9.86	8.85	3.46	0.01	0.00	-19.00	0.30	19.89
	7.49	5.89	2.49	2.11	3.23	-0.15	-3.73		

Analysts' Forecasts

	a0	a1	a2	a3	a4	a5	a6	Rsq	F Ratio
2001	-0.51	17.28	140.84	-1.06	0.01	-0.03	-8.63	0.44	36.00
	-0.22	11.21	10.73	-0.59	2.88	-2.62	-1.63		
2002	5.05	15.67	91.22	-4.06	0.00	0.02	-22.93	0.38	27.65
	2.48	11.23	7.66	-2.74	1.18	2.33	-2.87		
2003	7.25	14.47	45.60	3.47	0.01	0.00	-19.09	0.33	22.30
	3.56	9.42	4.68	2.20	2.36	-0.12	-3.89		

*T-stats below the coefficients in smaller font

THE JOURNAL OF FINANCE • VOL. XLVIII, NO. 1 • MARCH 1993

Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency

NARASIMHAN JEGADEESH and SHERIDAN TITMAN*

ABSTRACT

This paper documents that strategies which buy stocks that have performed well in the past and sell stocks that have performed poorly in the past generate significant positive returns over 3- to 12-month holding periods. We find that the profitability of these strategies are not due to their systematic risk or to delayed stock price reactions to common factors. However, part of the abnormal returns generated in the first year after portfolio formation dissipates in the following two years. A similar pattern of returns around the earnings announcements of past winners and losers is also documented.

A POPULAR VIEW HELD by many journalists, psychologists, and economists is that individuals tend to overreact to information.[1] A direct extension of this view, suggested by De Bondt and Thaler (1985, 1987), is that stock prices also overreact to information, suggesting that contrarian strategies (buying past losers and selling past winners) achieve abnormal returns. De Bondt and Thaler (1985) show that over 3- to 5-year holding periods stocks that performed poorly over the previous 3 to 5 years achieve higher returns than stocks that performed well over the same period. However, the interpretation of the De Bondt and Thaler results are still being debated. Some have argued that the De Bondt and Thaler results can be explained by the systematic risk of their contrarian portfolios and the size effect.[2] In addition, since the long-term losers outperform the long-term winners only in Januaries, it is unclear whether their results can be attributed to overreaction.

*Jegadeesh is from the Anderson Graduate School of Management, UCLA. Titman is from Hong Kong University of Science and Technology and the Anderson Graduate School of Management, UCLA. We would like to thank Kent Daniel, Ravi Jagannathan, Richard Roll, Hans Stoll, René Stulz, and two referees. We also thank participants of the Johnson Symposium held at the University of Wisconsin at Madison and seminar participants at Harvard, SMU, UBC, UCLA, Penn State, University of Michigan, University of Minnesota, and York University for helpful comments, and Juan Siu and Kwan Ho Kim for excellent research assistance.

[1]See for example, the academic papers by Kahneman and Tversky (1982), De Bondt and Thaler (1985) and Shiller (1981).

[2]See for example, Chan (1988), Ball and Kothari (1989), and Zarowin (1990). For an alternate view, see the recent paper by Chopra, Lakonishok, and Ritter (1992).

More recent papers by Jegadeesh (1990) and Lehmann (1990) provide evidence of shorter-term return reversals. These papers show that contrarian strategies that select stocks based on their returns in the previous week or month generate significant abnormal returns. However, since these strategies are transaction intensive and are based on short-term price movements, their apparent success may reflect the presence of short-term price pressure or a lack of liquidity in the market rather than overreaction. Jegadeesh and Titman (1991) provide evidence on the relation between short-term return reversals and bid-ask spreads that supports this interpretation. In addition, Lo and MacKinlay (1990) argue that a large part of the abnormal returns documented by Jegadeesh and Lehmann is attributable to a delayed stock price reaction to common factors rather than to overreaction.

Although contrarian strategies have received a lot of attention in the recent academic literature, the early literature on market efficiency focused on relative strength strategies that buy past winners and sell past losers. Most notably, Levy (1967) claims that a trading rule that buys stocks with current prices that are substantially higher than their average prices over the past 27 weeks realizes significant abnormal returns. Jensen and Bennington (1970), however, point out that Levy had come up with his trading rule after examining 68 different trading rules in his dissertation and because of this express skepticism about his conclusions. Jensen and Bennington analyze the profitability of Levy's trading rule over a long time period that was, for the most part, outside Levy's original sample period. They find that in their sample period Levy's trading rule does not outperform a buy and hold strategy and hence attribute Levy's result to a selection bias.

Although the current academic debate has focused on contrarian rather than relative strength trading rules, a number of practitioners still use relative strength as one of their stock selection criteria. For example, a majority of the mutual funds examined by Grinblatt and Titman (1989, 1991) show a tendency to buy stocks that have increased in price over the previous quarter. In addition, the Value Line rankings are known to be based in large part on past relative strength. The success of many of the mutual funds in the Grinblatt and Titman sample and the predictive power of Value Line rankings (see Copeland and Mayers (1982) and Stickel (1985)) provide suggestive evidence that the relative strength strategies may generate abnormal returns.

How can we reconcile the success of Value Line rankings and the mutual funds that use relative strength rules with the current academic literature that suggests that the opposite strategy generates abnormal returns? One possibility is that the abnormal returns realized by these practitioners are either spurious or are unrelated to their tendencies to buy past winners. A second possibility is that the discrepancy is due to the difference between the time horizons used in the trading rules examined in the recent academic papers and those used in practice. For instance, the above cited evidence favoring contrarian strategies focuses on trading strategies based on either very short-term return reversals (1 week or 1 month), or very long-term return reversals (3 to 5 years). However, anecdotal evidence suggests that practitioners who use relative strength rules base their selections on price movements over the past 3 to 12 months.[3] This paper provides an analysis of relative strength trading strategies over 3- to 12-month horizons. Our analysis of NYSE and AMEX stocks documents significant profits in the 1965 to 1989 sample period for each of the relative strength strategies examined. We provide a decomposition of these profits into different sources and develop tests that allow us to evaluate their relative importance. The results of these tests indicate that the profits are not due to the systematic risk of the trading strategies. In addition, the evidence indicates that the profits cannot be attributed to a lead-lag effect resulting from delayed stock price reactions to information about a common factor similar to that proposed by Lo and MacKinlay (1990). The evidence is, however, consistent with delayed price reactions to firm-specific information.

Further tests suggest that part of the predictable price changes that occur during these 3- to 12-month holding periods may not be permanent. The stocks included in the relative strength portfolios experience negative abnormal returns starting around 12 months after the formation date and continuing up to the thirty-first month. For example, the portfolio formed on the basis of returns realized in the past 6 months generates an average cumulative return of 9.5% over the next 12 months but loses more than half of this return in the following 24 months.

Our analysis of stock returns around earnings announcement dates suggests a similar bias in market expectations. We find that past winners realize consistently higher returns around their earnings announcements in the 7 months following the portfolio formation date than do past losers. However, in each of the following 13 months past losers realize higher returns than past winners around earnings announcements.

The rest of this paper is organized as follows: Section I describes the trading strategies that we examine and Section II documents their excess returns. Section III provides a decomposition of the profits from relative strength strategies and evaluates the relative importance of the different components. Section IV documents these returns in subsamples stratified on the basis of ex ante beta and firm size and Section V measures these profits across calendar months and over 5-year subperiods. The longer term performance of the stocks included in the relative strength portfolios is examined in Section VI and Section VII back tests the strategy over the 1927 to 1964

[3]For instance, one of the inputs used by Value Line to assign a timeliness rank for each stock is a price momentum factor computed based on the stock's past 3- to 12-month returns. Value Line reports that the price momentum factor is computed by "dividing the stock's latest 10-week average relative price by its 52-week average relative price." These timeliness ranks, according to Value Line, are "designed to discriminate among stocks on the basis of relative price performance over the next 6 to 12 months" (see Bernard (1984), pp. 52–53).

period. Section VIII examines the returns of past winners and past losers around earnings announcement dates and Section IX concludes the paper.

I. Trading Strategies

If stock prices either overreact or underreact to information, then profitable trading strategies that select stocks based on their past returns will exist. This study investigates the efficiency of the stock market by examining the profitability of a number of these strategies. The strategies we consider select stocks based on their returns over the past 1, 2, 3, or 4 quarters. We also consider holding periods that vary from 1 to 4 quarters. This gives a total of 16 strategies. In addition, we examine a second set of 16 strategies that skip a week between the portfolio formation period and the holding period. By skipping a week, we avoid some of the bid-ask spread, price pressure, and lagged reaction effects that underlie the evidence documented in Jegadeesh (1990) and Lehmann (1990).

To increase the power of our tests, the strategies we examine include portfolios with overlapping holding periods. Therefore, in any given month t, the strategies hold a series of portfolios that are selected in the current month as well as in the previous $K - 1$ months, where K is the holding period. Specifically, a strategy that selects stocks on the basis of returns over the past J months and holds them for K months (we will refer to this as a J-month/K-month strategy) is constructed as follows: At the beginning of each month t the securities are ranked in ascending order on the basis of their returns in the past J months. Based on these rankings, ten decile portfolios are formed that equally weight the stocks contained in the top decile, the second decile, and so on. The top decile portfolio is called the "losers" decile and the bottom decile is called the "winners" decile. In each month t, the strategy buys the winner portfolio and sells the loser portfolio, holding this position for K months. In addition, the strategy closes out the position initiated in month $t - K$. Hence, under this trading strategy we revise the weights on $\frac{1}{K}$ of the securities in the entire portfolio in any given month and carry over the rest from the previous month.

The profits of the above strategies were calculated for both a series of buy and hold portfolios and a series of portfolios that were rebalanced monthly to maintain equal weights. Since the returns for these two strategies were very similar (the buy and hold strategies yielded slightly higher returns) we present only the rebalanced returns which are also used in the event study presented in Section VI.

II. The Returns of Relative Strength Portfolios

This section documents the returns of the portfolio strategies described in the last section over the 1965 to 1989 period using data from the CRSP daily returns file.[4] All stocks with available returns data in the J months preceding the portfolio formation date are included in the sample from which the buy and sell portfolios are constructed.

Table I reports the average returns of the different buy and sell portfolios as well as the zero-cost, winners minus losers portfolio, for the 32 strategies described above. The returns of all the zero-cost portfolios (i.e., the returns per dollar long in this portfolio) are positive. All these returns are statistically significant except for the 3-month/3-month strategy that does not skip a week. Many of the individual t-statistics are sufficiently large to be significant even after considering the fact that we have conducted 32 separate tests. The probability of obtaining a single t-statistic as large as 4.28 (obtained with the 12-month/3-month strategy that skips a week) with 32 observations is less than 0.0006, as given by the Bonferroni inequality.[5]

The most successful zero-cost strategy selects stocks based on their returns over the previous 12 months and then holds the portfolio for 3 months. This strategy yields 1.31% per month (shown in Panel A) when there is no time lag between the portfolio formation period and the holding period and it yields 1.49% per month (shown in Panel B) when there is a 1-week lag between the formation period and the holding period.[6] The 6-month formation period produces returns of about 1% per month regardless of the holding period. These holding period returns are slightly higher when there is a 1-week lag between the formation period and the holding period (Panel B) than when the formation and holding periods are contiguous (Panel A).

Having established that the relative strength strategies are on average quite profitable, we now examine one specific strategy in detail, the 6-month/6-month strategy that does not skip a week between the portfolio formation period and the holding period. The results for this strategy are representative of the results for the other strategies.

III. Sources of Relative Strength Profits

This section presents two simple return-generating models that allow us to decompose the excess returns documented in the last section and identify the important sources of relative strength profits. The first model allows for factor-mimicking portfolio returns to be serially correlated but requires indi-

[4]The latest version of the CRSP daily returns file at the time this study was initiated covers the July 1962 to December 1989 period. Monthly returns were obtained by compounding the daily returns recorded in this data set. Since the 12-month/12-month strategy considered here requires lagged returns data over 23 months the first full calendar year for which we could examine portfolio returns is 1965.

[5]The Bonferroni inequality provides a bound for the probability of observing a t-statistic of a certain magnitude with N tests that are not necessarily independent.

[6]De Bondt and Thaler (1985) report 1-year holding period returns in their tables that are consistent with our findings here. However, they do not examine strategies based on 1-year horizons in any detail and based on their analysis of longer horizon strategies conclude that the market overreacts.

Table I

Returns of Relative Strength Portfolios

The relative strength portfolios are formed based on J-month lagged returns and held for K months. The values of J and K for the different strategies are indicated in the first column and row, respectively. The stocks are ranked in ascending order on the basis of J-month lagged returns and an equally weighted portfolio of stocks in the lowest past return decile is the *sell* portfolio and an equally weighted portfolio of the stocks in the highest return decile is the *buy* portfolio. The average monthly returns of these portfolios are presented in this table. The relative strength portfolios in Panel A are formed immediately after the lagged returns are measured for the purpose of portfolio formation. The relative strength portfolios in Panel B are formed 1 week after the lagged returns used for forming these portfolios are measured. The t-statistics are reported in parentheses. The sample period is January 1965 to December 1989.

		Panel A				Panel B			
	J	$K=$ 3	6	9	12	$K=$ 3	6	9	12
3	Sell	0.0108	0.0091	0.0092	0.0087	0.0083	0.0079	0.0084	0.0083
		(2.16)	(1.87)	(1.92)	(1.87)	(1.67)	(1.64)	(1.77)	(1.79)
3	Buy	0.0140	0.0149	0.0152	.0156	0.0156	0.0158	0.0158	0.0160
		(3.57)	(3.78)	(3.83)	(3.89)	(3.95)	(3.98)	(3.96)	(3.98)
3	Buy-sell	0.0032	0.0058	0.0061	0.0069	0.0073	0.0078	0.0074	0.0077
		(1.10)	(2.29)	(2.69)	(3.53)	(2.61)	(3.16)	(3.36)	(4.00)
6	Sell	0.0087	0.0079	0.0072	0.0080	0.0066	0.0068	0.0067	0.0076
		(1.67)	(1.56)	(1.48)	(1.66)	(1.28)	(1.35)	(1.38)	(1.58)
6	Buy	0.0171	0.0174	0.0174	0.0166	0.0179	0.0178	0.0175	0.0166
		(4.28)	(4.33)	(4.31)	(4.13)	(4.47)	(4.41)	(4.32)	(4.13)
6	Buy-sell	0.0084	0.0095	0.0102	0.0086	0.0114	0.0110	0.0108	0.0090
		(2.44)	(3.07)	(3.76)	(3.36)	(3.37)	(3.61)	(4.01)	(3.54)
9	Sell	0.0077	0.0065	0.0071	0.0082	0.0058	0.0058	0.0066	0.0078
		(1.47)	(1.29)	(1.43)	(1.66)	(1.13)	(1.15)	(1.34)	(1.59)
9	Buy	0.0186	0.0186	0.0176	0.0164	0.0193	0.0188	0.0176	0.0164
		(4.56)	(4.53)	(4.30)	(4.03)	(4.72)	(4.56)	(4.30)	(4.04)
9	Buy-sell	0.0109	0.0121	0.0105	0.0082	0.0135	0.0130	0.0109	0.0085
		(3.03)	(3.78)	(3.47)	(2.89)	(3.85)	(4.09)	(3.67)	(3.04)
12	Sell	0.0060	0.0065	0.0075	0.0087	0.0048	0.0058	0.0070	0.0085
		(1.17)	(1.29)	(1.48)	(1.74)	(0.93)	(1.15)	(1.40)	(1.71)
12	Buy	0.0192	0.0179	0.0168	0.0155	0.0196	0.0179	0.0167	0.0154
		(4.63)	(4.36)	(4.10)	(3.81)	(4.73)	(4.36)	(4.09)	(3.79)
12	Buy-sell	0.0131	0.0114	0.0093	0.0068	0.0149	0.0121	0.0096	0.0069
		(3.74)	(3.40)	(2.95)	(2.25)	(4.28)	(3.65)	(3.09)	(2.31)

vidual stocks to react instantaneously to factor realizations. This model is used to decompose relative strength profits into two components relating to systematic risk, which would exist in an efficient market, and a third component relating to firm-specific returns, which would contribute to relative strength profits only if the market were inefficient. The second return-generating model relaxes the assumption that stocks react instantaneously to the common factor. This model enables us to evaluate the possibility that the relative strength profits arise because of a lead-lag relationship in stock prices similar to that proposed by Lo and MacKinlay (1990) as a partial explanation for short horizon contrarian profits.

A. A Simple One-Factor Model

Consider the following one-factor model describing stock returns:[7]

$$\begin{aligned} r_{it} &= \mu_i + b_i f_t + e_{it}, \\ \mathrm{E}(f_t) &= 0 \\ \mathrm{E}(e_{it}) &= 0 \\ \mathrm{Cov}(e_{it}, f_t) &= 0, \qquad \forall i \\ \mathrm{Cov}(e_{it}, e_{jt-1}) &= 0, \qquad \forall i \neq j \end{aligned} \tag{1}$$

where μ_i is the unconditional expected return on security i, r_{it} is the return on security i, f_t is the unconditional unexpected return on a factor-mimicking portfolio, e_{it} is the firm-specific component of return at time t, and b_i is the factor sensitivity of security i. For the 6-month/6-month strategy that we consider in the rest of this paper the length of a period is 6 months.

The superior performance of the relative strength strategies documented in the last section implies that stocks that generate higher than average returns in one period also generate higher than average returns in the period that follows. In other words, these results imply that:

$$\mathrm{E}(r_{it} - \bar{r}_t | r_{it-1} - \bar{r}_{t-1} > 0) > 0$$

and

$$\mathrm{E}(r_{it} - \bar{r}_t | r_{it-1} - \bar{r}_{t-1} < 0) < 0,$$

where a bar above a variable denotes its cross-sectional average.

Therefore,

$$\mathrm{E}\{(r_{it} - \bar{r}_t)(r_{it-1} - \bar{r}_{t-1})\} > 0. \tag{2}$$

The above cross-sectional covariance equals the expected profits from the zero-cost contrarian trading strategy examined by Lehmann (1990) and Lo and MacKinlay (1990) that weights stocks by their past returns less the past equally weighted index returns. This weighted relative strength strategy (WRSS) is closely related to our strategy. The WRSS yields a profit of 4.5% per dollar long semiannually (t-statistic = 2.99) and the correlation between the returns of this strategy and that of the trading strategy examined in the last section is 0.95. The equally weighted decile portfolios are used in most of our empirical tests since they provide relatively more information than the WRSS. However, as the following analysis demonstrates, the closely related WRSS provides a tractable framework for analytically examining the sources of relative strength profits and evaluating the relative importance of each of these sources.

[7]Our analysis in this subsection is similar to that in Jegadeesh (1987) and Lo and MacKinlay (1990).

Given the one-factor model defined in (1), the WRSS profits given in expression (2) can be decomposed into the following three terms:

$$\mathrm{E}\{(r_{it} - \bar{r}_t)(r_{it-1} - \bar{r}_{t-1})\} = \sigma_\mu^2 + \sigma_b^2 \mathrm{Cov}(f_t, f_{t-1}) + \overline{\mathrm{Cov}}_i(e_{it}, e_{it-1}), \tag{3}$$

where σ_μ^2 and σ_b^2 are the cross-sectional variances of expected returns and factor sensitivities respectively.

The above decomposition suggests three potential sources of the relative strength profits. The first term in this expression is the cross-sectional dispersion in expected returns. Intuitively, since realized returns contain a component related to expected returns, securities that experience relatively high returns in one period can be expected to have higher than average returns in the following period. The second term is related to the potential to time the factor. If the factor portfolio returns exhibit positive serial correlation, the relative strength strategy will tend to pick stocks with high b's when the conditional expectation of the factor portfolio return is high. As the above expression demonstrates, the extent to which relative strength strategies generate profits because of the serial correlation of the factor portfolio return is a function of the cross-sectional variance of the b's. The last term in the above expression is the average serial covariance of the idiosyncratic components of security returns.

To assess whether the existence of relative strength profits imply market inefficiency, it is important to identify the sources of the profits. If the profits are due to either the first or the second term in expression (3) they may be attributed to compensation for bearing systematic risk and need not be an indication of market inefficiency. However, if the superior performance of the relative strength strategies is due to the third term, then the results would suggest market inefficiency.

B. The Average Size and Beta of Relative Strength Portfolios

This subsection considers the possibility that relative strength strategies systematically pick high-risk stocks and benefit from the first term in expression (3). Table II reports estimates of the two most common indicators of systematic risk, the post-ranking betas of the ten 6-month/6-month relative strength portfolios and the average capitalizations of the stocks in these portfolios. The betas of the extreme past returns portfolios are higher than the average beta for the full sample. In addition, since the beta of the portfolio of past losers is higher than the beta of the portfolio of past winners, the beta of the zero-cost winners minus losers portfolio is negative. The average capitalizations of the stocks in the different portfolios show that the highest and the lowest past returns portfolios consist of smaller than average stocks, with the stocks in the losers portfolios being smaller than the stocks in the winners portfolio. This evidence suggests that the observed relative strength profits are not due to the first source of profits in expression (3).

Table II

Betas and Market Capitalization of Relative Strength Portfolios

The relative strength portfolios are formed based on 6-month lagged returns and held for 6 months. The stocks are ranked in ascending order on the basis of 6-month lagged returns. The equally weighted portfolio of stocks in the lowest past return decile is portfolio P1, the equally weighted portfolio of stocks in the next decile is portfolio P2, and so on. The betas with respect to the value-weighted index and the average market capitalizations of the stocks included in these portfolios are reported here. The sample period is January 1965 to December 1989.

	Beta	Average Market Capitalization
P1	1.36	208.24
P2	1.19	480.07
P3	1.14	545.31
P4	1.11	618.85
P5	1.09	692.89
P6	1.08	702.51
P7	1.09	738.09
P8	1.12	758.87
P9	1.17	680.18
P10	1.28	495.13
P10–P1	−0.08	—

Additional evidence relating to the extent to which the dispersion in expected returns explains these profits is given in the next section.

C. The Serial Covariance of 6-Month Returns

This subsection examines the serial covariance of 6-month returns in order to assess the potential contribution of the second and third source of profits from our decomposition. Given the model expressed in (1), the serial covariance of an equally weighted portfolio of a large number of stocks is:[8]

$$\operatorname{cov}(\bar{r}_t, \bar{r}_{t-1}) = \bar{b}_i^2 \operatorname{Cov}(f_t, f_{t-1}). \tag{4}$$

If the source of relative strength profits is the serial covariance of factor-related returns then, from the above expression, the in-sample serial covariance of the equally weighted index returns is required to be positive. However, we find that the serial covariance of 6-month returns of the equally weighted index is negative (-0.0028) which, from the decomposition in expression (3), reduces the relative strength profits. This result indicates that the serial covariance of factor portfolio returns is unlikely to be the source of relative strength profits.

[8]The contribution of the serial covariances of e_{it} to the serial covariance of the equally weighted index becomes arbitrarily small as the number of stocks in the index becomes arbitrarily large.

The estimates of the serial covariance of market model residuals for individual stocks are on average positive (0.0012). This evidence suggests that the relative strength profits may arise from stocks underreacting to firm-specific information. However, this evidence is also potentially consistent with an alternative model in which some stocks react with a lag to factor realizations, and we address this possibility in the next subsection.

D. Lead-Lag Effects and Relative Strength Profits

This subsection examines whether the relative strength profits can arise from a lead-lag relationship in stock prices similar to that considered in Lo and MacKinlay (1990). In contrast to the model previously presented, the model in this subsection assumes that stocks can either overreact or underreact to the common factor but that the factor-mimicking portfolio returns are serially uncorrelated.

Consider the following return generating process:

$$r_{it} = \mu_i + b_{1i} f_t + b_{2i} f_{t-1} + e_{it}, \tag{5}$$

where b_{1i} and b_{2i} are sensitivities to the contemporaneous and lagged factor realizations. $b_{2i} > 0$ implies that stock i partly reacts to the factor with a lag as in Lo and MacKinlay and $b_{2i} < 0$ implies that the stock overreacts to contemporaneous factor realizations and this overreaction gets corrected in the subsequent period.

Given this model, the WRSS profits and the serial covariance of the equally weighted index are given by:

$$\mathrm{E}\{(r_{it} - \bar{r}_t)(r_{it-1} - \bar{r}_{t-1})\} = \sigma_\mu^2 + \delta\sigma_f^2 \tag{6}$$

and

$$\mathrm{cov}(\bar{r}_t, \bar{r}_{t-1}) = \bar{b}_1 \bar{b}_2 \sigma_f^2, \tag{7}$$

where $\bar{b}_1$ and $\bar{b}_2$ are cross-sectional averages of b_{1i} and b_{2i}, and,

$$\delta \equiv \frac{1}{N} \sum_{i=1}^{N} (b_{1i} - \bar{b}_1)(b_{2i} - \bar{b}_2).$$

From expression (6), when $\delta < 0$ the lead-lag relation has a negative effect on the profitability of the WRSS, or equivalently, a positive effect on contrarian profits as in Lo and MacKinlay. However, when $\delta > 0$, the lead-lag relation will generate positive relative strength profits. In addition, if $\bar{b}_2$ is positive (negative) then the equally weighted index returns will be positively (negatively) serially correlated. This parameter, however, does not affect the profitability of the WRSS.

If the lead-lag effect is an important source of relative strength profits, then the profit in any period will depend on the magnitude of factor portfolio return in the previous period. Formally, consider the expected WRSS profits conditional on the past factor portfolio return:

$$\mathrm{E}\{(r_{it} - \bar{r}_t)(r_{it-1} - \bar{r}_{t-1}) | f_{t-1}\} = \sigma_\mu^2 + \delta f_{t-1}^2. \tag{8}$$

In contrast, under model (1), the conditional expectation of the WRSS profits given in expression (3), assuming that the factor portfolio returns are normally distributed, is:

$$\mathrm{E}\{(r_{it} - \bar{r}_t)(r_{it-1} - \bar{r}_{t-1}) | f_{t-1}\} = \sigma_\mu^2 + \sigma_b^2 \rho f_{t-1}^2,$$

where ρ is the first order serial correlation of the factor portfolio returns.

Expression (8) implies that if the relative strength profits come entirely from the lead-lag effect in stock returns, then the magnitude of the profits should be positively related to the squared factor portfolio return in the previous period. Intuitively, if inefficient stock price reactions to factor realizations are important for the profitability of relative strength strategies, then large factor realizations should result in large WRSS profits. Alternatively, if the lead-lag effect does not contribute to the profits, then the observed negative serial covariance of the market index implies a negative relation between the magnitude of the WRSS profits and squared lagged factor portfolio returns.

To examine which of these predictions best explains the time-series variation in relative strength profits we estimate the following regression using the value-weighted index as a proxy for the factor portfolio:

$$r_{pt,6} = \alpha_i + \theta r_{mt,-6}^2 + u_{it},$$

where $r_{pt,6}$ is the 6-month return of the relative strength portfolio formed in month t based on 6-month lagged returns and $r_{mt,-6}$ is the demeaned return on the value-weighted index in the months $t-6$ through $t-1$. The estimates of θ and the corresponding autocorrelation-consistent t-statistic over the 1965 to 1989 sample period are -2.29 and -1.74 respectively. The estimates (t-statistic) of θ in the first and second half of this sample period are -2.55 (-2.65) and -1.83 (-2.52) respectively.[9] This reliably negative relation between the relative strength profits and lagged squared market returns is consistent with the model presented in the last subsection which assumed no lead-lag relationship and is inconsistent with the lead-lag model. This evidence indicates that the lead-lag effect is not an important source of relative strength profits and that the profitability of these strategies is therefore related to market underreaction to the firm-specific information.

[9] When this regression is fitted with the WRSS profits as the dependent variable, the estimate (t-statistic) of θ over 1965–1989 is -1.77 (-3.56) and the corresponding statistics in the two equal subperiods are -1.94 (-2.52) and -1.51 (-2.53).

IV. Profitability of Relative Strength Strategies Within Size- and Beta-Based Subsamples

In this section we examine the profitability of the 6-month/6-month strategy within subsamples stratified on the basis of firm size and ex ante estimates of betas. Specifically, we implement this strategy on three size-based subsamples (small, medium, and large), and three beta-based subsamples (low-beta, medium-beta, and high-beta stocks).

Measuring relative strength profits on size- and beta-based subsamples allows us to examine whether the profitability of the strategy is confined to any particular subsample of stocks. This analysis also provides additional evidence about the source of the observed relative strength profits. Since extant empirical evidence indicates that size and beta are related to both risk and expected returns,[10] the cross-sectional dispersion in expected returns should be less within these subsamples than in the full sample. Therefore, if the relative strength strategy profits are related to differences in expected returns, they will be less when they are implemented on stocks within each subsample rather than on all the stocks in the sample. The profits need not be reduced in these subsamples, however, if the profits of the strategies are due to serial covariances in idiosyncratic returns. In fact, if the profits are not factor-related, the strategies are likely to generate higher returns when they are implemented within the small-firm subsample that consists of less actively traded stocks and to generate lower returns when they are implemented within the large-firm subsample.

Table III presents the average returns of the 6-month/6-month strategy for each of the subsamples. The results in Panel A indicate that the observed abnormal returns are of approximately the same magnitude when the strategies are implemented on the various subsamples of stocks as when they are implemented on the entire sample. They do, however, appear to be somewhat related to firm size and beta; for the zero-cost, winners minus losers portfolio, the subsample with the largest firms generates lower abnormal returns than the other two subsamples and the returns in the subsamples segmented by beta are monotonically increasing in beta.[11] These findings indicate that the relative strength profits are not primarily due to the cross-sectional differences in the systematic risk of the stocks in the sample. This evidence suggests that the profits are due to the serial correlation in the firm-specific component of returns. Furthermore, these results indicate that the profitabil-

[10] See Fama and MacBeth (1973) and Banz (1981).

[11] One thing that is interesting to note here is that the average returns of low beta stocks are higher than the returns of the medium and high beta stocks. The average returns of stocks in the low, medium and high beta groups are 1.48%, 1.39%, and 1.16% respectively. These results, obtained with daily betas, should be contrasted with earlier findings of positive relations between monthly betas and average returns (e.g., Fama and MacBeth (1973)). The difference between our results using daily betas and the earlier results using monthly betas is due to the lower correlation between firm size and daily betas. Jegadeesh (1992) and Fama and French (1992) document that there is no reliable relation between monthly betas and average returns after controlling for firm size.

ity of the relative strength strategies is not confined to any particular subsample of stocks.

As a further test Panel B of Table III presents the risk-adjusted returns of the relative strength strategies implemented within the size- and beta-based subsamples. The risk-adjusted returns are estimated as the intercepts from the following market model regression:

$$r_{pt} - r_{ft} = \alpha_p + \beta_p(r_{mt} - r_{ft}) + e_{it}, \tag{9}$$

where r_{pt} is the return on the portfolio p, r_{mt} is the return on the value-weighted index, and r_{ft} is the interest rate on 1-month Treasury Bill. Consistent with the negative betas of the zero-cost strategies, the abnormal returns of the relative strength strategies estimated from these regressions slightly exceed the raw returns given in Table III (Panel A). With the exception of the F-statistics becoming somewhat more significant, the findings in Table III (Panel B) are virtually the same as those reported in Table III (Panel A).

An additional implication of the results in Table III (Panel B) is that the abnormal performance of the zero-cost portfolio is due to the buy side of the transaction rather than the sell side. The portfolio of past winners achieves significant positive abnormal return when the value-weighted index is used as the benchmark, while the abnormal return of the portfolio of past losers is not statistically significant with this benchmark. However, in unreported regressions that used the equally weighted index as the benchmark, the positive and the negative abnormal returns of the winners and losers portfolios were both statistically significant. The magnitude and statistical significance of the abnormal returns of the zero-cost, winners minus losers, portfolio (0.0115 with a t-statistic of 3.84) was slightly higher when the equally weighted index was used in place of the value-weighted index as the benchmark.

From a practical investment perspective, it is important to assess whether the relative strength strategies will be profitable after accounting for transaction costs. On average, the relative strength trading rule results in a turnover of 84.8% semiannually.[12] The risk-adjusted return of the relative strength trading rule after considering a 0.5% one-way transaction cost[13] is 9.29% per year, which is reliably different from zero. The risk-adjusted returns after transaction costs are also significantly positive in each of the three size-based subsamples.

[12] The average turnovers for the buy and sell sides of the zero-cost portfolio are 86.6% and 83.1% respectively. These percentages are significantly less than the 90% turnover that would be expected if the transition probabilities are equal across the return decile portfolios.

[13] Berkowitz, Logue, and Noser (1988) estimate one way transaction costs of 23 basis points for institutional investors, suggesting that the assumed transaction cost of 0.5% per trade is conservative.

V. Subperiod Analysis

A. Seasonal Patterns in Relative Strength Portfolio Returns

This section tests for possible seasonal effects in the performance of the relative strength portfolios. Based on earlier papers, e.g., Roll (1983), we have reason to expect that the relative strength strategies will not be successful in the month of January. Table IV reports the average returns of the zero-cost portfolio in each calendar month and the results here support this conjecture.

Table III

Returns of Size-Based and Beta-Based Relative Strength Portfolios

The relative strength portfolios are formed based on 6-month lagged returns and held for 6 months. The stocks are ranked in ascending order on the basis of 6-month lagged returns and the equally weighted portfolio of stocks in the lowest past return decile is portfolio P1, the equally weighted portfolio of stocks in the next decile is portfolio P2, and so on. Average monthly returns and excess returns of these portfolios and the returns of the relative strength portfolios formed using size-based and beta-based subsamples of securities are reported here. The subsample S1 contains the smallest firms, S2 contains the medium-sized firms, and S3 contains the largest firms. The subsamples β_1, β_2, and β_3 contain the firms with the smallest, medium, and the largest Scholes-Williams betas estimated from the returns data in the calendar year prior to portfolio formation. The sample period is January 1965 to December 1989.

Panel A: Average Monthly Returns							
	All	S1	S2	S3	β_1	β_2	β_3
P1	0.0079	0.0083	0.0047	0.0082	0.0129	0.0097	0.0052
	(1.56)	(1.35)	(0.99)	(2.22)	(2.92)	(2.01)	(0.95)
P2	0.0112	0.0117	0.0102	0.0098	0.0140	0.0128	0.0086
	(2.78)	(2.29)	(2.54)	(3.08)	(4.38)	(3.37)	(1.83)
P3	0.0125	0.0152	0.0125	0.0105	0.0132	0.0133	0.0102
	(3.40)	(3.23)	(3.34)	(3.53)	(4.59)	(3.77)	(2.28)
P4	0.0124	0.0163	0.0130	0.0105	0.0134	0.0128	0.0110
	(3.59)	(3.59)	(3.58)	(3.66)	(5.02)	(3.82)	(2.50)
P5	0.0128	0.0164	0.0134	0.0109	0.0135	0.0135	0.0121
	(3.87)	(3.74)	(3.83)	(3.85)	(5.14)	(4.15)	(2.86)
P6	0.0134	0.0174	0.0146	0.0102	0.0135	0.0142	0.0122
	(4.14)	(4.08)	(4.22)	(3.66)	(5.23)	(4.38)	(2.92)
P7	0.0136	0.0175	0.0143	0.0109	0.0136	0.0142	0.0126
	(4.19)	(4.13)	(4.12)	(3.90)	(5.09)	(4.43)	(3.01)
P8	0.0143	0.0174	0.0148	0.0111	0.0143	0.0146	0.0132
	(4.30)	(4.11)	(4.16)	(3.86)	(5.12)	(4.44)	(3.15)
P9	0.0153	0.0183	0.0154	0.0126	0.0165	0.0156	0.0141
	(4.36)	(4.28)	(4.11)	(4.17)	(5.34)	(4.56)	(3.28)
P10	0.0174	0.0182	0.0173	0.0157	0.0191	0.0176	0.0160
	(4.33)	(3.99)	(4.11)	(4.41)	(5.17)	(4.53)	(3.50)
P10–P1	0.0095	0.0099	0.0126	0.0075	0.0062	0.0079	0.0108
	(3.07)	(2.77)	(4.57)	(3.03)	(2.05)	(2.64)	(3.35)
F-Statistics[a]	2.83	2.65	4.51	4.38	2.51	1.99	1.69
p-Value	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.04)	(0.09)

Table III—*Continued*

Panel B: Excess Returns Using the CRSP Value-Weighted Index as the Market Proxy							
	All	S1	S2	S3	β_1	β_2	β_3
P1	−0.0030	−0.0029	−0.0062	−0.0020	0.0031	−0.0009	−0.0062
	(−0.89)	(−0.60)	(−2.11)	(−1.17)	(0.94)	(−0.28)	(−1.71)
P2	0.0011	0.0012	−0.0001	0.0000	0.0051	0.0029	−0.0024
	(0.43)	(0.31)	(−0.03)	(0.03)	(2.36)	(1.26)	(−0.87)
P3	0.0026	0.0051	0.0024	0.0009	0.0045	0.0035	−0.0007
	(1.24)	(1.46)	(1.18)	(0.93)	(2.45)	(1.83)	(−0.29)
P4	0.0026	0.0062	0.0030	0.0011	0.0048	0.0031	0.0000
	(1.48)	(1.90)	(1.57)	(1.24)	(2.98)	(1.83)	(0.01)
P5	0.0031	0.0064	0.0036	0.0014	0.0049	0.0038	0.0012
	(1.96)	(2.06)	(1.98)	(1.84)	(3.21)	(2.55)	(0.58)
P6	0.0037	0.0075	0.0048	0.0008	0.0048	0.0045	0.0013
	(2.55)	(2.51)	(2.74)	(1.13)	(3.46)	(3.12)	(0.69)
P7	0.0039	0.0075	0.0044	0.0015	0.0049	0.0045	0.0017
	(2.70)	(2.57)	(2.61)	(2.15)	(3.29)	(3.25)	(0.90)
P8	0.0045	0.0074	0.0048	0.0016	0.0054	0.0049	0.0023
	(3.01)	(2.56)	(2.76)	(2.12)	(3.53)	(3.29)	(1.19)
P9	0.0053	0.0082	0.0052	0.0029	0.0074	0.0057	0.0031
	(3.20)	(2.89)	(2.76)	(3.23)	(4.10)	(3.60)	(1.54)
P10	0.0070	0.0077	0.0067	0.0056	0.0094	0.0074	0.0048
	(3.24)	(2.56)	(2.91)	(3.50)	(4.10)	(3.47)	(2.02)
P10−P1	0.0100	0.0106	0.0129	0.0076	0.0063	0.0083	0.0111
	(3.23)	(2.97)	(4.69)	(3.08)	(2.09)	(2.76)	(3.42)
F-Statistics[b]	5.2910	5.4401	8.3713	4.7386	3.6045	4.0171	2.5872

[a]The *F*-statistics are computed under the hypothesis that the returns on portfolios P1 through P10 are jointly equal.

[b]The *F*-statistics are computed under the hypothesis that the abnormal returns on portfolios P1 through P10 are jointly equal to zero. All *F*-statistics are significant at the 1 percent level.

The relative strength strategy loses about 7% on average in each January but achieves positive abnormal returns in each of the other months.[14] The relative strength strategy realizes positive returns in 67% of the months, and 71% of the months when January is excluded (see Table V). The average return in non-January months is 1.66% per month.[15] Consistent with earlier papers, we find the magnitude of the negative January performance of the relative strength strategy to be inversely related to firm size. The negative

[14]It is possible that at least part of the negative January returns of the relative strength strategy is due to a tendency of past winners to trade at the ask prices and past losers to sell at the bid prices at the close of the last trading day in the year. See Keim (1989) for a discussion of bid-ask spread biases and the January effect.

[15]If we were to use our priors about the performance of relative strength strategies in January and reverse the buy and sell portfolios in that calendar month (taking a long position in the past losers and a short position in the past winners in January only), then the abnormal returns would be even larger. Such a strategy generates close to 25% per year in abnormal returns, and loses money (about −0.7%) only in 1 year out of the 25 years in the sample period.

Table IV

Returns on Size-Based Relative Strength Portfolios (P10–P1) by Calendar Months

The relative strength portfolios are formed based on 6-month lagged returns and held for 6 months. The stocks are ranked in ascending order on the basis of 6-month lagged returns and the equally weighted portfolio of stocks in the lowest past return decile is the *sell* portfolio and the equally weighted portfolio of stocks in the highest past return decile is the *buy* portfolio. This table reports the average monthly returns of the zero-cost, buy minus sell, portfolio in each calendar month. The average returns of the zero-cost portfolios formed using size-based subsamples of securities are also reported. The subsample S1 contains the smallest firms, S2 contains the medium-sized firms, and S3 contains the largest firms. The sample period is January 1965 to December 1989.

	All	S1	S2	S3
Jan.	−0.0686	−0.0797	−0.0347	−0.0161
	(−3.52)	(−3.36)	(−2.14)	(−1.28)
Feb.	0.0063	0.0089	0.0149	0.0099
	(0.85)	(0.81)	(2.44)	(1.35)
Mar.	0.0105	0.0196	0.0103	0.0108
	(1.37)	(2.08)	(1.49)	(1.49)
Apr.	0.0333	0.0323	0.0368	0.0215
	(7.39)	(5.35)	(7.29)	(4.91)
May	0.0102	0.0046	0.0091	0.0079
	(1.32)	(0.56)	(1.18)	(1.19)
June	0.0238	0.0237	0.0231	0.0185
	(3.86)	(3.50)	(3.23)	(2.59)
July	0.0075	0.0112	0.0084	0.0035
	(0.96)	(1.44)	(0.96)	(0.41)
Aug.	0.0027	0.0079	−0.0011	−0.0058
	(0.35)	(0.97)	(−0.14)	(−0.71)
Sept.	0.0116	0.0126	0.0137	0.0053
	(1.10)	(1.20)	(1.27)	(0.60)
Oct.	0.0137	0.0160	0.0151	0.0025
	(1.30)	(1.40)	(1.44)	(0.22)
Nov.	0.0372	0.0352	0.0331	0.0248
	(5.31)	(5.01)	(4.12)	(2.78)
Dec.	0.0264	0.0265	0.0224	0.0070
	(2.61)	(2.13)	(2.86)	(0.99)
Feb.–Dec.	0.0166	0.0181	0.0169	0.0096
	(6.67)	(6.47)	(6.83)	(4.00)
F-Statistics[a]	7.90	7.14	4.11	1.81
p-Value	(0.00)	(0.00)	(0.00)	(0.51)
F-Statistics[b]	2.04	1.23	1.91	1.28
p-Value	(0.03)	(0.27)	(0.04)	(0.24)

[a]The F-statistics are computed under the hypothesis that the returns on the zero-cost portfolio are jointly equal in all calendar months.

[b]The F-statistics are computed under the hypothesis that the returns on the zero-cost portfolios are jointly equal in the calendar months February through December.

Table V

Proportion of Positive Returns of Relative Strength Portfolios by Calendar Months

The relative strength portfolios are formed based on 6-month lagged returns and held for 6 months. The stocks are ranked in ascending order on the basis of 6-month lagged returns and the equally weighted portfolio of stocks in the lowest past return decile is the *sell* portfolio and the equally weighted portfolio of stocks in the highest past return decile is the *buy* portfolio. This table reports the proportion of months when the average return of the zero-cost, buy minus sell, portfolio is positive. This proportion for the zero-cost portfolio formed within each size-based subsample of securities is also reported. The subsample S1 contains the smallest firms, S2 contains the medium-sized firms, and S3 contains the largest firms. The sample period is January 1965 to December 1989.

	All	S1	S2	S3
Jan.	0.24	0.16	0.20	0.44
Feb.	0.60	0.60	0.76	0.60
Mar.	0.80	0.76	0.72	0.72
Apr.	0.96	0.92	0.96	0.80
May	0.68	0.68	0.72	0.56
June	0.76	0.64	0.76	0.72
July	0.56	0.68	0.56	0.52
Aug.	0.52	0.60	0.48	0.48
Sept.	0.80	0.72	0.80	0.68
Oct.	0.64	0.60	0.64	0.56
Nov.	0.84	0.84	0.84	0.68
Dec.	0.68	0.76	0.68	0.44
Feb.–Dec.	0.71	0.71	0.72	0.61
All	0.67	0.66	0.68	0.60

average relative strength return in January is not statistically significant for the subsample of large firms.

The findings in Table IV suggest that there is also a seasonal pattern outside January. For example, the returns are fairly low in August and are particularly high in April, November, and December. The F-statistics reported in this table indicate that these monthly differences outside January are statistically significant for the whole sample as well as for the sample of medium-size firms.

One of the interesting findings documented in this table is that the relative strength strategy produces positive returns in 96% (24 out of 25) of the Aprils. The large (3.33%) and consistently positive April returns may be related to the fact that corporations must transfer money to their pension funds prior to April 15 if the funds are to qualify for a tax deduction in the previous year. If these pension fund assets are primarily invested by portfolio managers who follow relative strength rules, then the winners portfolio may benefit from additional price pressure in this month. Similarly, the larger than average returns in November and December may in part be due to price pressure arising from portfolio managers selling their losers in these months for tax or window dressing reasons.

Table VI

Returns of Size-Based Relative Strength Portfolios: Subperiod Analysis

The relative strength portfolios are formed based on 6-month lagged returns and held for 6 months. The stocks are ranked in ascending order on the basis of 6-month lagged returns and the equally weighted portfolio of stocks in the lowest past return decile is the *sell* portfolio and the equally weighted portfolio of stocks in the highest past return decile is the *buy* portfolio. This table reports the average monthly returns of the zero-cost, buy minus sell, portfolio within 5-year subperiods. The average returns of the zero-cost portfolios formed using size-based subsamples of securities within subperiods are also reported. The subsample S1 contains the smallest firms, S2 contains the medium-sized firms, and S3 contains the largest firms. The sample period is January 1965 to December 1989.

Sample	Months	65–69	70–74	75–79	80–84	85–89
	All	0.0123	0.0109	−0.0044	0.0127	0.0162
		(1.94)	(1.23)	(−0.51)	(2.67)	(3.42)
All	Jan.	−0.0524	−0.1070	−0.1017	−0.0253	−0.0569
		(−1.28)	(−2.54)	(−1.31)	(−1.38)	(−2.76)
	Feb.–Dec.	0.0182	0.0217	0.0044	0.0161	0.0229
		(3.36)	(2.88)	(0.78)	(3.44)	(6.09)
	All	0.0082	0.0128	−0.0064	0.0153	0.0197
		(1.14)	(1.63)	(−0.58)	(2.61)	(2.89)
S1	Jan.	−0.0838	−0.0853	−0.1107	−0.0124	−0.1064
		(−1.60)	(−2.29)	(−1.09)	(−0.62)	(−4.45)
	Feb.–Dec.	0.0165	0.0217	0.0031	0.0179	0.0311
		(3.19)	(3.18)	(0.41)	(2.94)	(6.59)
	All	0.0177	0.0115	0.0018	0.0172	0.0146
		(3.08)	(1.57)	(0.24)	(3.38)	(3.40)
S2	Jan.	−0.0264	−0.0465	−0.0795	−0.0100	−0.0112
		(−1.05)	(−1.81)	(−1.16)	(−0.46)	(−0.48)
	Feb.–Dec.	0.0217	0.0168	0.0092	0.0197	0.0170
		(3.86)	(2.29)	(1.87)	(3.83)	(4.08)
	All	0.0129	0.0115	0.0018	0.0076	0.0035
		(2.71)	(1.62)	(0.35)	(1.41)	(0.73)
S3	Jan.	−0.0073	−0.0154	−0.0335	−0.0094	−0.0147
		(−0.32)	(−0.48)	(−0.77)	(−0.33)	(−0.78)
	Feb.–Dec.	0.0148	0.0139	0.0050	0.0092	0.0052
		(3.08)	(1.95)	(1.21)	(1.70)	(1.04)

B. Portfolio Returns Over 5-Year Subperiods

This section documents the returns of the 6-month/6-month zero-cost strategy in each of the five 5-year subperiods in the 1965 to 1989 sample period. The evidence in Table VI indicates that the returns of the strategy, when implemented on the entire sample of stocks, produces average returns that are positive in all but one time period (1975 to 1979). An analysis of this strategy applied to size-based subsamples indicates that the negative returns in the 1975 to 1979 time period is due primarily to the January returns of the small firms. The strategy yields positive profits in each of the 5-year time periods when it is implemented on the subsamples of large- and medium-size firms. In addition, the returns are positive in each of the 5-year periods as well as in each size-based subsample when the month of January is excluded.

VI. Performance of Relative Strength Portfolios in Event Time

In this section we examine the returns of the relative strength portfolio in event time. We track the average portfolio returns in each of the 36 months following the portfolio formation date.

This event study analysis provides both additional insights about the riskiness of the strategy and about whether the profits are due to overreaction or underreaction. Significant positive returns in months beyond the holding period would indicate that the zero-cost portfolio systematically selects stocks that have higher than average unconditional returns either because of their risk or for other reasons such as differential tax exposures. Significant negative returns of the zero-cost portfolio in the months following the holding period would suggest that the price changes during the holding period are at least partially temporary.

Table VII presents the average monthly and cumulative returns of the zero-cost portfolio in event time in the 36 months after the formation date.[16] With the exception of month 1, the average return in each month is positive in the first year. The average return is negative in each month in year 2 as well as in the first half of year 3 and virtually zero thereafter. The cumulative returns reach a maximum of 9.5% at the end of 12 months but decline to about 4% by the end of month 36.

The negative returns beyond month 12 indicate that the relative strength strategy does not tend to pick stocks that have high unconditional expected returns. The observed pattern of initially positive and then negative returns of the zero-cost portfolio also suggests that the observed price changes in the first 12 months after the formation period may not be permanent. Unfortunately, estimates of expected returns over 2-year periods are not very precise. As a result, the negative returns for the zero-cost portfolio in years 2 and 3 are not statistically significant (t-statistic of -1.27). Similarly, since the abnormal return over the entire 36-month period is not statistically different from zero, we cannot rule out the possibility that the positive returns over the first 12 months is entirely temporary.[17]

[16] Since overlapping returns are used to calculate the cumulative returns in event time, the autocorrelation-consistent Newey-West standard errors are used to compute the t-statistics for the cumulative returns (see Newey and West (1987)).

[17] Another reason why we find this evidence hard to interpret is that the entire negative return over this holding period occurs in Januaries. The returns beyond the first year are close to zero in non-January months.

Table VII

Performance of Relative Strength Portfolios in Event Time

The relative strength portfolios are formed based on 6-month lagged returns. The stocks are ranked in ascending order on the basis of 6-month lagged returns. The equally weighted portfolio of stocks in the lowest past return decile is the *sell* portfolio and the equally weighted portfolio of stocks in the highest past return decile is the *buy* portfolio. This table reports the average returns of the zero-cost, buy minus sell, portfolio in each month following the formation period. t is the month after portfolio formation. The sample period is January 1965 to December 1989. Autocorrelation-consistent estimates of standard errors are used to compute the t-statistics for cumulative returns.

t	Monthly Return	Cumulative Return	t	Monthly Return	Cumulative Return	t	Monthly Return	Cumulative Return
1	−0.0025	−0.0025	13	−0.0036	0.0915	25	−0.0035	0.0521
	(−0.59)	(−0.59)		(−1.12)	(3.35)		(−1.36)	(1.41)
2	0.0124	0.0099	14	−0.0039	0.0876	26	−0.0030	0.0492
	(3.29)	(1.37)		(−1.34)	(3.07)		(−1.14)	(1.22)
3	0.0116	0.0216	15	−0.0034	0.0842	27	−0.0024	0.0467
	(3.18)	(2.20)		(−1.21)	(2.89)		(−0.98)	(1.10)
4	0.0110	0.0326	16	−0.0038	0.0804	28	−0.0032	0.0435
	(3.19)	(2.67)		(−1.41)	(2.76)		(−1.33)	(0.98)
5	0.0093	0.0419	17	−0.0047	0.0757	29	−0.0032	0.0403
	(2.82)	(2.79)		(−1.74)	(2.70)		(−1.38)	(0.87)
6	0.0091	0.0510	18	−0.0056	0.0701	30	−0.0030	0.0373
	(2.94)	(2.92)		(−2.19)	(2.68)		(−1.31)	(0.77)
7	0.0134	0.0644	19	−0.0026	0.0675	31	−0.0001	0.0372
	(4.98)	(3.32)		(−1.14)	(2.75)		(−0.06)	(0.74)
8	0.0115	0.0759	20	−0.0032	0.0642	32	0.0008	0.0380
	(4.16)	(3.60)		(−1.35)	(2.73)		(0.41)	(0.73)
9	0.0085	0.0844	21	−0.0032	0.0611	33	0.0013	0.0394
	(3.07)	(3.73)		(−1.32)	(2.55)		(0.62)	(0.73)
10	0.0048	0.0892	22	−0.0034	0.0577	34	0.0008	0.0402
	(1.69)	(3.74)		(−1.39)	(2.21)		(0.36)	(0.71)
11	0.0045	0.0938	23	−0.0011	0.0566	35	0.0010	0.0412
	(1.55)	(3.77)		(−0.45)	(1.93)		(0.45)	(0.71)
12	0.0013	0.0951	24	−0.0010	0.0556	36	−0.0005	0.0406
	(0.43)	(3.67)		(−0.40)	(1.69)		(−0.24)	(0.67)

One possible explanation of the inverted U shape in the cumulative returns is that the risk of the strategy changes over event time. Perhaps, the strategy picks stocks that are initially very risky and the risk then diminishes with time. To assess this possibility we estimate the betas in each event month with respect to the value-weighted index and the equally weighted index. The beta of the zero-cost portfolio with respect to the value-weighted (equally weighted) index is initially −0.20 (−0.41) and then it steadily increases to 0.02 (−0.08). Although these results indicate that the risk of the zero-cost portfolio does change over time, the direction of change in risk goes counter to what would be required to explain the change in average returns.

VII. Back-Testing the Strategy

This section examines the extent to which the relative strength profits reported in the previous sections existed prior to 1965. Specifically, we replicate the test in Table VII, which tracks the performance of the 6-month relative strength portfolio in event time for both the 1927 to 1940 time period and the 1941 to 1964 time period. As Fama and French (1988) and others have noted, the market was extremely volatile and experienced a significant degree of mean reversion in the 1927 to 1940 period. In contrast, the market's volatility in the 1941 to 1964 period was similar to the volatility in the 1965 to 1989 period and the market index did not exhibit mean reversion in the post-1940 period.

Table VIII (Panel A) reports the returns of the 6-month relative strength strategy in the 36 event months over the 1927 to 1940 time period. The returns in this time period are significantly lower than the returns in the 1965 to 1989 period, but the patterns of returns across event months is somewhat similar. The month 1 returns are strongly negative on average (about −5%). The returns in months 2 through 10 are statistically insignificant, but the returns in the later months are substantially lower. The cumulative excess return equals −40.81% in month 36.

These negative cumulative returns are likely to be due to two factors: First, because of the greater volatility in this period, many of the firms in the loser's decile were close to bankruptcy and thus had very high betas over the holding periods. The beta of the zero-cost 6-month/6-month strategy is about −0.5 in this period and it is substantially higher following periods of market declines. The second factor relates to the market's mean reversion in this time period. As the decomposition in Subsection III.*A* and the regression results in Subsection III.*B* indicate, negative serial correlation in the market and large market movements will reduce the profits from relative strength strategies. This is because the relative strength strategy tends to select high- (low-) beta stocks following a market increase (decrease) and hence tends to perform poorly during market reversals. For example, following a 40% decline in the equally weighted index over the previous 6 months, the index rebounded with a 43% increase in July 1932. In this month the 6-month/6-month relative strength portfolio experienced a negative 40% return. In the following month the equally weighted index increased an additional 66% and the 6-month/6-month strategy lost 68%. In the 1930s there were four other months in which the 6-month/6-month strategy lost over 40%. Each occurred when the market increased substantially.

Panel B of Table VIII reports the returns in the 36 event months for the 1941 to 1964 period. The relative strength strategy returns over this time period are very similar to the returns in the more recent time period reported earlier. As in the 1965 to 1989 time period, the average return is slightly negative in month 1, significantly positive in month 2 through month 8, and negative in month 12 and beyond. In contrast to the findings for the 1965 to

1989 period, the positive cumulative return over the first 12 months dissipates almost entirely by month 24.

VIII. Stock Returns Around Earnings Announcement Dates

This section examines the returns of past winners and losers around their quarterly earnings announcement dates. By analyzing stock returns within a short window around the dissemination of important firm-specific information we have a sharp test that directly assesses the potential biases in market expectations. Consider, for example, the possibility that stock prices system-

Table VIII

Back-Testing the Strategy: Performance of Relative Strength Portfolios Prior to 1965

The relative strength portfolios are formed based on 6-month lagged returns. The stocks are ranked in ascending order on the basis of 6-month lagged returns. The equally weighted portfolio of stocks in the lowest past return decile is the *sell* portfolio and the equally weighted portfolio of stocks in the highest past return decile is the *buy* portfolio. This table reports the average returns of the zero-cost, buy minus sell, portfolio in each month following the formation period. t is the month after portfolio formation. Autocorrelation consistent estimates of standard errors are used to compute the t-statistics for cumulative returns.

Panel A: 1927–1940

t	Monthly Return	Cumulative Return	t	Monthly Return	Cumulative Return	t	Monthly Return	Cumulative Return
1	−0.0495	−0.0495	13	−0.0245	−0.1257	25	−0.0118	−0.3359
	(−3.72)	(−3.72)		(−2.60)	(−1.50)		(−1.41)	(−2.48)
2	−0.0143	−0.0639	14	−0.0166	−0.1423	26	−0.0067	−0.3427
	(−1.32)	(−2.21)		(−2.08)	(−1.69)		(−1.01)	(−2.53)
3	−0.0088	−0.0726	15	−0.0164	−0.1587	27	−0.0135	−0.3562
	(−0.87)	(−1.78)		(−1.87)	(−1.83)		(−1.82)	(−2.52)
4	−0.0048	−0.0775	16	−0.0200	−0.1787	28	−0.0082	−0.3644
	(−0.45)	(−1.60)		(−2.20)	(−2.01)		(−1.06)	(−2.47)
5	0.0061	−0.0713	17	−0.0131	−0.1919	29	−0.0125	−0.3769
	(0.60)	(−1.40)		(−1.80)	(−2.12)		(−1.37)	(−2.39)
6	0.0057	−0.0656	18	−0.0166	−0.2085	30	−0.0107	−0.3876
	(0.55)	(−1.22)		(−2.11)	(−2.07)		(−1.20)	(−2.29)
7	0.0092	−0.0564	19	−0.0161	−0.2245	31	−0.0018	−0.3894
	(0.83)	(−1.05)		(−1.90)	(−2.01)		(−0.20)	(−2.18)
8	0.0054	−0.0511	20	−0.0224	−0.2469	32	−0.0022	−0.3916
	(0.52)	(−0.92)		(−2.28)	(−2.03)		(−0.26)	(−2.07)
9	−0.0029	−0.0539	21	−0.0178	−0.2647	33	0.0008	−0.3908
	(−0.34)	(−0.94)		(−1.92)	(−2.04)		(0.11)	(−1.99)
10	−0.0065	−0.0604	22	−0.0213	−0.2860	34	−0.0025	−0.3933
	(−0.68)	(−0.90)		(−2.08)	(−2.14)		(−0.41)	(−1.97)
11	−0.0183	−0.0787	23	−0.0183	−0.3043	35	−0.0050	−0.3983
	(−1.74)	(−1.04)		(−1.74)	(−2.23)		(−0.89)	(−1.97)
12	−0.0225	−0.1012	24	−0.0198	−0.3241	36	−0.0098	−0.4081
	(−2.35)	(−1.27)		(−1.94)	(−2.41)		(−1.47)	(−2.01)

Table VIII—*Continued*

Panel B: 1941–1964

t	Monthly Return	Cumulative Return	t	Monthly Return	Cumulative Return	t	Monthly Return	Cumulative Return
1	−0.0035	−0.0035	13	−0.0068	0.0515	25	−0.0035	0.0014
	(−1.04)	(−1.04)		(−2.14)	(2.57)		(−1.32)	(0.04)
2	0.0069	0.0034	14	−0.0085	0.0429	26	−0.0027	−0.0013
	(2.32)	(0.59)		(−3.07)	(1.90)		(−1.08)	(−0.03)
3	0.0109	0.0143	15	−0.0059	0.0371	27	−0.0015	−0.0028
	(4.15)	(2.20)		(−2.40)	(1.54)		(−0.69)	(−0.07)
4	0.0098	0.0241	16	−0.0063	0.0308	28	−0.0003	−0.0030
	(3.81)	(3.15)		(−2.80)	(1.21)		(−0.14)	(−0.08)
5	0.0075	0.0316	17	−0.0080	0.0228	29	−0.0009	−0.0039
	(3.09)	(3.40)		(−3.70)	(0.86)		(−0.51)	(−0.11)
6	0.0049	0.0365	18	−0.0074	0.0153	30	−0.0001	−0.0040
	(1.97)	(3.42)		(−3.63)	(0.56)		(−0.03)	(−0.12)
7	0.0079	0.0444	19	−0.0033	0.0120	31	0.0017	−0.0023
	(3.24)	(3.82)		(−1.61)	(0.43)		(0.98)	(−0.08)
8	0.0062	0.0507	20	−0.0012	0.0108	32	0.0011	−0.0012
	(2.52)	(4.00)		(−0.61)	(0.38)		(0.69)	(−0.05)
9	0.0039	0.0546	21	−0.0016	0.0092	33	−0.0005	−0.0017
	(1.63)	(3.91)		(−0.81)	(0.31)		(−0.32)	(−0.10)
10	0.0022	0.0568	22	−0.0021	0.0071	34	−0.0006	−0.0023
	(0.96)	(3.73)		(−1.04)	(0.22)		(−0.37)	(−0.17)
11	0.0024	0.0592	23	−0.0008	0.0063	35	−0.0004	−0.0027
	(1.00)	(3.70)		(−0.35)	(0.19)		(−0.24)	(−0.20)
12	−0.0009	0.0583	24	−0.0014	0.0050	36	−0.0004	−0.0030
	(−0.34)	(3.40)		(−0.60)	(0.14)		(−0.28)	(−0.20)

atically underreact to information about future earnings. In this case, the stock returns for past winners, which presumably had favorable information revealed in the past, should realize positive returns around the time when their earnings are actually announced. Similarly, past losers should realize negative returns around the time their earnings are announced.[18] The quarterly earnings announcement dates used in this analysis are obtained from the COMPUSTAT quarterly industrial database. The sample period for this part of the study is 1980 to 1989, the period covered by the 1990 COMPUSTAT quarterly file. On average, there are 429.2 available quarterly earnings announcements per month with matched stock return data.

Our tests again separate firms into deciles based on their prior 6-month returns. The 3-day returns (days −2 to 0) of the individual stocks in these groups are then calculated around each of their quarterly earnings announcements that occur within 36 months of the date at which the stocks are ranked according to their past returns. Table IX reports the differences between the

[18] Chopra, Lakonishok, and Ritter (1992) use a similar approach to evaluate the evidence of long horizon overreaction documented by De Bondt and Thaler (1985). See also Bernard and Thomas (1990).

Table IX

Quarterly Earnings Announcement Date Returns

The stocks are ranked in ascending order on the basis of 6-month lagged returns. The stocks in the lowest past return decile are called the *losers* group and the stocks in the highest past return decile is called the *winners* group. The differences between the 3-day returns (returns on days −2 to 0) around quarterly earnings announcements for stocks in the winners group and the losers group are reported here ($r_t^w - r_t^l$). t is the month after the ranking date. The sample period is January 1980 to December 1989.

t	$r_t^w - r_t^l$	t	$r_t^w - r_t^l$	t	$r_t^w - r_t^l$
1	0.0055	13	−0.0055	25	−0.0002
	(2.75)		(−2.56)		(−0.11)
2	0.0082	14	−0.0080	26	−0.0021
	(4.41)		(−3.89)		(−1.02)
3	0.0082	15	−0.0071	27	−0.0032
	(4.36)		(−4.04)		(−1.68)
4	0.0090	16	−0.0097	28	−0.0028
	(4.88)		(−5.75)		(−1.31)
5	0.0059	17	−0.0062	29	−0.0015
	(3.16)		(−2.90)		(−0.62)
6	0.0058	18	−0.0060	30	−0.0021
	(3.14)		(−2.96)		(−1.10)
7	0.0013	19	−0.0031	31	−0.0027
	(0.62)		(−1.63)		(−1.52)
8	0.0000	20	−0.0017	32	−0.0021
	(−0.02)		(−0.82)		(−1.13)
9	−0.0020	21	0.0006	33	−0.0020
	(−1.07)		(0.27)		(−1.05)
10	−0.0031	22	−0.0005	34	−0.0017
	(−1.60)		(−0.29)		(−0.91)
11	−0.0039	23	−0.0001	35	−0.0022
	(−2.23)		(−0.05)		(−1.29)
12	−0.0053	24	0.0012	36	−0.0059
	(−2.75)		(0.63)		(−2.91)

average announcement period returns for the winners and losers deciles in each of the 36 months following the ranking date. The pattern of announcement date returns presented in this table is consistent with the pattern of the zero-cost portfolio returns reported in Table VII. For the first 6 months the announcement date returns of the past winners exceed the announcement date returns of the past losers by over 0.7% on average, and is statistically significant in each of these 6 months. Since there are on average 2 quarterly earnings announcements per firm within a 6-month period, the returns around the earnings announcements represents about 25% of the zero-cost portfolio returns over this holding period.

The negative announcement period returns in later months are consistent with the negative relative strength portfolio returns beyond month 12 documented earlier (see Table VII). From months 8 through 20 the differences in announcement date returns are negative and are generally statistically significant. The announcement period returns are especially significant in months 11 through 18 where they average about -0.7%. In the later months the differences between the announcement period returns of the winners and losers are generally negative but are close to zero.

The predictability of stock returns around quarterly earnings announcements documented in Table IX is similar to the recent findings of Bernard and Thomas (1990). Bernard and Thomas find that average returns around quarterly earnings announcement dates are significantly positive following a favorable earnings surprise in the previous quarter. This is consistent with the positive announcement returns we see in the first 7 months in Table IX. Bernard and Thomas also find that the average return around earnings announcement dates is significantly negative 4 quarters after a positive earnings surprise. The significant negative returns around earnings announcement dates in months 11 through 18 are consistent with this finding.

IX. Conclusions

Trading strategies that buy past winners and sell past losers realize significant abnormal returns over the 1965 to 1989 period. For example, the strategy we examine in most detail, which selects stocks based on their past 6-month returns and holds them for 6 months, realizes a compounded excess return of 12.01% per year on average. Additional evidence indicates that the profitability of the relative strength strategies are not due to their systematic risk. The results of our tests also indicate that the relative strength profits cannot be attributed to lead-lag effects that result from delayed stock price reactions to common factors. The evidence is, however, consistent with delayed price reactions to firm-specific information.

The returns of the zero-cost winners minus losers portfolio were examined in each of the 36 months following the portfolio formation date. With the exception of the first month, this portfolio realizes positive returns in each of the 12 months after the formation date. However, the longer-term performances of these past winners and losers reveal that half of their excess returns in the year following the portfolio formation date dissipate within the following 2 years.

The returns of the stocks in the winners and losers portfolios around their earnings announcements in the 36 months following the formation period were also examined and a similar pattern was found. Specifically, stocks in the winners portfolio realize significantly higher returns than the stocks in the losers portfolio around the quarterly earnings announcements that are made in the first few months following the formation date. However, the announcement date returns in the 8 to 20 months following the formation date are significantly higher for the stocks in the losers portfolio than for the stocks in the winners portfolio.

The evidence of initial positive and later negative relative strength returns suggests that common interpretations of return reversals as evidence of overreaction and return persistence (i.e., past winners achieving positive returns in the future) as evidence of underreaction are probably overly simplistic. A more sophisticated model of investor behavior is needed to explain the observed pattern of returns. One interpretation of our results is that transactions by investors who buy past winners and sell past losers move prices away from their long-run values temporarily and thereby cause prices to overreact. This interpretation is consistent with the analysis of DeLong, Shleifer, Summers, and Waldman (1990) who explore the implications of what they call "positive feedback traders" on market price. Alternatively, it is possible that the market underreacts to information about the short-term prospects of firms but overreacts to information about their long-term prospects. This is plausible given that the nature of the information available about a firm's short-term prospects, such as earnings forecasts, is different from the nature of the more ambiguous information that is used by investors to assess a firm's longer-term prospects.

The evidence in this paper does not allow us to distinguish between these two hypotheses about investor behavior. In addition, there are probably other explanations for these results. Given that our results suggest that investor expectations are systematically biased, further research that attempts to identify explanations for these empirical regularities would be of interest.

REFERENCES

Ball, Ray and S. P. Kothari, 1989, Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns, *Journal of Financial Economics* 25, 51–74.

Banz, Rolf, 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Berkowitz, Stephen A., Dennis E. Logue, and Eugene A. Noser, 1988, The total costs of transactions on the NYSE, *Journal of Finance* 43, 97–112.

Bernard, Arnold, 1984, *How to Use the Value Line Investment Survey: A Subscriber's Guide* (Value Line, New York).

Bernard, Victor and Jacob Thomas, 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics* 13, 305–340.

Chan, K. C., 1988, On the contrarian investment strategy, *Journal of Business* 61, 147–163.

Chopra, Navin, Josef Lakonishok, and Jay Ritter, 1992, Measuring abnormal returns: Do stocks overreact? *Journal of Financial Economics* 31, 235–268.

Copeland, Thomas and David Mayers, 1982, The Value Line enigma (1965–1978): A case study of performance evaluation issues, *Journal of Financial Economics* 10, 289–321.

De Bondt, Werner F. M. and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793–805.

———, 1987, Further evidence of investor overreaction and stock market seasonality, *Journal of Finance* 42, 557–581.

De Long, Bradford J., Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldman, 1990, Positive Feedback Investment Strategies and Destabilizing Rational Speculation, *Journal of Finance* 45, 379–395.

Fama, Eugene and Kenneth French, 1988, Permanent and temporary components of stock prices, *Journal of Political Economy* 96, 246–273.

———, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene, and J. D. MacBeth, 1973, Risk return and equilibrium: Empirical test, *Journal of Political Economy* 81, 607–636.

Grinblatt, Mark and Sheridan Titman, 1989, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of Business* 62, 394–415.

———, 1991, Performance measurement without benchmarks: An examination of mutual fund returns, Working paper, University of California at Los Angeles.

Jegadeesh, Narasimhan, 1987, Predictable behavior of security returns and tests of asset pricing models, Ph.D. dissertation, Columbia University.

———, 1990, Evidence of predictable behavior of security returns, *Journal of Finance* 45, 881–898.

———, 1992, Does market risk really explain the size effect?, *Journal of Financial and Quantitative Analysis* 10, 337–351.

——— and Sheridan Titman, 1991, Short horizon return reversals and the bid-ask spread, Working paper, University of California at Los Angeles.

Jensen, Michael and George Bennington, 1970, Random walks and technical theories: Some additional evidence, *Journal of Finance* 25, 469–482.

Kahneman, D. and A. Tversky, 1982, Intuitive prediction: Biases and corrective procedures, in D. Kahneman, P. Slovic, and A. Tversky, eds.: *Judgement Under Uncertainty: Heuristics and Biases* (Cambridge University Press, London).

Keim, Donald, 1989, Trading patterns, bid-ask spreads, and estimated security returns: The case of common stocks at calendar turning points, *Journal of Financial Economics* 25, 75–98.

Lehmann, Bruce, 1990, Fads, martingales and market efficiency, *Quarterly Journal of Economics* 105, 1–28.

Levy, Robert, 1967, Relative strength as a criterion for investment selection, *Journal of Finance* 22, 595–610.

Lo, Andrew and Craig MacKinlay, 1990, When are contrarian profits due to stock market overreaction?, *Review of Financial Studies* 3, 175–205.

Newey, W. K. and K. D. West, 1987, A Simple positive definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–705.

Roll, Richard, 1983, Vas ist Das? *Journal of Portfolio Management*, 18–28.

Scholes, Myron and Joseph Williams, 1977, Estimating betas from nonsynchronous data, *Journal of Financial Economics* 5, 309–327.

Shiller, Robert J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71, 421–436.

Stickel, Scott, 1985, The effect of Value Line Investment Survey rank changes on common stock prices, *Journal of Financial Economics* 14, 121–144.

Zarowin, Paul, 1990, Size, seasonality, and stock market overreaction, *Journal of Financial and Quantitative Analysis* 25, 113–125.

The Journal of Financial Research • Vol. XIX, No. 1 • Pages 59–74 • Spring 1996

THE COSTS OF RAISING CAPITAL

Inmoo Lee, Scott Lochhead, Jay Ritter
University of Illinois at Urbana-Champaign

Quanshui Zhao
City University of Hong Kong

Abstract

We report the average costs of raising external debt and equity capital for U.S. corporations from 1990 to 1994. For initial public offerings (IPOs) of equity, the direct costs average 11.0 percent of the proceeds. For seasoned equity offerings (SEOs), the direct costs average 7.1 percent. For convertible bonds, the direct costs average 3.8 percent. For straight debt issues, the direct costs average 2.2 percent, although they are strongly related to the credit rating of the issue. All classes of securities exhibit economies of scale, although they are less pronounced for straight debt issues. IPOs also incur a substantial indirect cost due to short-run underpricing. Most large equity offers include an international tranche, although debt issues do not.

I. Introduction

In this article we present the average costs of raising external capital for U.S. corporations from 1990 to 1994. Specifically, we report the average spreads on public equity offerings and debt offerings, along with the other direct costs of raising capital, as a percentage of the proceeds. We find substantial economies of scale for initial public offerings (IPOs) of equity and seasoned equity offerings (SEOs). We also find substantial economies of scale for both straight bond offerings and convertible bond offerings. Spreads on bond offerings are highly sensitive to the credit rating of the offering. This article is descriptive in nature; no theories are tested. Its purpose is to provide benchmark numbers for use by issuers of securities. We do not address why firms issue the securities they do. This much broader corporate finance question would have to address taxes, corporate control, debt capacity, long-run performance patterns, investment-financing interactions, etc.

We would like to thank Charles Calomiris and Tim Loughran for useful comments on an earlier draft.

II. Data and Terminology

Securities Data Company's (SDC) New Issues database is the primary source of information. After downloading SDC's data, we identified outliers and checked suspicious numbers in other publicly available sources. The New Issues database includes publicly placed firm commitment offerings only. In all of our tables, we exclude ADRs and unit offerings.[1] We restrict our sample to securities offered by domestic operating companies, and so exclude closed-end fund and real estate investment trust (REIT) offerings. We also exclude rights offerings and shelf registrations.[2]

We use security offerings from January 1990 to December 1994, a five-year period of relatively low inflation. Consequently, we do not make any inflation adjustments; all proceeds are the nominal proceeds. Proceeds reflect the gross proceeds raised in the U.S. and do not include money raised from the exercise of overallotment options or an international tranche, if any. In the case of equity offerings, the proceeds include the amount raised from both primary and secondary components. Primary shares are those being sold by the company, thereby increasing the number of shares outstanding. Secondary shares are those being sold by existing shareholders (managers, venture capitalists, etc.), which neither increase the number of shares outstanding nor provide capital for the company. Many IPOs include both primary and secondary components, with the fraction that is primary generally higher for younger companies. A few IPOs, sometimes involving spin-offs from parent companies, are pure secondaries. All of our SEOs involve primary shares; we exclude "registered secondaries," in which the entire issue is composed of shares being sold by existing shareholders, from our SEO sample.

For our sample of bond offerings, we exclude issues with a maturity date of one year or less. Our sample includes both zero-coupon, original-issue discount bonds, and coupon bonds. We include serial, floating-rate, and reset bonds, as

---

[1]ADRs are American Depository Receipts (also called American Depository Shares) that are traded in the United States for foreign issuers. Unit offerings are bundles of securities (frequently, a share plus a warrant to buy a share at some exercise price), commonly issued in small IPOs by young, speculative companies taken public by less-prestigious investment bankers.

[2]Rights offerings give existing shareholders the right to buy the securities offered. While they are common in many countries, rights offerings have been rare in the United States during the last twenty years. See Smith (1977), Hansen and Pinkerton (1982), and Hansen (1988) for a discussion of rights offerings. Shelf registrations are offerings whereby a company meeting certain qualifications is permitted to issue securities without issuing a prospectus (taking the securities "off the shelf" and selling them). In our sample period, shelf equity offerings are practically nonexistent, although there are many bond offerings (typically smaller issues) using shelf registrations that we exclude.

well as traditional coupon bonds.[3] We exclude mortgage-backed bonds. For zero-coupon and original-issue discount bonds that are sold for less than their par value, our percentage spreads and costs are based upon the offer price, and not the face value. Our convertible bond sample includes only issues that are convertible into shares of the issuing company. Exchangeable bonds, where the bond is convertible into shares of a different company, are not in our sample. None of our convertible bonds has a maturity date of less than five years.

We refer to new equity issues by publicly traded companies as seasoned equity offerings, reserving the use of "secondary" to identify the source of shares. Among practitioners, the term "secondary offering" is frequently used to refer to an SEO. Seasoning refers to whether the security being offered is already publicly traded; IPOs are unseasoned new issues. For that matter, the term "new issues" is sometimes used to refer to any security offering, and sometimes used to refer to equity IPOs alone. Although a new bond issue is an unseasoned new issue, and therefore a debt initial public offering, we use the term IPO to refer to unseasoned equity offerings exclusively.

Gross spreads are the commissions paid to investment bankers when securities are issued. Since buyers do not pay commissions on new security issues, these spreads implicitly reflect both the buyer and seller commissions. Other direct costs include the legal, auditing, and printing costs associated with putting together a prospectus.

III. Evidence

Average Spreads and Total Direct Costs

In Table 1 we report the average investment banker commissions (gross spreads) and other direct expenses for four classes of securities: IPOs, SEOs, convertible bonds, and straight bonds. In addition to reporting the average direct costs for each class, we also classify issues by proceeds categories. By going across a row, a reader can see how the expenses vary by security type, holding proceeds constant. By going down a column, a reader can see the magnitude of the economies of scale for a given type of security. Also reported is the number of observations in each category.

In Table 1 the median IPO is \$24.4 million, the median SEO is \$33.8 million, the median convertible bond is \$75 million, and the median straight

---

[3]Serial bonds have the individual bonds maturing on different dates, with the coupons varying depending upon the maturity date. Reset and floating-rate bonds have the interest rate changing periodically, with the new interest rate determined either by an auction (reset) or a formula (floaters).

TABLE 1. Direct Costs as a Percentage of Gross Proceeds for Equity (IPOs and SEOs) and Straight and Convertible Bonds Offered by Domestic Operating Companies, 1990–94.

	Equity								Bonds							
	IPOs				SEOs				Convertible Bonds				Straight Bonds			
Proceeds[a] ($ millions)	N[b]	GS[c]	E[d]	TDC[e]	N	GS	E	TDC	N	GS	E	TDC	N	GS	E	TDC
2–9.99	337	9.05	7.91	16.96	167	7.72	5.56	13.28	4	6.07	2.68	8.75	32	2.07	2.32	4.39
10–19.99	389	7.24	4.39	11.63	310	6.23	2.49	8.72	14	5.48	3.18	8.66	78	1.36	1.40	2.76
20–39.99	533	7.01	2.69	9.70	425	5.60	1.33	6.93	18	4.16	1.95	6.11	89	1.54	0.88	2.42
40–59.99	215	6.96	1.76	8.72	261	5.05	0.82	5.87	28	3.26	1.04	4.30	90	0.72	0.60	1.32
60–79.99	79	6.74	1.46	8.20	143	4.57	0.61	5.18	47	2.64	0.59	3.23	92	1.76	0.58	2.34
80–99.99	51	6.47	1.44	7.91	71	4.25	0.48	4.73	13	2.43	0.61	3.04	112	1.55	0.61	2.16
100–199.99	106	6.03	1.03	7.06	152	3.85	0.37	4.22	57	2.34	0.42	2.76	409	1.77	0.54	2.31
200–499.99	47	5.67	0.86	6.53	55	3.26	0.21	3.47	27	1.99	0.19	2.18	170	1.79	0.40	2.19
500–up	10	5.21	0.51	5.72	9	3.03	0.12	3.15	3	2.00	0.09	2.09	20	1.39	0.25	1.64
Total	1767	7.31	3.69	11.00	1593	5.44	1.67	7.11	211	2.92	0.87	3.79	1092	1.62	0.62	2.24

Notes: Closed-end funds (SIC 6726), REITs (SIC 6798), ADRs, and unit offerings are excluded from the sample. Rights offerings for SEOs are also excluded. Bond offerings do not include securities backed by mortgages and issues by Federal agencies (SIC 6011, 6019, 6111, and 999B). Only firm commitment offerings and nonshelf-registered offerings are included. Standard Industrial Classification (SIC) codes are from Securities Data Co. (SDC).

[a]Total proceeds raised in the United States, excluding proceeds from the exercise of overallotment options (SDC variable: PROCDS).
[b]Number of issues.
[c]Gross spreads as a percentage of total proceeds (including management fee, underwriting fee, and selling concession) (SDC variable: GPCTP).
[d]Other direct expenses as a percentage of total proceeds (including registration fee and printing, legal, and auditing costs) (SDC variables: EXPTH/(PROCDS)*10).
[e]Total direct costs as a percentage of total proceeds (total direct costs are the sum of gross spreads and other direct expenses).

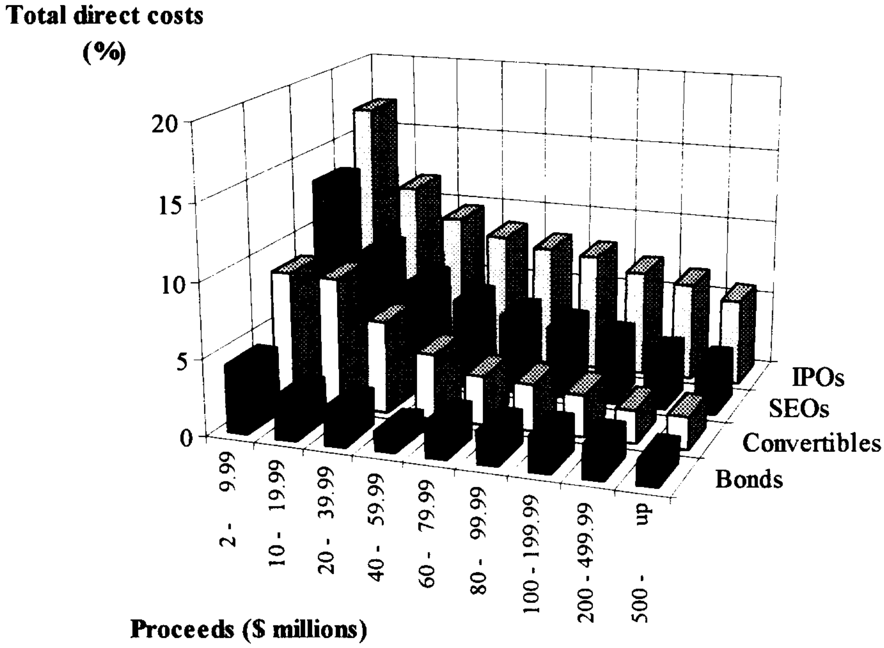


Figure I. Total Direct Costs as a Percentage of Gross Proceeds. The total direct costs for initial public offerings (IPOs), seasoned equity offerings (SEOs), convertible bonds, and straight bonds are composed of underwriter spreads and other direct expenses. Closed-end funds (SIC 6726), REITs (SIC 6798), ADRs, and unit offerings are excluded. Rights offerings for SEOs are also excluded. Bond offerings do not include securities backed by mortgages and issues by federal agencies (SIC 6011, 6019, 6111, and 999B). Only firm commitment offerings and nonshelf-registered offerings are included. The numbers plotted are reported in Table 1 for issues from 1990 to 1994.

bond is $100 million. For both IPOs and SEOs, substantial economies of scale exist in both the gross spreads and the other expenses.

For SEOs, the lack of any diseconomies, even for offerings over $500 million, is inconsistent with the findings of Hansen and Torregrosa (1992), who report diseconomies of scale for offers over $100 million. Hansen and Torregrosa use a sample of SEOs from 1978–86, in contrast to our 1990–94 sample period. Our conjecture is that while diseconomies of scale may have existed for very large issues before the mid 1980s, a structural change has probably occurred since then, possibly because of the market's greater experience with absorbing large numbers of big offerings. While they are not in our sample, the large number of multibillion dollar privatizations that have occurred around the world in the last decade have made megaofferings routine events.

In all of our tables, we report the averages based upon the number of observations for which we have data. For the gross spreads, SDC reports numbers for our entire sample. For the other direct expenses, however, many observations are missing. Consequently, the averages for the expenses are based upon a

TABLE 2. Direct Costs of Raising Capital, 1990–94: Utility versus Nonutility Companies.

	Equity						Bonds					
	IPOs			SEOs			Convertible			Straight		
Proceeds[a] ($ millions)	N[b]	GS[c]	TDC[d]	N	GS	TDC	N	GS	TDC	N	GS	TDC
Panel A. Nonutility Offerings Only												
2–9.99	332	9.04	16.97	154	7.91	13.76	4	6.07	8.75	29	2.07	4.53
10–19.99	388	7.24	11.64	278	6.42	9.01	12	5.54	8.65	47	1.70	3.28
20–39.99	528	7.01	9.70	399	5.70	7.07	16	4.20	6.23	63	1.59	2.52
40–59.99	214	6.96	8.71	240	5.17	6.02	28	3.26	4.30	76	0.73	1.37
60–79.99	78	6.74	8.21	131	4.68	5.31	47	2.64	3.23	84	1.84	2.44
80–99.99	47	6.46	7.88	60	4.35	4.84	12	2.54	3.19	104	1.61	2.25
100–199.99	101	6.01	7.01	137	3.97	4.36	55	2.34	2.77	381	1.83	2.38
200–499.99	44	5.65	6.49	50	3.27	3.48	26	1.97	2.16	154	1.87	2.27
500–up	10	5.21	5.72	8	3.12	3.25	3	2.00	2.09	19	1.28	1.53
Total	1742	7.31	11.01	1457	5.57	7.32	203	2.90	3.75	957	1.70	2.34
Panel B. Utility Offerings Only												
2–9.99	5	9.40	16.54	13	5.41	7.68	0	—	—	3	2.00	3.28
10–19.99	1	7.00	8.77	32	4.59	6.21	2	5.13	8.72	31	0.86	1.35
20–39.99	5	7.00	9.86	26	4.17	4.96	2	3.88	5.18	26	1.40	2.06
40–59.99	1	6.98	11.55	21	3.69	4.12	0	—	—	14	0.63	1.10
60–79.99	1	6.50	7.55	12	3.39	3.72	0	—	—	8	0.87	1.13
80–99.99	4	6.57	8.24	11	3.68	4.11	1	1.13	1.34	8	0.71	0.98
100–199.99	5	6.45	7.96	15	2.83	2.98	2	2.50	2.74	28	1.06	1.42
200–499.99	3	5.88	7.00	5	3.19	3.48	1	2.50	2.65	16	1.00	1.40
500–up	0	—	—	1	2.25	2.31	0	—	—	1	3.50	na[e]
Total	25	7.15	10.14	136	4.01	4.92	8	3.33	4.66	135	1.04	1.47

Notes: Closed-end funds (SIC 6726), REITs (SIC 6798), ADRs, and unit offerings are excluded from the sample. Rights offerings for SEOs are also excluded. Bond offerings do not include securities backed by mortgages and issues by Federal agencies (SIC 6011, 6019, 6111, and 999B). Only firm commitment offerings and nonshelf-registered offerings are included. Standard Industrial Classification (SIC) codes are from Securities Data Co. (SDC).

[a]Total proceeds raised in the United States, excluding proceeds from the exercise of overallotment options (SDC variable: PROCDS).
[b]Number of issues.
[c]Gross spreads as a percentage of total proceeds (including management fee, underwriting fee, and selling concession) (SDC variable: GPCTP).
[d]Other direct expenses as a percentage of total proceeds (including registration fee and printing, legal, and auditing costs) (SDC variables: EXPTH/(PROCDS)*10).
[e]Not available because of missing data on other direct expenses.

110138-OPC-POD-60-179

more limited number of observations.[4] For computing the average total direct costs in Table 1 (and other tables), we add the average gross spread and the average other expenses. In Figure I we show the average total direct costs for the four classes of securities, categorized by their gross proceeds.

The Appendix table reports the interquartile ranges for both the gross spreads and the total direct costs. (We report the interquartile range of the offerings for which we have complete data.) The largest variability of spreads occurs for bonds. As we document below, this can largely be explained based on differences in the credit quality of the issues.

Utility versus Nonutility Offerings

In Table 2 we report the direct costs of raising capital after categorizing offerings into utility and nonutility offerings. During the early 1990s, utilities were relatively minor issuers, representing roughly 10 percent of SEOs and straight bond offerings, and less than 5 percent of IPOs and convertibles. Spreads and direct costs are lower for utilities than for nonutilities. This pattern, previously documented by Bhagat and Frost (1986), may be partly due to the use of competitive bidding, rather than negotiated deals, for choosing an investment banker. Alternatively, it may be partly due to the relative noncomplexity of typical utility offerings.

Debt Offerings and Credit Quality

In Table 3 we report the costs of raising debt capital after categorizing issues by whether they are investment grade or noninvestment grade.[5] Following industry practice, we classify offerings as investment grade issues if they have a Standard & Poor's credit rating of BBB- or higher.[6]

Inspection of Table 3 discloses that for both convertibles and straight bonds, spreads are lower for investment-grade issues. For straight bonds, this difference is especially pronounced. Note that for issues raising less than $60

---

[4]If the offerings with missing expense information have systematically higher or lower expenses than those for which SDC reports information, our procedure would result in biased estimates of average expenses. To check this, for a sample of bond offerings in 1994 that are missing expense information, we used the Securities and Exchange Commission's Edgar electronic database (http://www.sec.gov/cgi-bin/srch-edgar) to find the expense information. The expenses for these issues are representative of those for which SDC reports information, suggesting our numbers do not have important biases.

[5]Following the practice of SDC, we report as separate offerings two bond issues by the same company on the same day if they have different maturity dates, provided they are not explicitly serial bonds. For example, on September 22, 1994, Southern Pacific Transport issued two bonds, one with proceeds of $8.1 million with a coupon rate of 7.61 percent, and the other with proceeds of $8.8 million and a coupon rate of 7.77 percent. We treat these as two distinct offerings.

[6]The highest credit rating is AAA, followed by AA, A, BBB, BB, B, C, and D, in order of their perceived default probabilities. These ratings are further partitioned by pluses and minuses.

TABLE 3. Average Gross Spreads and Total Direct Costs for Domestic Debt Issues, 1990–94.

	Convertible Bonds						Straight Bonds					
	Investment Grade[a]			Noninvestment Grade[b]			Investment Grade			Noninvestment Grade		
Proceeds[c] ($ millions)	N[d]	GS[e]	TDC[f]	N	GS	TDC	N	GS	TDC	N	GS	TDC
2–9.99	0	—	—	0	—	—	14	0.58	2.19	0	—	—
10–19.99	0	—	—	1	4.00	5.67	56	0.50	1.19	2	5.13	7.41
20–39.99	1	1.75	2.75	9	3.29	4.92	64	0.86	1.48	9	3.11	4.42
40–59.99	3	1.92	2.43	19	3.37	4.58	78	0.47	0.94	9	2.48	3.35
60–79.99	4	1.31	1.76	41	2.76	3.37	49	0.61	0.98	43	3.07	3.84
80–99.99	2	1.07	1.34	10	2.83	3.48	65	0.66	0.94	47	2.78	3.75
100–199.99	20	2.03	2.33	37	2.51	3.00	181	0.57	0.81	222	2.75	3.44
200–499.99	17	1.71	1.87	10	2.46	2.70	60	0.50	0.93	105	2.56	2.96
500–up	3	2.00	2.09	0	—	—	11	0.39	0.57	9	2.60	2.90
Total	50	1.81	2.09	127	2.81	3.53	578	0.58	0.94	446	2.75	3.42

Notes: Closed-end funds (SIC 6726), REITs (SIC 6798), ADRs, and unit offerings are excluded from the sample. Bond offerings do not include securities backed by mortgages and issues by Federal agencies (SIC 6011, 6019, 6111, and 999B). Only nonshelf-registered offerings are included. Standard Industrial Classification (SIC) codes are from Securities Data Co. (SDC).

[a]Firms with a BBB- or higher Standard & Poor's credit rating.
[b]Firms with a BB+ or lower Standard & Poor's credit rating.
[c]Total proceeds raised in the United States, excluding proceeds from the exercise of overallotment options (SDC variable: PROCDS).
[d]Number of issues.
[e]Gross spreads as a percentage of total proceeds (including management fee, underwriting fee, and selling concession) (SDC variable: GPCTP).
[f]Other direct expenses as a percentage of total proceeds (including registration fee and printing, legal, and auditing costs) (SDC variables: EXPTH/(PROCDS)*10).

million, very few noninvestment-grade issues exist. This reflects that smaller issues with lower credit quality are commonly placed privately, and thus do not appear in our sample.

This correlation of credit quality and issue size also explains why in Tables 1 and 2 straight bond issues do not appear to display large economies of scale: as the issue size increases, the credit quality of public issuers decreases, masking some of the economies of scale. Still, in Table 3, where we hold credit quality constant, the economies of scale for debt issues are more modest than those for equity issues in Tables 1 and 2. The correlation between issue size and credit quality also explains why the average spread is so low for bonds with \$40-\$59.9 million in proceeds. The average spread of only seventy-two basis points in Table 1 reflects that for this issue size, economies of scale are largely realized, while, at the same time, very few noninvestment-grade issuers exist. For smaller offerings, the lack of economies of scale keeps the average spread high. For larger offerings, the high proportion of noninvestment-grade issues pushes

TABLE 4. Direct and Indirect Costs, in Percent, of Equity IPOs, 1990–94.

Proceeds[a] ($ millions)	Gross Spreads[b]	Other Expenses[c]	Total Direct Costs[d]	Average Initial Return[e]	Average Direct and Indirect Costs[f]
2–9.99	9.05	7.91	16.96	16.36	25.16
10–19.99	7.24	4.39	11.63	9.65	18.15
20–39.99	7.01	2.69	9.70	12.48	18.18
40–59.99	6.96	1.76	8.72	13.65	17.95
60–79.99	6.74	1.46	8.20	11.31	16.35
80–99.99	6.47	1.44	7.91	8.91	14.14
100–199.99	6.03	1.03	7.06	7.16	12.78
200–499.99	5.67	0.86	6.53	5.70	11.10
500–up	5.21	0.51	5.72	7.53	10.36
Total	7.31	3.69	11.00	12.05	18.69

Notes: There are 1,767 domestic operating company IPOs in the sample. The first four columns express costs as a percentage of the offer price, and the last column expresses costs as a percentage of the market price.

[a]Total proceeds raised in the United States, excluding proceeds from the exercise of overallotment options (SDC variable: PROCDS).
[b]Gross spreads as a percentage of total proceeds (including management fee, underwriting fee, and selling concession) (SDC variable: GPCTP).
[c]Other direct expenses as a percentage of total proceeds (including registration fee and printing, legal, and auditing costs) (SDC variables: EXPTH/(PROCDS)*10).
[d]Total direct costs as a percentage of total proceeds (the average total direct costs are the sum of average gross spreads and average other direct expenses).
[e]Initial return = 100*{[closing price one day after the offering date (SDC variable: PR1DAY)/offering price (SDC variable: P)] − 1}. If PR1DAY is missing, PR2DAY is used.
[f]Total direct and indirect costs = $(d + e)/(1 + e/100)$, computed for each issue individually (excluding firms with other expenses or initial returns missing), and then averaged, where d is the percentage of total direct costs, and e is the percentage initial return.

the average spread up. In other words, the average spread of only seventy-two basis points for this category is not a typographical error.

Although not reported in any table, the average maturity of bond offerings is about ten years for all of the proceeds categories and investment grades.

Initial Public Offerings

In Table 4 we report not only the direct costs for IPOs, but also the indirect costs of short-run underpricing.[7] Inspection of the table reveals that, consistent with previous findings, IPOs are underpriced on average. With average direct costs of 11.0 percent and average initial returns of 12.0 percent, a typical

[7]We compute the average initial return only for those offerings for which SDC reports the market price at the end of the first day of trading or, if this is missing, at the end of the second day of trading. In computing the average direct and indirect cost, we compute this number for each individual firm for which we have the gross spread, other expenses, and the initial return, and then compute the average.

issuer with an offer price of $10.00 receives net proceeds of $8.90 on a share that trades at $11.20. Taking the difference between the market price and the amount realized of $8.90, the total direct and indirect costs amount to $2.30, which is 20.5 percent of the market value of $11.20. In Table 4 the average direct and indirect cost as a percentage of market value is 18.7 percent, since the average that is reported is the average of this percentage for each firm. (The average ratio of costs to market value is different from the ratio of the averages.) This number is less than the 21.2 percent that Ritter (1987) reports for firm commitment offerings from 1977 to 1982 for several reasons. First, our 1990–94 sample period reveals less underpricing than in 1977-1982. Second, we exclude offerings of less than $2 million, whereas he includes them. Third, spreads have experienced some downward movement the past fifteen years.[8] Still, the direct and indirect costs of going public are substantial.[9]

Note that we may be understating the extent of the economies of scale. This is because we are not including the value of any warrants granted to underwriters as part of their compensation. These warrants are common among small, speculative offerings underwritten by less-prestigious underwriters. Their inclusion would boost the average costs of the smallest offerings, but not the larger offerings. For evidence on the quantitative effect of this omission, see Barry, Muscarella, and Vetsuypens (1991) and Dunbar (1995).

While the average gross spread on IPOs is 7.31 percent, we find a large "bunching" at exactly 7.00 percent. Most issues with proceeds of $20–$60 million have a spread of exactly 7 percent, as shown in the Appendix table.

For IPOs, we include the indirect cost of underpricing in Table 4, but we do not include this as a cost for other security offerings. This is because of the lack of economically important underpricing effects for other offerings. Smith (1977) documents underpricing of 0.5 percent for SEOs. We suspect that much of this represents the practice of pricing the offering at the bid price, rather than the mean of the bid and the ask price, and the tendency to round down to the nearest eighth or integer. For example, if a stock traded at $30.125 bid and $30.375 ask, it would be common to set a $30.00 offer price. Depending upon which price had been the most recent transaction price, this would be measured as underpricing of either 0.4 percent or 1.2 percent. Barclay and Litzenberger (1988) report excess returns of 1.5 percent for SEOs during the month after issuing. Since companies typically issue after a large stock price run-up, it is not clear how much of this 1.5 percent is due to momentum effects, and how

[8]Calomiris and Raff (1995) report that for convertible bonds, the average spread in 1963–65 was 3.7 percent and in 1971–72 it was 3.2 percent. Our 1990–94 sample has an average spread of 2.9 percent.

[9]Beatty and Welch (1996) report the average direct and indirect costs for a sample of 980 IPOs from 1992 to 1994. Whereas we aggregate auditing, legal, printing, and other direct expenses, they report audit expenses and legal expenses separately. For all proceeds classes, legal expenses are slightly higher than auditor expenses.

TABLE 5. Number of Issues Containing an International Tranche for Domestic Operating Companies That Are Issuing, 1990–94.

	Equity				Bonds			
	IPOs Int'l Tranche?[a]		SEOs Int'l Tranche?		Convertible Int'l Tranche?		Straight Int'l Tranche?	
Proceeds ($ millions)	Yes	No	Yes	No	Yes	No	Yes	No
2–9.99	2	335	4	163	0	4	1	31
10–19.99	12	377	12	298	1	13	0	78
20–39.99	45	488	36	389	3	15	0	89
40–59.99	40	175	42	219	0	28	4	86
60–79.99	33	46	45	98	1	46	8	84
80–99.99	25	26	30	41	9	4	2	110
100–199.99	81	25	72	80	22	35	14	395
200–499.99	39	8	48	7	14	13	13	157
500–up	10	0	8	1	2	1	2	18
Total	287	1480	297	1296	52	159	44	1048

Notes: Closed-end funds (SIC 6726), REITs (SIC 6798), ADRs, and unit offerings are excluded from the sample. Rights offerings for SEOs are also excluded. Bond offerings do not include securities backed by mortgages and issues by Federal agencies (SIC 6011, 6019, 6111, and 999B). Only firm commitment offerings and nonshelf-registered offerings are included. Standard Industrial Classification (SIC) codes are from Securities Data Co. (SDC).

[a]If (TOTDOLAMT/PROCDS) > 1.05, the issue is treated as having an international tranche. TOTDOLAMT is the total proceeds raised globally, and PROCDS is the total proceeds raised in the United States.

much is due to issue effects. Kang and Lee (1996) document that convertible bonds are underpriced by about 1 percent on average. Straight bonds, especially those with high credit ratings, seem to be underpriced very little.

International Tranches

In Table 5 we report the frequency with which domestic operating companies include an international tranche in their offerings. Recall that we are excluding Eurobonds from our debt offerings and ADRs from our equity offerings. Inspection of the table reveals that equity offerings and convertibles that raise less than $60 million in domestic trading rarely include an international tranche. Straight debt offerings, no matter what their size, rarely include an international tranche. Now, foreign investors can always participate in a domestic offering regardless of whether it is explicitly marketed overseas. Thus, the existence/nonexistence of an international tranche largely reflects the degree to which the selling efforts are expanded to find international buyers. Domestic operating companies issuing debt with foreign buyers in mind frequently issue Eurobonds.[10]

Overallotment Options

The Rules of Fair Practice of the National Association of Security Dealers (NASD) permit firm commitment offerings to include an overallotment option, where more securities can be sold if demand is strong.[11] Since August 1983, the size of this overallotment option has been limited to 15 percent of the issue size. Investment bankers typically have thirty days to exercise this option. In practice, investment bankers typically presell at least 115 percent of the offering, and then stand ready to buy back the incremental 15 percent if demand is weak when some of the buyers immediately sell their securities (a practice known as "flipping").[12]

The NASD Rules of Fair Practice require that investment bankers sell securities at or below the stated offer price. Normally, all of the securities are sold at the offer price, but occasionally, if demand is weak, the investment banker winds up selling some of the securities below the offer price. In this arrangement the underwriter writes a put option to the issuing firm, with the value of this put included in the gross spread. The overallotment option can be viewed as a call option that the issuing firm has written, where investors hold this call.

On securities sold through the exercise of overallotment options, investment bankers collect the same gross spread as on the rest of the issue. However, since the direct expenses do not change, these fixed costs are spread over a larger issue size. Thus, the total direct cost numbers that we report would be lower if overallotment options were included in the gross proceeds. On the other hand, since overallotment options are generally exercised only if the issue is underpriced, the value of this call option is a cost to the issuing firm that we do not include in our total cost calculations.

In Table 6 we report the frequency with which overallotment options are used and the frequency with which they are exercised. Inspection of the table reveals that in recent years, essentially all IPOs have included an overallotment option. The vast majority of SEOs and convertibles include an overallotment option, but straight bond issues rarely do.

---

[10]The relative yields on Eurobonds versus domestic bonds also play a role in the decision of what to issue (see Kim and Stulz (1988)).

[11]Overallotment options are sometimes called Green Shoe options. The Green Shoe Company was apparently the first company to use one.

[12]See Schultz and Zaman (1994) for evidence on the exercise of overallotment options on IPOs. With IPOs, if the underwriter expects aftermarket demand to be weak, 135 percent of the issue may be presold, with the underwriter's taking a naked short position equal to the amount exceeding 115 percent of the offering. This allows the underwriter to support, or stabilize, the price by buying back the increment in open market purchases. These shares are then treated as if they were never issued. If the underwriter expects the price to jump, typically only 115 percent of the issue size will be presold, to avoid losing money on a naked short position.

TABLE 6. Number of Issues Containing an Overallotment Option, for Domestic Operating Companies That Are Issuing, 1990–94.

	Equity								Bonds							
	IPOs Overallotment Option?				SEOs Overallement Option?				Convertible Overallotment Option?				Straight Overallotment Option?			
	Yes			No[d]	Yes			No	Yes			No	Yes			No
	Sold?				Sold?				Sold?				Sold?			
Proceeds ($ millions)	Yes[a]	No[b]	?[c]		Yes	No	?		Yes	No	?		Yes	No	?	
2–9.99	159	115	51	12	100	41	21	5	0	0	4	0	1	0	4	27
10–19.99	198	151	40	0	209	58	38	5	1	2	8	3	2	1	4	71
20–39.99	306	164	60	3	269	100	49	7	4	2	8	4	6	0	9	74
40–59.99	123	67	25	0	173	50	33	5	6	6	13	3	1	0	1	88
60–79.99	45	27	7	0	81	37	21	4	21	6	16	4	3	0	0	89
80–99.99	25	17	9	0	44	9	15	53	10	0	3	0	0	1	1	10
100–199.99	54	34	16	2	96	24	28	4	23	2	28	4	4	1	3	401
200–499.99	21	17	8	1	35	4	14	2	7	2	15	3	3	1	1	165
500–up	6	0	3	1	6	2	1	0	0	0	3	0	0	0	1	19
Total	937	592	219	19	1013	325	220	35	72	20	98	21	20	4	24	1044

Notes: Closed-end funds (SIC 6726), REITs (SIC 6798), ADRs, and unit offerings are excluded from the sample. Rights offerings for SEOs are also excluded. Bond offerings do not include securities backed by mortgages and issues by Federal agencies (SIC 6011, 6019, 6111, and 999B). Only firm commitment offerings and nonshelf-registered offerings are included. Standard Industrial Classification (SIC) codes are from Securities Data Co. (SDC).

[a]If OVERAMT > 0 and OVERC = Yes, where OVERAMT is the amount that can be raised through the overallotment option and OVERC is "Yes" if any overallotment option is exercised.
[b]If OVERAMT > 0 and OVERC = No.
[c]If OVERAMT > 0 and OVERC = Missing.
[d]If OVERAMT = "-"; this may include offerings with missing data on OVERAMT.

APPENDIX. Interquartile Range of Direct Costs as a Percentage of Gross Proceeds for Equity (IPOs and SEOs) and Straight and Convertible Bonds Offered by Domestic Operating Companies, 1990–94.

	Equity				Bonds			
	IPOs		SEOs		Convertible Bonds		Straight Bonds	
Proceeds[a] ($ millions)	GS[b]	TDC[c]	GS	TDC	GS	TDC	GS	TDC
2–9.99	8.00–10.00	14.34–19.23	6.50–10.00	10.03–16.16	5.45–6.69	7.38–10.04	0.64–3.38	3.47–6.21
10–19.99	7.00–7.14	9.94–12.44	5.74–6.94	7.42–9.63	4.25–6.00	6.65–9.70	0.35–2.90	1.55–5.68
20–39.99	7.00–7.00	8.82–10.09	5.22–6.00	6.19–7.57	3.00–5.00	4.56–6.50	0.57–3.00	1.10–4.55
40–59.99	7.00–7.00	8.23–9.00	4.73–5.48	5.26–6.31	2.88–3.50	3.63–4.65	0.15–0.71	0.91–2.88
60–79.99	6.55–7.00	7.69–8.51	4.24–5.00	4.51–5.70	2.50–3.00	2.83–3.54	0.65–3.00	0.94–3.64
80–99.99	6.21–6.85	7.26–8.44	3.87–4.75	4.22–5.38	2.25–3.00	2.56–3.66	0.63–2.76	0.94–3.70
100–199.99	5.72–6.47	6.43–7.49	3.15–4.47	3.38–4.89	2.15–2.75	2.36–3.19	0.65–2.75	1.01–3.55
200–499.99	5.29–5.86	5.92–6.78	2.79–3.58	2.92–3.79	1.25–2.50	1.40–2.69	0.65–2.63	1.43–3.16
500–up	5.00–5.37	5.33–5.95	2.75–3.00	2.82–3.17	1.00–2.50	1.11–2.60	0.29–2.75	1.05–3.18
Total	7.00–7.05	8.57–12.04	4.51–6.08	5.12–8.20	2.25–3.00	2.66–3.96	0.60–2.75	1.02–3.60

Notes: Closed-end funds (SIC 6726), REITs (SIC 6798), ADRs, and unit offerings are excluded from the sample. Rights offerings for SEOs are also excluded. Bond offerings do not include securities backed by mortgages and issues by Federal agencies (SIC 6011, 6019, 6111, and 999B). Only firm commitment offerings and nonshelf-registered offerings are included. Standard Industrial Classification (SIC) codes are from Securities Data Co. (SDC).

[a]Total proceeds raised in the United States, excluding proceeds from the exercise of overallotment options (SDC variable: PROCDS).
[b]Gross spreads as a percentage of total proceeds (including management fee, underwriting fee, and selling concession) (SDC variable: GPCTP).
[c]Total direct costs as a percentage of total proceeds (total direct costs are the sum of gross spreads and other direct expenses).

110138-OPC-POD-60-187

The frequency with which overallotment options are exercised varies across security type. In Table 6 we use the SDC classification where an overallotment option is considered to be exercised as long as at least part of it is exercised. In practice, most overallotment options are for 15 percent of the issue size. Most commonly, either all or none of the additional shares are sold, but sometimes only part of the overallotment option is exercised. On securities sold as part of an overallotment option, the spread is the same as on the rest of the issue.

IV. Conclusions

Firms have many choices for financing their activities: internal versus external, private versus public, and debt versus equity. This article focuses on public external financing and documents the cost of this financing from 1990 to 1994. We report the direct costs of raising capital for IPOs, SEOs, convertible bonds, and straight bonds. These are, respectively, 11.0 percent, 7.1 percent, 3.8 percent, and 2.2 percent of the proceeds. We find substantial economies of scale for all types of securities, although for straight bond offerings, these are largely exhausted for proceeds over $40 million. Spreads on bonds are sensitive to credit quality, with gross spreads more than 200 basis points higher on noninvestment-grade issues. Except for bonds, most large issues include an international tranche.

References

Barclay, M. J. and R. H. Litzenberger, 1988, Announcement effects of new equity issues and the use of intraday price data, *Journal of Financial Economics* 21, 71–99.

Barry, C., C. J. Muscarella, and M. R. Vetsuypens, 1991, Underwriter warrants, underwriter compensation, and the costs of going public, *Journal of Financial Economics* 29, 113–35.

Beatty, R P. and I. Welch, 1996, Issuer expenses and legal liability in initial public offerings, *Journal of Law and Economics*, Forthcoming.

Bhagat, S. and P. A. Frost, 1986, Issuing costs to existing shareholders in competitive and negotiated underwritten public utility equity offerings, *Journal of Financial Economics* 15, 233–59.

Calomiris, C. W. and D. M. G. Raff, 1995, The evolution of market structure, information, and spreads in American investment banking, in M. B. Bordo and R. Sylla, eds., *Anglo-American Finance: Financial Markets and Institutions in 20th Century North America and the U.K.* (Business One-Irwin, Homewood, IL), 103–60.

Dunbar, C. G., 1995, The use of warrants as underwriter compensation in initial public offerings, *Journal of Financial Economics* 38, 59–78.

Hansen, R. S., 1988, The demise of the rights issue, *Review of Financial Studies* 1, 289–309.

Hansen, R. S. and J. Pinkerton, 1982, Direct equity financing: A resolution of a paradox, *Journal of Finance* 37, 651–65.

Hansen, R. S. and P. Torregrosa, 1992, Underwriter compensation and corporate monitoring, *Journal of Finance* 47, 1537–55.

Kang, J. and Y. Lee, 1996, The pricing of convertible debt offerings, *Journal of Financial Economics*, Forthcoming.

Kim, Y. C. and R. M. Stulz, 1988, The Eurobond market and corporate financial policy: A test of the clientele hypothesis, *Journal of Financial Economics* 22, 189–205.

Ritter, J. R., 1987, The costs of going public, *Journal of Financial Economics* 19, 269–81.

Schultz, P. H. and M. A. Zaman, 1994, Aftermarket support and the pricing of initial public offerings, *Journal of Financial Economics* 35, 199–219.

Smith, C. W., 1977, Alternative methods for raising capital: Rights versus underwritten offerings, *Journal of Financial Economics* 5, 273–307.

Journal of Financial Economics 7 (1979) 163–195. © North-Holland Publishing Company

THE EFFECT OF PERSONAL TAXES AND DIVIDENDS ON CAPITAL ASSET PRICES

Theory and Empirical Evidence

Robert H. LITZENBERGER*

Stanford University, Stanford, CA 94305, USA

Krishna RAMASWAMY*

Bell Telephone Laboratories, Murray Hill, NJ 07974, USA

Received July 1978, revised version received March 1979

This paper derives an after tax version of the Capital Asset Pricing Model. The model accounts for a progressive tax scheme and for wealth and income related constraints on borrowing. The equilibrium relationship indicates that before-tax expected rates of return are linearly related to systematic risk and to dividend yield. The sample estimates of the variances of observed betas are used to arrive at maximum likelihood estimators of the coefficients. The results indicate that, unlike prior studies, there is a strong positive relationship between dividend yield and expected return for NYSE stocks. Evidence is also presented for a clientele effect.

1. Introduction

The effect of dividend policy on the prices of equity securities has been an issue of interest in financial theory. The traditional view was that investors prefer a current, certain return in the form of dividends to the uncertain prospect of future dividends. Consequently, they bid up the price of high yield securities relative to low yield securities [see Cottle, Dodd and Graham (1962) and Gordon (1963)]. In their now classic paper Miller and Modigliani (1961) argued that in a world without taxes and transactions costs the dividend policy of a corporation, given its investment policy, has no effect on the price of its shares. In a world where capital gains receive preferential treatment relative to dividends, the Miller–Modigliani 'irrelevance proposition' would seem to break down. They argue, however, that since tax rates vary across investors each corporation would attract to itself a clientele of investors that most desired its dividend policy. Black and Scholes (1974) assert that corporations would adjust their payout policies until in equilib-

*We thank Roger Clarke, Tom Foregger, Bill Schwert, William Sharpe, and the referee, Michael Brennan, for helpful comments, and Jim Starr for computational assistance. Any remaining errors are the authors' responsibility.

rium the spectrum of policies offered would be such that any one firm is unable to affect the price of its shares by (marginal) changes in its payout policy.

In the absence of taxes, capital asset pricing theory suggests that individuals choose mean-variance efficient portfolios. Under personal income taxes, individuals would be expected to choose portfolios that are mean-variance efficient in after-tax rates of return. However, the tax laws in the United States are such that some economic units (for example, corporations) would seem to prefer dividends relative to capital gains. Other units (for example, non-profit organizations) pay no taxes and would be indifferent to the level of yield for a given level of expected return. The resulting effect of dividend yield on common stock prices seems to be an empirical issue.

Brennan (1973) first proposed an extended form of the single period Capital Asset Pricing Model that accounted for the taxation of dividends. Under the assumption of proportional individual tax rates (not a function of income), certain dividends, and unlimited borrowing at the riskless rate of interest (among others) he derived the following equilibrium relationship:

$$E(\tilde{R}_i)-r_f=b\beta_i+\tau(d_i-r_f), \tag{1}$$

where $\tilde{R}_i$ is the before tax total return to security i, β_i is its systematic risk, $b=[E(\tilde{R}_m)-r_f-\tau(d_m-r_f)]$ is the after-tax excess rate of return on the market portfolio, r_f is the return on a riskless asset, d_i is the dividend yield on security i, and the subscript m denotes the market portfolio. τ is a positive coefficient that accounts for the taxation of dividends and interest as ordinary income and taxation of capital gains at a preferential rate.

In empirical tests [of the form (1)] to date, the evidence has been inconsistent. Black and Scholes (1974, p. 1) conclude that

> '...it is not possible to demonstrate that the expected returns on high yield common stocks differ from the expected returns on low yield common stocks either before or after taxes.'

Alternatively, stated in terms of the Brennan model, their tests were not sufficiently powerful either to reject the hypothesis that $\tau=0$ or to reject the hypothesis that $\tau=0.5$. Rosenberg and Marathe (1978) attribute the lack of power in the Black–Scholes tests to (a) the loss in efficiency from grouping stocks into portfolios and (b) the inefficiency of their estimating procedures, which are equivalent to Ordinary Least Squares. Using an instrumental variables approach to the problem of errors in variables and a more complete specification of the variance-covariance matrix (of disturbances in the regression), Rosenberg and Marathe find that the dividend term is statistically significant. Both the Rosenberg and Marathe and the Black and Scholes studies use an average dividend yield from the prior twelve month

period as a surrogate for the expected dividend yield. Since most dividends are paid quarterly, their proxy understates the expected dividend yield in ex-dividend months and overstates it in those months that a stock does not go ex-dividend, thereby reducing the efficiency of the estimated coefficient on the dividend yield term. Both studies (Rosenberg and Marathe in using instrumental variables, and Black–Scholes in grouping) sacrifice efficiency to achieve consistency.

The present paper derives an after-tax version of the Capital Asset Pricing Model that accounts for a progressive tax scheme and both wealth and income related constraints on borrowing. Alternative econometric procedures are used to test the implications of this model. Unlike prior tests of the CAPM, the tests here use the variance of the observed betas to arrive at maximum likelihood estimators of the coefficients. Consistent estimators are obtained without loss of efficiency. Also, for ex-dividend months the expected dividend yield based on prior information is used, and for other months the expected dividend yield is set equal to zero. While the estimate of the coefficient of dividend yield is of the same order of magnitude as that found in Black and Scholes, and lower than that found by Rosenberg and Marathe, the t-value is substantially larger, indicating a substantial increase in efficiency. Furthermore, the tests are consistent with the existence of a clientele effect, indicating that the aversion for dividends relative to capital gains is lower for high yield stocks and higher for low yield stocks. This is consistent with the Elton and Gruber (1970) empirical results on the ex-dividend behavior of common stocks.

2. Theory

This section derives a version of the Capital Asset Pricing Model that accounts for the tax treatment of dividend and interest income under a progressive taxation scheme. Two types of constraints on individual borrowing are imposed. The first constrains the maximum interest on riskless borrowing to be equal to the individual's dividend income, and the second is a margin requirement that restricts the fraction of security holdings that may be financed through borrowing. In previous published work, Brennan (1973) derives an after-tax version of the Capital Asset Pricing Model with unlimited borrowing and with constant tax rates which may vary across individuals.[1] Under his model when interest on borrowing exceeds dividend income the investor would pay a negative tax. The theoretical model

[1]Brennan (1970) also derives a model with a progressive tax scheme. However, he neither considers constraints on borrowing nor the limiting of interest deduction on margin borrowing to dividend income. Consideration of the limit on the interest tax deduction to dividend income combined with a positive capital gains tax would result in a preference for dividends by those individuals whose interest payments exceed their dividend income.

developed here may be viewed as an extension of the Brennan analysis to account for constraints on borrowing along with a progressive tax scheme. Special cases of the model are examined, where the income related constraint and/or the margin constraint on individual borrowing are removed.

The following assumptions are made:

(A.1) Individuals' Von Neumann–Morgenstern utility functions are monotone increasing strictly concave functions of after-tax end of period wealth.

(A.2) Security rates of return have a multivariate normal distribution.

(A.3) There are no transactions costs, and no restrictions on the short sale of securities, and individuals are price takers.

(A.4) Individuals have homogeneous expectations.

(A.5) All assets are marketable.

(A.6) A riskless asset, paying a constant rate r_f, exists.

(A.7) Dividends on securities are paid at the end of the period and are known with certainty at the beginning of the period.

(A.8) Income taxes are progressive and the marginal tax rate is a continuous function of taxable income.

(A.9) There are no taxes on capital gains.

(A.10) Constraints on individuals' borrowing are of the form:

(i) A constraint that the interest on borrowing cannot exceed dividend income, called the income constraint on borrowing, and/or

(ii) a margin constraint that the individual's net worth be at least a given fraction of the market value of his holdings of risky securities.

Assumptions (A.1) through (A.6) are standard assumptions of the Capital Asset Pricing Model. Assumptions (A.1) and (A.2) taken together imply that preferences can be described over the mean and the variance of after-tax end of period wealth. Under these conditions individuals prefer more mean return and are averse to the variance of return. The individual's marginal rate of substitution between the mean and variance of after-tax end of period wealth, at the optimum, can be written as the ratio of his global risk tolerance to his initial period wealth. That is, if $u_k(W_1^k)$ is the kth individual's utility function in terms of after-tax end of period wealth, $f^k(\mu_k, \sigma_k^2)$ is his objective function in terms of the mean and variance of the after-tax portfolio return, and W^k is his initial wealth,

$$f_1^k / -2f_2^k = \theta^k / W^k, \tag{2}$$

where $\theta^k = -E(u')/E(u'')$ is the individual's global risk tolerance at the optimum [see Gonzalez-Gaverra (1973) and Rubinstein (1973)]. (A.7) implies

that dividends are announced at the beginning of the period and paid at its end. Since firms display relatively stable dividend policies this may be a reasonable approximation for a monthly holding period.

Assumption (A.8) closely resembles the tax treatment of ordinary dividends in the U.S. The $100 dividend exclusion is ignored, since the small magnitude of the exclusion implies that for the majority of stockholders the marginal tax rate applicable to ordinary income is the same as that applied to dividends. Assumption (A.9) abstracts from the effects of capital gains taxes. Since capital gains are taxed only upon realization, their treatment in a single period model is not possible. It is, however, straightforward to model a capital gains tax on an accrual basis [see Brennan (1973)]. Since most capital gains go unrealized for long periods, this would tend to overstate the effect of the actual tax. Noting that the ratio of realizations to accruals is small, and that capital gains are exempt from tax when transferred by inheritance, Bailey (1969) has argued that the effective tax is rather small.

Under assumption (A.8), the kth individual's average tax rate, t^k, is a non-decreasing function of his taxable end of period income Y_1^k,

$$\begin{aligned} t^k &= g(Y_1^k), \\ g(0) = 0, \qquad g'(Y_1^k) &= 0 \quad \text{for} \quad Y_1^k \leqq 0, \\ &> 0 \quad \text{for} \quad Y_1^k > 0. \end{aligned} \tag{3}$$

The kth individual's marginal tax rate, written T^k, is the first derivative of taxes paid with respect to taxable income. This is equal to the average tax rate plus the product of taxable income and the derivative of the average tax rate,

$$T^k \equiv d(t^k Y_1^k)/dY_1^k = t^k + Y_1^k g'(Y_1^k). \tag{4}$$

The margin constraint in assumption (A.10–ii) resembles institutional margin restrictions. By (A.10–i), borrowing is constrained up to a point where interest paid equal dividends received. This constraint incorporates the casual empirical observation that loan applications require information on income (which this constraint accounts for) in addition to information on wealth (which the margin constraint accounts for). One or both of the constraints may be binding, for a given individual. This formulation allows the analysis of an equilibrium with both constraints, with only one of them imposed or with no borrowing constraints.

The following notation is employed:

R_i = the total before tax rate of return on security i, equal to the ratio of the value of the security at the end of the period plus dividends over its current value, less one.

d_i = the dividend yield on security i, equal to the dollar dividend divided by the current price,
X_i^k = the fraction of the kth individual's wealth invested in the ith risky asset, $i = 1, 2, \ldots, N$ (a negative value is a short sale),
X_f^k = the fraction of the kth individual's wealth invested in the safe asset (a negative value indicates borrowing),
$\tilde{R}_p^k$ = the before-tax rate of return on the kth individual's portfolio,
W^k = the kth individual's initial wealth, and
$f^k(\mu_k, \sigma_k^2)$ = the kth individual's expected utility function defined over the mean and variance of after-tax portfolio return, μ_k and σ_k^2, respectively.

The kth individual's ordinary income is then

$$Y_1^k = W^k\left(\sum_i X_i^k d_i + X_f^k r_f\right). \tag{5}$$

The mean after-tax return on the individual's portfolio is

$$\mu_k = \sum X_i^k E(\tilde{R}_i) + X_f^k r_f - t^k\left(\sum_i X_i^k d_i + X_f^k r_f\right). \tag{6}$$

and under assumption (A.7) the variance of after-tax return is

$$\begin{aligned}\sigma_k^2 &= \sum_j \sum_i X_i^k X_j^k \operatorname{cov}(\tilde{R}_i - d_i t^k, \tilde{R}_j - d_j t^k)\\ &= \sum_j \sum_i X_i^k X_j^k \operatorname{cov}(\tilde{R}_i, \tilde{R}_j).\end{aligned} \tag{7}$$

By assumption (A.10-i) the income constraint on borrowing is

$$W^k\left\{\sum_i X_i^k d_i + X_f^k r_f\right\} \geqq 0, \tag{8}$$

and the margin constraint on borrowing is

$$W^k\left\{(1-\alpha)\sum_i X_i^k + X_f^k\right\} \geqq 0, \tag{9}$$

where α; $0 < \alpha < 1$, is the margin requirement on the individual. As pointed out earlier, one or both of these constraints may be binding.

The kth individual's optimization problem is stated in terms of the

following Lagrangian:

$$\begin{aligned}\mathscr{L}^k \equiv\ & f^k(\mu_k, \sigma_k^2) + \lambda_1^k\left[1 - \sum_i X_i^k - X_f^k\right]\\ & + \lambda_2^k\left[\sum_i X_i^k d_i + X_f^k r_f - S_2^k\right] + \lambda_3^k\left[(1-\alpha)\sum_i X_i^k + X_f^k - S_3^k\right],\end{aligned} \tag{10}$$

where

λ_1^k = the Lagrange multiplier on the kth individual's budget;
λ_2^k, S_2^k = the Lagrange multiplier and non-negative slack variable for the income related constraint on the kth individual's borrowing, respectively (when the constraint is binding $\lambda_2^k > 0$ and $S_2^k = 0$, and when it is not binding $\lambda_2^k = 0$ and $S_2^k \geqq 0$), and
λ_3^k, S_3^k = the Lagrange multipler and non-negative slack variables for the margin constraint on the kth individual's borrowing, respectively; again if the constraint is binding (not binding), $\lambda_3^k > (=)\ 0$ and $S_3^k = (\geqq)\ 0$.

The stationary points satisfy the following first order conditions:

$$\begin{aligned}\frac{\partial \mathscr{L}^k}{\partial X_i^k} =\ & f_1^k\{E(\tilde{R}_i) - [t^k + Y_1^k g'(Y_1^k)]d_i\} - \lambda_1^k + \lambda_2^k d_i\\ & + \lambda_3^k(1-\alpha) + 2f_2^k \sum_j X_j^k \operatorname{cov}(\tilde{R}_i, \tilde{R}_j) = 0, \qquad i = 1, 2, \ldots, N,\end{aligned} \tag{11}$$

$$\frac{\partial \mathscr{L}^k}{\partial X_f^k} = f_1^k\{r_f - [t^k + Y_1^k g'(Y_1^k)]r_f\} - \lambda_1^k + \lambda_2^k r_f + \lambda_3^k = 0, \tag{12}$$

where $f_1^k \equiv \partial f^k(\mu_k, \sigma_k^2)/\partial \mu_k$, $f_2^k \equiv \partial f^k(\mu_k, \sigma_k^2)/\partial \sigma_k^2$. The other first order conditions are the constraints and specify the signs of the Lagrangian multipliers and are omitted here. The progressive nature of the tax scheme [assumption (A.8)] ensures that the mean variance efficient frontier in after-tax terms is concave, and this together with risk aversion from assumption (A.8) is sufficient to guarantee the second order conditions for a maximum.

Recall the following relationships: (i) the marginal tax rate, $T^k = [t^k + Y_1^k g'(\cdot)]$, (ii) the covariance $\sum_j X_j^k \operatorname{cov}(\tilde{R}_i, \tilde{R}_j) = \operatorname{cov}(\tilde{R}_i, \tilde{R}_p^k)$, and (iii) the global risk tolerance $\theta^k = W^k(f_1^k / -2f_2^k)$. Subtracting relation (12) from relation (11) and re-arranging terms yields

$$\begin{aligned}\{E(\tilde{R}_i) - r_f] =\ & \alpha(\lambda_3^k / f_1^k) + (W^k/\theta^k)\operatorname{cov}(\tilde{R}_i, \tilde{R}_p^k)\\ & + [T^k - (\lambda_2^k / f_1^k)](d_i - r_f).\end{aligned} \tag{13}$$

Relation (13) must be satisfied for the individual's portfolio optimum.

Market equilibrium requires that relation (13) holds for all individuals, and that markets clear. For markets to clear all assets have to be held which implies the conservation relation (14) that requires the value weighted average of all individuals' portfolios be equal to the market portfolio,

$$\sum_k (W^k/W^m)\tilde{R}_p^k = \tilde{R}_m, \tag{14}$$

or

$$\sum_k W^k \tilde{R}_p^k = W^m \tilde{R}_m,$$

where

$$\sum_k W^k \equiv W^m.$$

Multiplying both sides of relation (13) by θ^k, summing over all individuals, using the conservation relation (14) and re-arranging terms yields

$$E(\tilde{R}_i) - r_f = a + b\beta_i + c(d_i - r_f), \tag{15}$$

where

$$\beta_i \equiv \operatorname{cov}(\tilde{R}_i, \tilde{R}_m)/\operatorname{var}(\tilde{R}_m),$$

$$a \equiv \alpha \sum_k (\theta^k/\theta^m)(\lambda_1^k/f_1^k),$$

$$b \equiv \operatorname{var}(\tilde{R}_m)(W^m/\theta^m),$$

$$c \equiv \sum_k (\theta^k/\theta^m)[T^k - (\lambda_2^k/f_1^k)],$$

$$\theta^m \equiv \sum_k \theta^k.$$

The term 'a', the intercept of the implied security market plane, is the fractional margin requirement α times the weighted average of the ratios of individual shadow prices on the margin constraint and the expected marginal utility of mean return. The weights, (θ^k/θ^m), are proportional to individuals' global risk tolerances. When $\alpha > 0$ and the constraint is binding for some individuals, $\lambda_1^k > 0$ for some k, a is positive. In the absence of margin requirements ($\alpha = 0$) or when the margin constraint is not binding for all individuals, ($\lambda_1^k = 0$ for all k), $a = 0$.

Interpreting eq. (15), 'a' is the excess return on a zero beta portfolio (relative to the market) whose dividend yield is equal to the riskless rate, i.e., $a = E(\tilde{R}_z) - r_f$. The term 'b', the coefficient on beta is equal to the product of the variance of the rate of return on the market portfolio and global market relative risk aversion, i.e., $b = \operatorname{var}(\tilde{R}_m)(W^m/\theta^m)$. Since relation (15) also holds for the market portfolio, b may be alternatively expressed as $b = [E(\tilde{R}_m) - r_f - c(d_m - r_f) - a]$. If '$c$' is interpreted as a tax rate, b may be viewed as the expected after-tax rate of return on a hedge portfolio which is long the market portfolio and short a portfolio having a zero beta and a dividend yield equal to the riskless rate of interest; i.e., $b = [E(\tilde{R}_m) - E(\tilde{R}_z) - c(d_m - d_z)]$. The term '$c$' is a weighted average of individual's marginal tax rates $(\sum_k(\theta^k/\theta^m)T^k)$, less the weighted average of the individual's ratios of the shadow price on the income related borrowing constraint and the expected marginal utility of mean portfolio return $\sum_k(\theta^k/\theta^m)(\lambda_2^k/f_1^k)$. For the cases where the income related margin constraint is either non-existent or non-binding for all individuals, c is simply the weighted average of marginal tax rates, and is positive. Otherwise, the sign of 'c' depends on the magnitudes of these two terms. Define B as the set of indices of those individuals k for whom the income related constraint is binding: and define N (*not* B) as the set of indices for which the constraint is non-binding. Now for $k \in B$, $\lambda_2^k > 0$, $Y_1^k = 0$ and $T^k = t^k = 0$. And for $k \in N$, $\lambda_2^k = 0$, $Y_1^k \geqq 0$ and $T^k \geqq t^k \geqq 0$. Hence

$$c = \sum_{k \in N} \frac{\theta^k}{\theta^m} T^k - \sum_{k \in B} \frac{\theta^k}{\theta^m} \frac{\lambda_2^k}{f_1^k}. \tag{16}$$

The individuals in N may be viewed as a clientele that prefers capital gains to dividends. The individuals in B may be viewed as a clientele that shows a preference for dividends: in the context of this model, these individuals wish to borrow more than the income related constraint allows them, and increased dividends serve to increase their debt capacity without additional tax obligations. To this point corporate dividend policies have been treated as exogenous in this model.

Now consider supply adjustments by value maximizing firms. If $c > 0$ ($c < 0$) firms could increase their market values by decreasing (increasing) cash dividends and increasing (decreasing) share repurchases or decreasing (increasing) external equity flotations. Value maximizing firms (in absence of any restrictions the IRS may impose) would adjust the supply of dividends until an equilibrium was obtained where

$$\sum_{k \in N} (\theta^k/\theta^m)T^k = \sum_{k \in B} (\theta^k/\theta^m)(\lambda_2^k/f_1^k). \tag{17}$$

When condition (17) is satisfied an individual firm's dividend decision does not affect its market value, $c=0$ and dividend yield has no effect on the before tax rate of return on any security.[2]

Under unrestricted supply effects, $c=0$ and the equilibrium relationship (15) reduces to the before tax zero beta version of the Capital Asset Pricing Model:

$$E(\tilde{R}_i)=(a+r_f)(1-\beta_i)+E(\tilde{R}_m)\beta_i. \tag{18}$$

Note that this obtains in the presence of taxes. Long (1975) has studied conditions under which the before tax and after-tax mean variance efficient frontiers are identical for any individual. He does not, however, study the equilibrium as is done here: for even though the before tax and after-tax individual mean variance frontiers are not identical, (18) demonstrates that prices are found as if there is no tax effect.

In the case where there are no margin constraints, $a=0$, and relation (18) reduces to the before tax traditional Sharpe–Lintner version of the Capital Asset Pricing Model,

$$E(\tilde{R}_i)=r_f+[E(\tilde{R}_m)-r_f]\beta_i. \tag{19}$$

Return now to the case where the income related borrowing constraint is absent. Then, in (16), $c=\sum_k T^k(\theta^k/\theta^m)\equiv T^m$, the 'market' marginal tax bracket: and the relation reduces to an after-tax version of the Black (1972), Lintner (1965), Vasicek (1971) zero beta model,

$$E(\tilde{R}_i)-T^m d_i=[r_f(1-T^m)+a](1-\beta_i)+(E(\tilde{R}_m)-T^m d_m)\beta_i. \tag{20}$$

When there is no margin constraint or when it is non-binding for all individuals, $a=0$, and relation (20) reduced to an after-tax version of the Sharpe (1964), Lintner (1965) model,

$$E(\tilde{R}_i)-T^m d_i=[r_f(1-T^m)]+[E(\tilde{R}_m)-T^m d^m-r_f(1-T^m)]\beta_i. \tag{21}$$

However, in none of these cases is T^m a weighted average of individual

[2]Note, however, that this equilibrium, where dividends do not affect before tax returns, may not exist. For example, the income constraint may be binding for no one even when dividends are zero. If all individuals had the same endowments and had the same utility functions this constraint would be non-binding for all individuals.

This argument is in the spirit of the 'supply effect' alluded to in Black and Scholes (1974). Unlike the recent argument in Miller and Scholes (1977) for a zero dividend effect, the present argument does not depend on an artificial segmentation of accumulators and non-accumulators, and the existence of tax-sheltered lending opportunities with zero administrative costs. The major problem with the argument here is that with the existence of two distinct clienteles, one preferring higher dividends and the other preferring lower dividends, shareholders would not agree on the direction in which firms should change their dividend. Thus the assertion of value maximizing behavior by firms does not have a strong theoretical basis.

average tax rates. It is only when taxes are simply proportional to income that $T^k=t^k$, and relation (21) is identical to the equilibrium implied by Brennan (1973), who assumes a constant tax rate that may differ across investors.

3. Empirical tests

From the theory, the equilibrium specification to be tested is

$$E(\tilde{R}_i)-r_f=a+b\beta_i+c(d_i-r_f). \tag{22}$$

The hypotheses are $a>0$, $b>0$ and in the absence of the income related constraint on borrowing $c>0$.

In obtaining econometric estimates of a, b and c, two problems arise. The first is that expectations are not directly observed. The usual procedure is to assume that expectations are rational and that the parameters a, b and c are constant over time; the realized returns are used on the left-hand side

$$\tilde{R}_{it}-r_{ft}=\gamma_0+\gamma_1\beta_{it}+\gamma_2(d_{it}-r_{ft})+\tilde{\varepsilon}_{it}, \qquad i=1,2,\ldots,N_t, \quad t=1,2,\ldots,T, \tag{23}$$

where $\tilde{R}_{it}$ is the return of security i in period t, β_{it} and d_{it} are the systematic risk and the dividend yield of security i in period t respectively. The disturbance term $\tilde{\varepsilon}_{it}$ is $\tilde{R}_{it}-E(\tilde{R}_{it})$, the deviation of the realized return from its expected value. The coefficients γ_0, γ_1 and γ_2 correspond to a, b and c. The variance of the column vector of disturbance terms, $\tilde{\varepsilon}\equiv\{\tilde{\varepsilon}_{it}: i=1,2,\ldots,N_t, t=1,\ldots,T\}$, is not proportional to the identity matrix, since contemporaneous covariances between security returns are non-zero and return variances differ across securities. (Note that in order to conserve space '$\{\cdot\}$' is used to denote a column vector.) This means that ordinary least squares (OLS) estimators are inefficient, for either a cross-sectional regression in month t, or a pooled time series and cross-sectional regression. The computed variance of the OLS estimator (based on the assumption that the variance of $\tilde{\varepsilon}$ is proportional to the identity matrix) is not equal to the true variance of the estimator.

The second problem is that the true population β_{it}'s are unobservable. The usual procedure uses an estimate from past data, and this estimate has an associated measurement error. This means that the OLS estimates will be biased and inconsistent. The method used in tackling these problems is discussed in this section.

To fix matters, assume that data exist for rates of return, true betas and for dividend yields in periods t, $i=1,2,\ldots,N_t$ securities in each period t, $t=1,\ldots,T$. Define the vector of realized excess returns as

$$\bar{R} \equiv \{\bar{R}_1 \bar{R}_2, \ldots, \bar{R}_t, \ldots, \bar{R}_T\},$$

where

$$\bar{R}_t \equiv \{(\tilde{R}_{1t} - r_{ft})(\tilde{R}_{2t} - r_{ft})(\tilde{R}_{it} - r_{ft}), \ldots, (\tilde{R}_{N_t t} - r_{ft})\},$$

and the matrices X of explanatory variables as

$$X \equiv \{X_1 X_2, \ldots, X_t, \ldots, X_T\},$$

where

$$X_t \equiv \begin{bmatrix} 1 & \beta_{1t} & (d_{1t} - r_{ft}) \\ 1 & \beta_{2t} & (d_{2t} - r_{ft}) \\ \vdots & \vdots & \vdots \\ 1 & \beta_{N_t t} & (d_{N_t t} - r_{ft}) \end{bmatrix}.$$

By defining the vector of regression coefficients as $\Gamma = \{\gamma_0 \gamma_1 \gamma_2\}$ one can write the pooled time series and cross-sectional regression as

$$\bar{R} = X\Gamma + \tilde{\varepsilon}, \tag{24}$$

where

$$\tilde{\varepsilon} \equiv \{\tilde{\varepsilon}_1 \tilde{\varepsilon}_2, \ldots, \tilde{\varepsilon}_t, \ldots, \tilde{\varepsilon}_T\},$$

and

$$\tilde{\varepsilon}_t \equiv \{\tilde{\varepsilon}_{1t} \tilde{\varepsilon}_{2t}, \ldots, \tilde{\varepsilon}_{it}, \ldots, \tilde{\varepsilon}_{N_t t}\}.$$

It is assumed that

$$E(\tilde{\varepsilon}) = 0,$$

and that

$$E(\tilde{\varepsilon}_t \tilde{\varepsilon}_t') = V_t,$$

some symmetric positive definite matrix of order $(N_t \times N_t)$. It is also assumed that security returns are serially uncorrelated, so that

$$E(\tilde{\varepsilon}_{it} \tilde{\varepsilon}_{js}) = 0 \quad \text{for} \quad t \neq s.$$

This means that the variance–covariance matrix $V \equiv E(\tilde{\varepsilon}\tilde{\varepsilon}')$ is block diagonal, with the off-diagonal blocks being zero. The matrices V_t appears along the diagonal of V.

It is well known that the estimator for Γ which is linear in $\bar{R}$, unbiased and has minimum variance is unique, and is given by the Aitken or Generalized Least Squares estimator (GLS),

$$\hat{\Gamma} = (X'V^{-1}X)^{-1}X'V^{-1}\bar{R}. \tag{25}$$

From the block diagonal nature of V, it follows that V^{-1} is also block diagonal. The matrices V_t^{-1}, $t = 1, 2, \ldots, T$, appear along the diagonal of V^{-1}, with the off-diagonal blocks being zero. Assuming that Γ is an intertemporal constant, Γ can be estimated by efficiently pooling T independent GLS estimates of Γ, namely $\hat{\Gamma}_1, \hat{\Gamma}_2, \ldots, \hat{\Gamma}_t, \ldots, \hat{\Gamma}_T$, obtained by using cross-sectional data in periods $1, 2, \ldots, t, \ldots, T$,

$$\hat{\Gamma}_t = (X_t' V_t^{-1} X_t)^{-1} X_t' V_t^{-1} \bar{R}_t, \qquad t = 1, 2, \ldots, T. \tag{26}$$

That is, the monthly estimators $\hat{\gamma}_{kt}$ for γ_k, $k = 0$, 1 or 2, are serially uncorrelated, and the pooled GLS estimator $\hat{\gamma}_k$ is found as the weighted mean of the monthly estimates, where the weights are inversely proportional to the variances of these estimates,

$$\hat{\gamma}_k = \sum_{t=1}^{T} Z_{kt} \hat{\gamma}_{kt}, \tag{27}$$

$$\operatorname{var}(\hat{\gamma}_k) = \sum_{t=1}^{T} Z_{kt}^2 \operatorname{var}(\hat{\gamma}_{kt}), \tag{28}$$

$$Z_{kt} = [\operatorname{var}(\hat{\gamma}_{kt})]^{-1} \Big/ \sum_t [\operatorname{var}(\hat{\gamma}_{kt})]^{-1}. \tag{29}$$

For some of the results presented in section 4 each $\hat{\gamma}_{kt}$ is assumed to be drawn from a stationary distribution, and the estimates of $\hat{\gamma}_k$ and its variance are

$$\hat{\gamma}_k = \sum_{t=1}^{T} (\hat{\gamma}_{kt}/T), \tag{30}$$

$$\hat{\sigma}^2(\hat{\gamma}_k) = \left[\sum_{t=1}^{T} (\hat{\gamma}_{kt} - \hat{\gamma}_k)^2 / T(T-1)\right], \qquad k = 0, 1, 2. \tag{31}$$

A useful portfolio interpretation can be given to each of the GLS estimators $\hat{\Gamma}_t$ in (26). Choose any matrix numbers of order $N_t \times N_t$, say W_t^{-1}, such that $(X_t' W_t^{-1} X_t)^{-1}$ exists. Construct an estimator, using cross-sectional data in period t, as

$$(X_t' W_t^{-1} X_t)^{-1} X_t' W_t^{-1} \tilde{R}_t. \tag{32}$$

This estimator is linear in $\tilde{R}_t$ and unbiased for Γ. This estimator is a linear combination of realized security excess returns in period t. From the fact that

$$(X_t' W_t^{-1} X_t)^{-1} X_t' W_t^{-1} X_t = I, \tag{33}$$

where I is the identity matrix, it follows that the estimator for γ_0 in (32) is the realized excess return on a zero beta portfolio having a dividend yield equal to the riskless rate. Similarly, the estimator for γ_1 is the realized excess return on a hedge portfolio that has a beta of one and dividend yield equal to zero; and that for γ_2 is the realized excess return on a hedge portfolio having a zero beta and a dividend yield equal to unity. This interpretation[3] can be given to any estimator of the form (32). When W_t^{-1} (or, equivalently, the portfolio weights discussed above) is chosen so as to minimize the variance of the portfolio return, the resulting estimator is the GLS estimator. This is because portfolio estimates as in (32) are linear and unbiased by construction, and by the Gauss–Markov theorem the GLS estimator is the unique minimum variance estimator among linear unbiased estimators [see Amemiya (1972)].

It is not possible to specify the elements of the variance–covariance matrix V_t a priori. The task of estimating these elements is greatly simplified by assuming that the Sharpe single index model is a correct description of the return generating process. The process that generates returns at the beginning of period t is assumed to be as follows:

$$\tilde{R}_{it} = \alpha_{it} + \beta_{it} \tilde{R}_{mt} + \tilde{\varepsilon}_{it}, \qquad i = 1, 2, \ldots, N_t, \tag{34}$$

$$\begin{aligned} \operatorname{cov}(\tilde{\varepsilon}_{it}, \tilde{\varepsilon}_{jt}) &= 0, & i \neq j, \\ &= s_{ii}, & i = j, \end{aligned} \tag{35}$$

$$\alpha_{it} = E(\tilde{R}_{it} \mid R_{mt} = 0).$$

With this specification the element in the ith row and the jth column of V_t, written as $V_t(i,j)$, is given by

$$\begin{aligned} V_t(i,j) &= \beta_{it}\beta_{jt}\sigma_{mm}, & i \neq j, \\ &= \beta_{it}^2 \sigma_{mm} + s_{ii}, & i = j, \end{aligned} \qquad i, j = 1, 2, \ldots, N_t, \tag{36}$$

[3] For a similar interpretation, see Rosenberg and Marathe (1978).

where

$$\sigma_{mm} \equiv \operatorname{var}(\tilde{R}_{mt}).$$

Under these conditions the GLS estimator of Γ obtained by using data in period t reduces to

$$\hat{\Gamma}_t = (X_t' \Omega_t^{-1} X_t)^{-1} X_t' \Omega_t^{-1} R_t, \tag{37}$$

where Ω_t is a diagonal matrix of order $(N_t \times N_t)$, whose element in the ith row and jth column is given by

$$\begin{aligned} \Omega_t(i,j) &= 0, & i \neq j, \\ &= s_{ii}, & i = j, \end{aligned} \qquad i, j = 1, 2, \ldots, N_t. \tag{38}$$

In appendix A it is shown that this estimator is the GLS estimator for Γ. That is, under the assumptions of the single index model, the estimator minimizes the 'residual risk' of three portfolio returns, subject to the constraint that the expected returns on these portfolios are γ_0, γ_1 and γ_2 respectively. This estimator can be constructed as a heteroscedastic transformation on $\tilde{R}_t$ and X_t. Define the matrix P_t of order $(N_t \times N_t)$ whose elements are given by

$$\begin{aligned} P_t(i,j) &= \phi / s_i \equiv \phi / \sqrt{s_{ii}}, & i = j \\ &= 0, & i \neq j, \end{aligned} \tag{39}$$

where ϕ is a positive scalar. Then $\hat{\Gamma}_t$ can also be arrived at from the OLS regression on the transformed variables,

$$\tilde{R}_t^* = X_t^* \Gamma + \tilde{\varepsilon}_t^*, \tag{40}$$

where

$$\tilde{R}_t^* = P\tilde{R}_t \quad \text{and} \quad X_t^* = PX_t.$$

This is equivalent to deflating the variables in the ith rows of $\tilde{R}_t$ and X_t by a factor proportional to the residual standard error s_i. Note that Black and Scholes (1974), who used the portfolio approach, assumed in addition to the single index model that the 'residual' risks of all securities were equal; that is, they assumed that $s_{ii} = s^2$ for all i. Therefore, the Black–Scholes estimator reduces to OLS on the untransformed variables.

Errors in variables. Since true population β_{it} variables are unobserved, estimates of this variable, $\tilde{\beta}_{it}$ are obtained from historical data. The estimated beta is assumed equal to the true beta plus a measurement error $\tilde{v}_{it}$,

$$\tilde{\beta}_{it} = \beta_{it} + \tilde{v}_{it}. \tag{41}$$

The presence of measurement error causes a misspecification in OLS and GLS estimators, and the resulting estimates of Γ are biased and inconsistent [see, for example, Johnston (1972), for a discussion of the bias in the coefficients of a variable without error, here dividend yield, see Fisher (1977)]. The estimates $\tilde{\beta}_{it}$ are obtained from a regression of $\tilde{R}_{i\tau}$ on the return of the market portfolio $\tilde{R}_{m\tau}$ from data prior to period t,

$$\tilde{R}_{i\tau} = \alpha_{it} + \beta_{it}\tilde{R}_{m\tau} + \tilde{e}_{i\tau}, \qquad \tau = t-60,\, t-59, \ldots, t-1. \tag{42}$$

Since the single index model is assumed, $\mathrm{cov}(\tilde{e}_{i\tau}, \tilde{e}_{j\tau}) = 0$ and hence $\mathrm{cov}(\tilde{v}_{it}, \tilde{v}_{jt}) = 0$. If the joint probability distribution between security rates of return and market return is stationary, the variance of the measurement error $\mathrm{var}(\tilde{v}_{it})$ is proportional to the variance of the residual risk term $\mathrm{var}(\tilde{e}_{i\tau})$, for each i. Since month t is not used in this time series regression, $\mathrm{cov}(\tilde{e}_{it}, \tilde{v}_{it}) = 0$. Note that this time series regression yields a measured beta, $\tilde{\beta}_{it}$, its variance $\mathrm{var}(\tilde{v}_{it})$ and the variance of the residual risk term $\mathrm{var}(\tilde{e}_{it}) = s_{ii}$.

Consistent with prior empirical studies, the assumption $E(\tilde{e}_{i\tau}) \doteq 0$ has been made. However, it is recognized that if the 'market return' used in (42) is not the true market return, then the estimate of β_{it} may be biased, as has been observed by Sharpe (1977), Mayers (1972) and Roll (1977).

Because of errors in variables, most previous empirical tests have grouped stocks into portfolios. Since errors in measurement in betas for different securities are less than perfectly correlated, grouping risky assets into portfolios would reduce the asymptotic bias in OLS estimators. However, grouping results in a reduction of efficiency caused by the loss of information. The efficiency of the OLS estimator of the coefficient of a single independent variable is proportional to the cross sectional variation in that independent variable (beta). For the two independent variables case (dividend yield and beta), Stehle (1976) has shown that the efficiency of the OLS estimator of the coefficient of a given independent variable, using grouped data, is proportional to the cross-sectional variation in that variable unexplained by the variation in the other independent variable. Since the within group variation in dividend yield unexplained by beta is eliminated, the efficiency of the estimate of the dividend yield coefficient using grouped data is lower than that using all the data.[4] For this reason the present study

[4] The variance of the OLS estimator of the second independent variable (dividend yield) is equal to the variance of the error term divided by the portion of its variation that is unexplained by the first independent variable (beta). Therefore, unless the independent variables are

does not use the grouping approach to errors in variables. Instead, use is made of the measurement error in beta to arrive at a consistent estimator for Γ.

In constructing the GLS estimator $\hat{\Gamma}_t$ in (37), each variable has been deflated by a factor proportional to the residual standard deviation. The factor of proportionality was an arbitrary positive scalar. The structure of our problem is such that the standard error of measurement in $\tilde{\beta}_{it}$, $\sigma_i = (\mathrm{var}(\tilde{v}_{it}))^{\frac{1}{2}}$, is proportional to the standard deviation of residual risk, $s_i = (\mathrm{var}(\tilde{e}_{i\tau}))^{\frac{1}{2}}$. That is, if the time series regression model satisfies the OLS assumptions,

$$\sigma_i = s_i \Big/ \left(\sum_{\tau=t-60}^{t-1} (\tilde{R}_{m\tau} - \bar{R}_m)^2 \right)^{\frac{1}{2}}, \tag{43}$$

where $\bar{R}_m$ is the sample mean of the market return in the prior 60 month period.[5] Assume that σ_i is known and let

$$\phi = s_i / \sigma_i, \tag{44}$$

in the definition of P in (39). Thus each variable in the rows of $\tilde{R}_t$ and X_t is now deflated by the standard deviation of the measurement error in β_{it}. If $\tilde{\beta}_{it}$ is used in place of β_{it} (unobserved), the measurement error in the deflated independent variable, $\tilde{\beta}^* = \tilde{\beta}_{it}/\sigma_i$ will now have unit variance.

Call the matrix of regressors used $\tilde{X}_t^*$, which is simply X_t^* with $\tilde{\beta}_{it}$ replacing β_{it}. Then

$$\tilde{X}_t^* = X_t^* + \begin{bmatrix} 0 & \tilde{v}_{1t}/\sigma_1 & 0 \\ 0 & \tilde{v}_{2t}/\sigma_2 & 0 \\ \vdots & \vdots & \vdots \\ 0 & \tilde{v}_{N_t t}/\sigma_{N_t} & 0 \end{bmatrix}, \tag{45}$$

where $\mathrm{var}(\tilde{v}_{it}/\sigma_i) = 1$. Then the computed overall estimator

uncorrelated sequential grouping procedures as used by Black and Scholes (1974) are inefficient relative to grouping procedures that maximize the between group variation in dividend yield that is unexplained by the between group variation in beta.

[5] In the actual estimation, risk premiums were used. That is, $\tilde{R}_{i\tau} - r_{f\tau}$ was regressed on $\tilde{R}_{m\tau} - r_{f\tau}$ to estimate β_{it} as explained in section 4 below. Thus in the computation in (43), $(\tilde{R}_{m\tau} - r_{f\tau} - \overline{R_m - r_f})^2$ is used in place of $(\tilde{R}_{m\tau} - \bar{R}_m)^2$.

$$\hat{\Gamma} = \sum_{t=1}^{T} (\hat{\Gamma}_t/T), \tag{46}$$

where

$$\hat{\Gamma}_t = (\tilde{X}_t^{*\prime} \tilde{X}_t^*)^{-1} \tilde{X}_t^{*\prime} \tilde{R}_t^* \tag{47}$$

is inconsistent. This is because

$$\operatorname*{plim}_{N_t} \hat{\Gamma}_t = \left(\Sigma_{X_t^* X_t^*} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right)^{-1} \frac{X_t^{*\prime} \tilde{R}_t}{N_t}, \tag{48}$$

where

$$\Sigma_{X_t^* X_t^*} = \operatorname*{plim}_{N_t} \frac{X_t^{*\prime} X_t^*}{N_t}.$$

This says that each cross sectional estimator is biased even in large samples. Hence the overall estimator, being an arithmetic mean of the cross-sectional estimators, is inconsistent.

Consider the following estimator in each cross sectional month:

$$\hat{\Gamma}_t = \left(\frac{\tilde{X}_t^{*\prime} \tilde{X}_t^*}{N_t} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right)^{-1} \frac{\tilde{X}_t^{*\prime} R_t^*}{N_t}. \tag{49}$$

Then

$$\operatorname*{plim}_{N_t} \hat{\Gamma} = \frac{X_t^{*\prime} \tilde{R}_t}{X_t^{*\prime} X_t^*}, \tag{50}$$

and

$$E\left(\operatorname*{plim}_{N_t} \hat{\Gamma}_t \right) = \frac{X_t^{*\prime} E(\tilde{R}_t^*)}{X_t^{*\prime} X_t^*} = \Gamma. \tag{51}$$

Thus each cross-sectional estimator is unbiased, in large samples, for Γ.

Note that a portfolio interpretation can also be given to (47). Since

$$\operatorname*{plim}_{N_t} \left(\frac{\tilde{X}_t^{*\prime} \tilde{X}_t^*}{N_t} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right)^{-1} \frac{\tilde{X}_t^{*\prime} X_t^*}{N_t} = I, \tag{52}$$

it follows that the estimator for γ_0 in (47) is the realized excess return on a normal portfolio that has, in probability limit, a zero beta and a dividend yield equal to the riskless rate. Similarly the estimator for γ_1 (or γ_2) is the realized excess return on a hedge portfolio that has, in probability limit, a beta of one (or zero) and a dividend yield equal to zero (or unity).

The overall estimator,

$$\hat{\Gamma} = \sum_{t=1}^{T} (\hat{\Gamma}_t/T), \tag{53}$$

combines T independent estimates, and is consistent,

$$\operatorname*{plim}_{T} \left[\operatorname*{plim}_{N_t} \sum_{t=1}^{T} (\hat{\Gamma}_t/T) \right] = \Gamma. \tag{54}$$

It is shown in appendix B that, if $\tilde{v}_{it}$ and $\tilde{\varepsilon}_{it}$ are jointly normal and independent, then $\hat{\Gamma}_t$ is the maximum likelihood estimator (MLE) for Γ, using data in period t.

4. Data and results

Data on security rates of return (R_{it}) were obtained from the monthly return tapes supplied by the Center for Research in Security Prices (CRSP) at the University of Chicago. The same service provides the monthly return on a value weighted index of all the securities on the tape, and this index was used as the market return (R_{mt}) for the time series regressions. From January 1931 until December 1951, the monthly return on high grade commercial paper was used as the return on the riskless asset (r_{ft}); from January 1952 until December 1977 the return on a Treasury Bill (with one month to maturity) was used for r_{ft}. Estimates of each security's beta, $\hat{\beta}_{it}$, and its associated standard error were obtained from regressions of the security excess return on the market excess return for 60 months prior to t,

$$\tilde{R}_{i\tau} - r_{f\tau} = \alpha_{it} + \beta_{it}(\tilde{R}_{m\tau} - r_{f\tau}) + \tilde{\varepsilon}_{i\tau}, \qquad \tau = t-60, t-59, \ldots, t-1. \tag{55}$$

This was repeated for all securities on the CRSP tapes from $t=1$ (January 1936) to $t = T = 504$ (December 1977). January 1936 was chosen as the initial month for (subsequent) cross-sectional regressions because that was when dividends first became taxable.

To conduct the cross-sectional regression, the dividend yield variable (d_{it}) was computed from the CRSP monthly master file. This is

$$d_{it} = 0,$$

if in month t, security i did not go ex-dividend; or if it did, it was a non-recurring dividend not announced prior to month t;

$$d_{it} = D_{it}/P_{it-1},$$

if in month t, security i went ex-dividend, and the dollar taxable dividend D_{it} per share was announced prior to month t; and

$$d_{it} = \hat{D}_{it}/P_{it-1},$$

if in month t security i went ex-dividend and this was a recurring dividend not previously announced. Here $\hat{D}_{it}$ was the previous (going back at most 12 months), recurring, taxable dividend per share, adjusted for any changes in the number of shares outstanding in the interim; where P_{it-1} is the closing price in month $t-1$.

This construction assumes that the investor knows at the end of each month whether or not the subsequent month is an ex-dividend month for a recurring dividend. However, the surrogate for the dividend is based only on information that would have been available ex ante to the investor.

The cross-sectional regressions in each month provide a sequence of estimates $\{(\hat{\gamma}_{0t}, \hat{\gamma}_{1t}, \hat{\gamma}_{2t}),\ t = 1, 2, \ldots, 504\}$. Three such sequences are available: the first uses OLS, the second uses GLS and the third uses maximum likelihood estimation. The econometric procedures developed in section 3 apply equally well to the single variable regression, excess returns on beta alone. This corresponds to a test of the two factor Capital Asset Pricing Model, as in Black, Jensen and Scholes (1972) and Fama and MacBeth (1973),

$$\tilde{R}_{it} - r_{ft} = \gamma_0' + \gamma_1' \beta_{it} + \tilde{u}_{it}, \qquad i = 1, 2, \ldots, N_t, \quad t = 1, 2, \ldots, 504, \qquad (56)$$

where $\tilde{u}_{it}$ is the deviation of $\tilde{R}_{it}$ from its expected value. These cross sectional regressions provide three sequences $\{(\hat{\gamma}_{0t}', \hat{\gamma}_{1t}'),\ t = 1, 2, \ldots, 504\}$, the first using OLS, the second using GLS and the third using maximum likelihood estimation.

The estimated coefficients were shown to be realized excess rates of return on portfolios (with certain characteristics)[6] in month t. It is assumed that the excess rates of return on these portfolios are stationary and serially uncorrelated. Under these conditions the most efficient estimators of the

[6] See section 3, and also appendix A.

expected excess return on these portfolios would be the unweighted means of the monthly realized excess returns. The sample variance of the mean is computed as the time series sample variance of the respective portfolio returns divided by the number of months,

$$\hat{\gamma}_k = \sum_{t=1}^{504} \hat{\gamma}_{kt} \Big/ 504, \qquad k = 0, 1, 2, \qquad (57)$$

$$\operatorname{var}(\hat{\gamma}_k) = \sum_{t=1}^{504} (\hat{\gamma}_{kt} - \hat{\gamma}_k)^2 / (504 \cdot 503). \qquad (58)$$

A similar computation is made for $\hat{\gamma}_0'$ and $\hat{\gamma}_1'$.

The three sets of estimators of γ_0, γ_1 and γ_2 (and of γ_0' and γ_1') and their respective t-statistics for the overall period January 1936 to December 1977 are provided in Panel A (Panel B) of table 1.

Table 1

Pooled time series and cross section estimates of the after-tax and the before-tax CAPM: 1936–1977.[a]

	Panel A: After-tax model			Panel B: Before-tax model	
Procedure	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_0'$	$\hat{\gamma}_1'$
OLS	0.00616 (4.37)	0.00268 (1.51)	0.227 (6.33)	0.00681 (4.84)	0.00228 (1.26)
GLS	0.00446 (3.53)	0.00344 (1.87)	0.234 (8.24)	0.00516 (4.09)	0.00302 (1.63)
MLE	0.00363 (2.63)	0.00421 (1.86)	0.236 (8.62)	0.00443 (3.22)	0.00369 (1.62)

[a] *Notes*: The after-tax version corresponds to the regression

$$\tilde{R}_{it} - r_{ft} = \gamma_0 + \gamma_1 \beta_{it} + \gamma_2 (d_{it} - r_{ft}) + \tilde{\varepsilon}_{it}, \quad i = 1, 2, \ldots, N_t, \ t = 1, 2, \ldots, T.$$

The before-tax version corresponds to the regression

$$\tilde{R}_{it} - r_{ft} = \gamma_0' + \gamma_1' \beta_{it} + \tilde{\mu}_{it}, \quad i = 1, 2, \ldots, N_t, \ t = 1, 2, \ldots, T.$$

Each regression above is performed across securities in a given month. This gives estimates $\{\hat{\gamma}_{0t}, \hat{\gamma}_{1t}, \hat{\gamma}_{2t};\ t = 1, 2, \ldots, T\}$ and $\{\hat{\gamma}_{0t}', \hat{\gamma}_{1t}';\ t = 1, 2, \ldots, T\}$. The reported coefficients are arithmetic averages of this time series: for example,

$$\hat{\gamma}_1 = \sum_{t=1}^{T} \frac{\hat{\gamma}_{1t}}{T},$$

where $T = 504$, t-statistics are in parentheses under each coefficient, and they refer to $t(\hat{\gamma}_j)$, where $j = 1, 2, 3$.

110138-OPC-POD-60-200

The OLS and GLS estimators are biased and inconsistent due to measurement error in beta. The maximum likelihood estimators are consistent: consistency is a large sample property and for this study the monthly cross sectional regressions have between 600 and 1200 firms, and there were 504 months.[7] In Panel A, table 1, the MLE estimator of γ_1 is about 60 percent greater than the corresponding GLS estimator. Consistent with prior studies, the MLE estimator of γ_1 is significantly positive, indicating that investors are risk averse. Also consistent with prior studies, the MLE estimator of γ_0 is significantly positive. In Panel B, tests of the two factor model are presented. Note that in both panels, the GLS procedure results in an increase in the efficiency of the estimator of γ_1, which is $\hat{\gamma}_1$ ($\hat{\gamma}'_1$) in Panel A (Panel B). Consistent with prior tests of the traditional version of the Capital Asset Pricing Model, the null hypothesis that $\gamma'_0 = 0$ is rejected. Consistent with investor risk aversion $\hat{\gamma}'_1$ is significantly positive at the 0.1 level. Explanations for a positive intercept ($\gamma_0 > 0$) include, in addition to margin constraints on borrowing, misspecification of the market porfolio [see Mayers (1972), Sharpe (1977) and Roll (1977)], or beta serving as a surrogate for systematic skewness [see Kraus and Litzenberger (1976)].

The coefficient of the excess dividend yield variable, $\hat{\gamma}_2$, (Panel A) is highly significant under all the estimating procedures. The standard errors of the GLS and maximum likelihood estimators of γ_2 are about 25 percent smaller than that of the OLS estimator. The magnitude of the coefficient indicates that for every dollar of taxable return investors require between 23 and 24 cents of additional before tax return.

While the finding of a significant dividend coefficient contrasts with the Black–Scholes (1974) finding of an insignificant dividend effect, the magnitude of the coefficient in table 1 is consistent with their study. The dividend yield (independent) variable they used was $(d_i - d_m)/d_m$, where d_m was the average dividend yield on stocks. Since the coefficient they found was 0.0009, and the average annual yield in their period of study (1936–1966) was 0.048, their estimate of γ_2 can be approximated by 0.0009/(0.048/12), or 0.225.

It has been assumed that the variance of the estimator of Γ is constant over time. If, due to the quarterly patterns in the incidence of dividend payments, the variances of the estimators are not constant, the equally weighted estimators in (50) are inefficient relative to an estimator that accounts for any seasonal pattern in the variance. Since dividends are usually paid once every quarter, it is possible to compute three independent estimates of Γ by averaging the coefficients obtained in only the first, only the second and only the third month of each quarter. These three estimates of Γ may be weighted by the inverse of their variances to obtain a more efficient estimator. This is provided in table 2. As can be seen from this table,

[7] Consistency here is with respect to the overall estimator so one takes probability limits with respect to t and with respect to N_t. See section 3.

the overall estimator for γ_2 is very close to the MLE estimate in table 1. The estimate of the standard error of $\hat{\gamma}_2$ is approximately the same for the first two months, but about 30 percent less for the third month.

Table 2

Pooled time series and cross section estimates of the after-tax CAPM: 1936–1977 (based on quarterly dividend patterns).[a]

Month of quarter	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
First	0.00748 (0.00234)	0.00770 (0.00379)	0.28932 (0.05418)
Second	0.00212 (0.00232)	0.00071 (0.00335)	0.23531 (0.05034)
Third	0.00134 (0.00248)	0.00399 (0.00453)	0.18940 (0.03534)
Overall estimate	0.00373 (0.00137)	0.00383 (0.00219)	0.22335 (0.02552)

[a] *Notes*: The after-tax version corresponds to the regression

$$\tilde{R}_{it} - r_{ft} = \gamma_0 + \gamma_1 \beta_{it} + \gamma_2 (d_{it} - r_{ft}), \quad i = 1, 2, \ldots, N_t.$$

This regression is performed across securities in a given month t. Maximum likelihood estimation is used. The reported coefficients are arithmetic averages of the coefficients obtained over time (see note to table 1). The first three rows use the estimates from only the first, only the second and only the third months of each quarter. There are 168 months' estimates in each row. Standard errors are in parentheses under each coefficient. The 'overall estimates' use the estimates in each row above, weighted inversely by their variances.

It may be inappropriate to treat γ_2 as an intertemporal constant: in the absence of income related constraints on borrowing, γ_2 is a weighted average of individuals' marginal tax rates, which may have changed over time. Assume that investors have utility functions that display decreasing absolute risk aversion and non-decreasing relative risk aversion. Assume in addition that the distribution of wealth is independent of individual utility functions. Under these conditions the weight of the marginal tax rates of individuals in the higher tax brackets would be greater than that of individuals in lower tax brackets. Holland (1962) has shown that from 1936 to 1960 there was no pronounced upward trend in the marginal tax rates of individuals with taxable income in excess of $25,000. To examine empirically whether there is evidence of an upward trend in γ_2 over time, the maximum likelihood results are presented for six subperiods in table 3. The estimators of γ_2 for the subperiods were consistently positive and, except for the 1/1955 to 12/1961 period, significantly different from zero. There does not appear to be a trend to the estimate.

Table 3

Pooled time series and cross section estimates of the after-tax CAPM (for 6 subperiods).*

Period	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
1/36–12/40	−0.00287 (−0.52)	0.00728 (0.65)	0.335 (2.64)
1/41–12/47	0.00454 (1.44)	0.00703 (1.59)	0.408 (7.35)
1/48–12/54	0.00528 (2.77)	0.00617 (1.45)	0.158 (4.37)
1/55–12/61	0.01355 (5.62)	−0.00316 (−0.78)	0.018 (0.32)
1/62–12/68	−0.00164 (−0.47)	0.01063 (1.95)	0.171 (2.33)
1/69–12/77	0.00166 (0.47)	−0.00045 (−0.09)	0.329 (6.00)

*Notes: The after-tax version corresponds to the regression

$$\tilde{R}_{it} - r_{ft} = \gamma_0 + \gamma_1\beta_{it} + \gamma_2(d_{it} - r_{ft}) + \tilde{\varepsilon}_{it}, \quad i = 1, 2, \ldots, N_t, \; t = 1, 2, \ldots, T.$$

Maximum likelihood estimation is used for the cross sectional regression. The reported coefficients are arithmetic averages of the coefficients estimated in the months in the period (see note to table 1). t-statistics are in parentheses under each coefficient.

It is possible that the positive coefficient on dividend yield is not a tax effect and that in non-ex-dividend months the effect completely reverses itself. If dividends are paid quarterly there would be twice as many non-ex-dividend months as ex-dividend months. Thus, a complete reversal would require a negative effect on returns in each non-ex-dividend month that is half the absolute size of the effect in an ex-dividend month. It is also possible that a stock's dividend yield is a proxy for the covariance of its return with classes of assets not included in the value weighted index of NYSE stocks used to calculate betas in the present study. If the coefficient on dividend yield is entirely due to the effects of omitted assets, the effect in non-ex-dividend months should be positive and the same size as the effect in ex-dividend months.

In order to test whether there is a reversal effect or a re-inforcing effect in non-ex-dividend months the following cross-sectional regression was estimated:

$$\tilde{R}_{it} - r_{ft} = \gamma_0 + \gamma_1\tilde{\beta}_{it} + \gamma_2\{\delta_{it}d^0_{it} - r_{ft}\} + \gamma_3\{(1-\delta_{it})d^0_{it}\} + \tilde{\varepsilon}_{it}, \quad i = 1, 2, \ldots, N_t \qquad (59)$$

where

$$d^0_{it} = D_{it}/P_{it-1},$$

if a dividend was announced prior to month t, to go ex-dividend in month t;

$$d^0_{it} = \hat{D}_{it}/P_{it-1}$$

otherwise; and

$$\delta_{it} = 1,$$

if month t was an ex-dividend month for a recurring dividend;

$$\delta_{it} = 0,$$

otherwise.

The variable $(1-\delta_{it})d^0_{it}$ is intended to pick up the effect of a dividend payment in subsequent, non-ex-dividend months. The variable $\delta_{it}d^0_{it}$ is identical to d_{it}, the variable used earlier. If dividends are paid quarterly, and γ_3 is negative and has an absolute value half the size of γ_2, then one can conclude that there is a complete reversal over the course of the quarter so that there is no net tax effect. On the other hand, if there is no reversal, γ_3 should not be significantly negative.

The MLE estimates of the coefficients in (52) are presented in table 4. The estimated value of $\hat{\gamma}_3$ is positive and significantly different from zero: this rejects the hypothesis that there is complete reversal.

The significant positive γ_3 is evidence of a re-inforcing effect in non-ex-dividend months. If the coefficient on dividend yield is entirely attributable

Table 4

Pooled time series and cross section test of the reversal effect of dividend yield: 1936–1977.*

$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
0.00184 (1.29)	0.00493 (2.17)	0.32784 (7.31)	0.10321 (2.87)

*Notes: The regression performed in each month is

$$\tilde{R}_{it} - r_{ft} = \gamma_0 \gamma_1\beta_{it} + \gamma_2\{\delta_{it}d^0_{it} - r_{ft}\} + \gamma_3(1-\delta_{it})d^0_{it} + \tilde{\varepsilon}_{it}, \quad i = 1, 2, \ldots, N_t, \; t = 1, 2, \ldots, T.$$

Maximum likelihood estimation is used for the cross-sectional regression. The reported coefficients are arithemetic averages of the coefficients in each month (see note to table 1). t-statistics are in parentheses under each coefficient.

to the effect of omitted assets γ_3' should be the same order of magnitude as γ_2'. If the effect in ex-dividend months exceeds the combined effect in the subsequent two non-ex-dividend months γ_2 should be more than twice as large as γ_3. $\hat{\gamma}_2 - 2\hat{\gamma}_3$ is 0.1214 and has a t-value of 2.79. Thus, the effect in an ex-dividend month is more than twice the size of the effect in a non-ex-dividend month. This evidence suggests that the coefficient on dividend yield in ex-dividend months is not solely attributable to the effects of missing assets and that the effect in an ex-dividend month exceeds the combined effect in the subsequent two non-ex-dividend months. If the effect in non-ex-dividend months is asserted to be entirely due to the effect of missing assets, the difference $\hat{\gamma}_2 - \hat{\gamma}_3 = 0.225$ is an estimate of the tax effect. However, further theoretical work on the combined effects of transaction costs and personal taxes in a multi-period valuation framework is required to be able to understand the cause of a significant yield effect in non-ex-dividend months. For the present it seems reasonable to conclude that 0.225 is a lower bound estimate of the tax effect.[8]

The empirical evidence presented by Elton and Gruber (1970) on the ex-dividend behavior of common stocks suggests that the coefficient on the excess dividend yield term may be a decreasing function of yield. The theoretical rationale for this effect is that investors in low (high) tax brackets invest in high (low) dividend yield stocks: a possible explanation is that institutional restrictions on short sales results in a segmentation of security holdings according to investors' tax brackets. To provide a simple test of this 'clientele' effect, the coefficient c in (22) is hypothesized to be a linear decreasing function of the ith security's dividend yield. That is c, which is now dependent on i, is written c_i and given by

$$c_i = k - hd_i, \tag{60}$$

where $k, h > 0$, and the hypothesized relationship is

$$E(\tilde{R}_i) - r_f = a + b\beta_i + (k - hd_i)(d_i - r_f). \tag{61}$$

The econometric model is

[8]It might be argued that the persistent dividend effect is due to the fact that the dividend variable used incorporates knowledge of the ex-dividend month, which the investor may not have. To test whether this introduces spurious correlations between yields and returns the variable $(d_m^e/3)$ was used in the cross-sectional regression (23). The variable does not incorporate knowledge of the ex-dividend month except when it was announced. It is divided by 3 so as to distribute the yield over the three months of every quarter. The overall estimate (1936–1977) of $\hat{\gamma}_2$ is 0.39, with a t-value of 3.57: one cannot attribute the earlier results due to knowledge of ex-dividend months. This is consistent with the Rosenberg and Marathe (1978) study. Note that this estimate is lower than the total effect in table 4, which is $\hat{\gamma}_2 + 2\hat{\gamma}_3 = 0.52$. The lower estimate is attributable to constraining the coefficient on yield to be the same in non-ex-dividend months and ex-dividend months.

$$\begin{aligned}\tilde{R}_{it} - r_{ft} = {} & \gamma_0 + \gamma_1\beta_{it} + \gamma_2(d_{it} - r_{ft}) \\ & + \gamma_4 d_{it}(d_{it} - r_{ft}) + \tilde{\varepsilon}_{it}, \qquad i = 1, 2, \ldots, N_t,\end{aligned} \tag{62}$$

where the estimate of k is γ_2 and that for $-h$ is γ_4. The maximum likelihood approach is used in each cross sectional regression, and the pooled estimates presented in table 5.

Table 5

Pooled time series and cross section test of the clientele effect: 1936–1977.[a]

$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_4$
0.00365 (2.65)	0.00425 (1.88)	0.336 (6.60)	−6.92 (−1.70)

[a]*Notes*: This corresponds to the following cross-sectional regression in each month:

$$\tilde{R}_{it} - r_{ft} = \gamma_0 + \gamma_1\beta_{it} + \gamma_2(d_{it} - r_{ft}) + \gamma_4 d_{it}(d_{it} - r_{ft}) + \tilde{\varepsilon}_{it} \qquad \begin{array}{l} i = 1, 2, \ldots, N_t \\ t = 1, 2, \ldots, T. \end{array}$$

Maximum likelihood estimation is used for the cross-sectional regression. The reported coefficients are arithmetic averages of the coefficients in each month (see note in table 1). t-statistics are in parentheses under each coefficient.

Consistent with the existence of a clientele effect, the maximum likelihood estimate of γ_2 is significantly positive and that of γ_4 is significantly negative, both at the 0.05 level. The magnitude of $\hat{\gamma}_4$ suggests that for every percentage point in yield the implied tax rate for ex-dividend months declines by 0.069. For example, if the annual yield was 4 percent, the implied tax rate would be approximately 0.336–6.92 (0.04/4) = 0.268, assuming quarterly payments. The empirical evidence supporting a clientele effect suggests the need for further research that rigorously derives an equilibrium model that incorporates institutional restrictions on short sales, along with personal taxes.

5. Conclusion

In this paper, an after-tax version of the Capital Asset Pricing Model is derived. The model extends the Brennan after-tax version of the CAPM to incorporate wealth and income related constraints on borrowing along with a progressive tax scheme. The wealth related constraint on borrowing causes the expected return on a zero-beta portfolio (having a dividend yield equal to the riskless rate) to exceed the riskless rate of interest. The income related constraint tends to offset the effect that personal taxes have on the equilibrium structure of share prices. The equilibrium relationship indicates that the before tax expected return on a security is linearly related to its systematic risk and to its dividend yield. Unrestricted supply adjustments in corporate dividends would result in the before tax version of the CAPM, in a world where dividends and interest are taxed as ordinary income. If income related constraints are non-binding and/or corporate supply adjustments are restricted, the before tax return on a security would be an increasing linear function of its dividend yield.

Unlike prior tests of the CAPM that used grouping or instrumental variables to correct for measurement error in beta, this paper uses the sample estimate of the variance of observed betas to arrive at maximum likelihood estimates of the coefficients in the relations tested. Unlike prior studies of the effect of dividend yields on asset prices, which used average monthly yields as a surrogate for the expected yield in both ex-dividend and non-ex-dividend months, the expected dividend yield based on prior information is used for ex-dividend months and is set to zero for other months.

The results indicate that there is a strong positive relationship between before tax expected returns and dividend yields of common stocks. The coefficient of the dividend yield variable was positive, less than unity, and significantly different from zero. The data indicates that for every dollar increase in return in the form of dividends, investors require an additional 23 cents in before tax return. There was no noticeable trend in the coefficient over time. A test was constructed to determine whether the effect of dividend yield reverses itself in non-ex-dividend months, and this hypothesis was rejected. Indeed, the data indicates that the effect of a dividend payment on before tax expected returns is positive in both the ex-dividend month and in the subsequent non-ex-dividend months. However, the combined effect in the subsequent non-ex-dividend months is significantly less than the effect in the ex-dividend month.

Evidence is also presented for a clientele effect: that is, that stockholders in higher tax brackets choose stocks with low yields, and vice versa. Further work is needed to derive a model that implies the existence of such clienteles and to test its implications.

Appendix A

In this appendix it is shown that the estimator for Γ, given by

$$\hat{\Gamma}_t = (X_t' \Omega_t^{-1} X_t)^{-1} X_t' \Omega_t^{-1} \tilde{R}_t,$$

using data in period t, is the Generalized Least Squares (GLS) estimator for Γ under the assumption of the single index model. It was shown in section 3 of the paper that each estimated coefficient corresponds to the realized excess

return of a specific portfolio. Suppose portfolio weights $\{h_{it}; i=1,2,\ldots,N_t\}$ are chosen in each period, for investment in assets $i=1,2,\ldots,N_t$. Using eq. (23) from the text the excess return on such a portfolio is given by

$$\sum_i h_{it}(\tilde{R}_{it}-r_{ft}) = \gamma_0\left(\sum_i h_{it}\right) + \gamma_1\left(\sum_i h_{it}\beta_{it}\right) + \gamma_2\left[\sum_i h_{it}(d_{it}-r_{ft})\right] + \sum_i h_{it}\tilde{\varepsilon}_{it}.$$

The expected excess return on this portfolio is

$$\begin{aligned}
&\gamma_0 \quad \text{if} \quad \sum_i h_{it}=1, \quad \sum_i h_{it}\beta_{it}=0, \quad \sum_i h_{it}(d_{it}-r_{ft})=0,\\
&\gamma_1 \quad \text{if} \quad \sum_i h_{it}=0, \quad \sum_i h_{it}\beta_{it}=1, \quad \sum_i h_{it}(d_{it}-r_{ft})=0,\\
&\gamma_2 \quad \text{if} \quad \sum_i h_{it}=0, \quad \sum_i h_{it}\beta_{it}=0, \quad \sum_i h_{it}(d_{it}-r_{ft})=1.
\end{aligned}$$

Under the assumption of the single index model, the variance of the return on such a portfolio is, from eq. (36) in the text,

$$\operatorname{var}\left(\sum_i h_{it}(\tilde{R}_{it}-r_{ft})\right) = \left(\sum_i h_{it}\beta_{it}\right)^2 \sigma_{mm} + \sum_i h_{it}^2 s_{it}.$$

Suppose one wishes to minimize the variance of the excess return on such a portfolio subject to the condition that the expected excess return on the portfolio is, in turn, γ_0, γ_1 or γ_2. This condition enforces $\sum_i h_{it}\beta_{it}$ to be either zero or unity. Hence minimizing

$$\left(\sum_i h_{it}\beta_{it}\right)^2 \sigma_{mm} + \sum_i h_{it}^2 s_{it},$$

subject to the unbiasedness condition, is equivalent to minimizing

$$\sum_i h_{it}^2 s_{it},$$

the 'residual risk' of the portfolio subject to the unbiasedness condition. Thus, one is using the residual risk of the portfolio as the minimand and enforcing the unbiasedness condition. By construction, Ω_t is the diagonal matrix of the residual variances s_{it}, and by construction, $\hat{\Gamma}_t$ is linear and unbiased for Γ. The variance of the estimator has been minimized under the single index model. But by the Gauss–Markov theorem, the GLS estimator [using the full matrix V_t in (36) as the variance–covariance matrix] is the unique minimum variance estimator among linear and unbiased estimators. Hence $\hat{\Gamma}_t$ is the GLS estimator for Γ, under the assumption of the single index model.

Appendix B

In this section, it is shown that under certain conditions, $\hat{\hat{\Gamma}}_t$ in (49) is the maximum likelihood estimator for Γ in period t.

First, note that there are no errors in the measurement of β, then if security returns are multivariate normal, then the GLS estimator in (37) is also the maximum likelihood estimator [see Johnston (1972)].

Suppose now there are errors in the measurement of β. Then one can use the transformation P defined in (39), with $\phi = s_i/\sigma_i$, to write the model as

$$\tilde{R}^*_{it} = \gamma_0 p^*_{it} + \gamma_1 \beta^*_{it} + \gamma_2 d^*_{it} + \tilde{\varepsilon}^*_{it}, \tag{B.1}$$

and the observed beta as

$$\tilde{\beta}^*_{it} = \beta^*_{it} + \tilde{v}^*_{it}, \tag{B.2}$$

where

$$\tilde{R}^*_{it} = (\tilde{R}_{it} - r_{ft})/\sigma_i, \quad p^*_{it} = 1/\sigma_i, \quad \beta^*_{it} = \beta_{it}/\sigma_i,$$
$$\tilde{\beta}^*_{it} = \tilde{\beta}_{it}/\sigma_i, \quad d^*_{it} = (d_{it} - r_{ft})/\sigma_i, \quad \tilde{\varepsilon}^*_{it} = \tilde{\varepsilon}_{it}/\sigma_i,$$

and

$$\tilde{v}^*_{it} = \tilde{v}_{it}/\sigma_i.$$

Define the variable

$$m_{xy} = \sum_{i=1}^{N_t} x_{it} y_{it} \Big/ N_t, \tag{B.3}$$

as the raw co-moment for a given sequence $\{(x_{it}, y_{it}),\ i = 1, 2, \ldots, N_t\}$. Then from (B.1) and (B.2),

$$m_{\tilde{R}^* p^*} = \gamma_0 m_{p^* p^*} + \gamma_1 m_{\beta^* p^*} + \gamma_2 m_{d^* p^*} + m_{\tilde{\varepsilon}^* p^*}, \tag{B.4}$$

$$\begin{aligned} m_{\tilde{R}^* \tilde{\beta}^*} = {} & \gamma_0 [m_{p^* \beta^*} + m_{p^* \tilde{v}^*}] + \gamma_1 [m_{\beta^* \beta^*} + m_{\tilde{v}^* \beta^*}] \\ & + \gamma_2 [m_{d^* \beta^*} + m_{d^* \tilde{v}^*}] + m_{\tilde{\varepsilon}^* \beta^*} + m_{\tilde{\varepsilon}^* \tilde{v}^*}, \end{aligned} \tag{B.5}$$

$$m_{\tilde{R}^* d^*} = \gamma_0 m_{p^* d^*} + \gamma_1 m_{\beta^* d^*} + \gamma_2 m_{d^* d^*} + m_{\tilde{\varepsilon}^* d^*}, \tag{B.6}$$

$$m_{\tilde{\beta}^* p^*} = m_{\beta^* p^*} + m_{\tilde{v}^* p^*}, \tag{B.7}$$

$$m_{\tilde{\beta}^* \tilde{\beta}^*} = m_{\beta^* \beta^*} + 2 m_{\beta^* \tilde{v}^*} + m_{\tilde{v}^* \tilde{v}^*}, \tag{B.8}$$

$$m_{d^* \tilde{\beta}^*} = m_{d^* \beta^*} + m_{d^* \tilde{v}^*}. \tag{B.9}$$

In these six equations, take expectations and use the fact that

$$\begin{aligned} & E(\tilde{v}^*_{it}) = E(\tilde{\varepsilon}^*_{it}) = 0, \\ & E(\tilde{v}^*_{it} \tilde{\varepsilon}^*_{it}) = 0, \\ & E(\tilde{v}^*_{it} \tilde{v}^*_{it}) = E[\tilde{v}^2_{it}/\sigma^2_i] = 1. \end{aligned} \tag{B.10}$$

The left-hand side of each of (B.4) through (B.9), after taking expectations, corresponds to the population co-moments of the subscripted variables.

If $\tilde{v}_{it}$ and $\tilde{\varepsilon}_{it}$ are independently normally distributed, then the corresponding sample moment is a maximum likelihood estimator of the population parameter. Replace these expected values by their maximum likelihood estimates. There are now six equations for the six unknown parameters γ_0, γ_1, γ_2, $m_{\beta^* \beta^*}$, $m_{\beta^* d^*}$, and $m_{\beta^* p^*}$. They can be solved for the coefficients of interest from the following 'normal' equations, which are in terms of observed sample estimates.

$$m_{\tilde{R}^* p^*} = \gamma_0 m_{p^* p^*} + \gamma_1 m_{\tilde{\beta}^* p^*} + \gamma_2 m_{d^* p^*}, \tag{B.11}$$

$$m_{\tilde{R}^* \tilde{\beta}^*} = \gamma_0 m_{p^* \tilde{\beta}^*} + \gamma_1 (m_{\tilde{\beta}^* \tilde{\beta}^*} - 1) + \gamma_2 m_{d^* \tilde{\beta}^*}, \tag{B.12}$$

$$m_{\tilde{R}^* d^*} = \gamma_0 m_{p^* d^*} + \gamma_1 m_{\tilde{\beta}^* d^*} + \gamma_2 m_{d^* d^*}, \tag{B.13}$$

and are themselves maximum likelihood [see Mood et al. (1974, p. 285)].

The solution to this set gives estimates $\hat{\hat{\gamma}}_{kt}$, $k = 0, 1, 2$, which are embodied in (49). They are functions of maximum likelihood estimates. Note that in addition to (B.4) through (B.9), one could write an equation for $m_{\tilde{R}^* \tilde{R}^*}$:

$$\begin{aligned} m_{\tilde{R}^* \tilde{R}^*} = {} & \gamma_0^2 m_{p^* p^*} + \gamma_1^2 m_{\beta^* \beta^*} + \gamma_2^2 m_{d^* d^*} + 2\gamma_0\gamma_1 m_{p^* \beta^*} \\ & + 2\gamma_0\gamma_2 m_{p^* d^*} + 2\gamma_1\gamma_2 m_{\beta^* d^*} + 2\gamma_0 m_{p^* \tilde{\varepsilon}^*} + 2\gamma_1 m_{\beta^* \tilde{\varepsilon}^*} \\ & + 2\gamma_2 m_{d^* \tilde{\varepsilon}^*} + m_{\tilde{\varepsilon}^* \tilde{\varepsilon}^*}. \end{aligned} \tag{B.14}$$

If we take expectations, using (B.10) and the fact that

$$E(m_{\varepsilon^*\varepsilon^*})=E\left(\sum_{i=1}^{N_t}\frac{\tilde{\varepsilon}_{it}^2}{\sigma_i^2 N_t}\right)$$

$$=\frac{1}{N_t}\sum_{i=1}^{N_t}\frac{E(\tilde{\varepsilon}_{it}^2)}{\sigma_i^2}=\frac{1}{N_t}\cdot N_t\phi^2=\phi^2,$$

we have

$$E(m_{g^*g^*})=\gamma_0^2 m_{p^*p^*}+\gamma_1^2 m_{\beta^*\beta^*}+\gamma_2^2 m_{d^*d^*}+2\gamma_0\gamma_1 m_{p^*\beta^*}$$
$$+2\gamma_0\gamma_2 m_{p^*d^*}+2\gamma_1\gamma_2 m_{\beta^*d^*}+\phi^2, \qquad \text{(B.15)}$$

where ϕ^2 is assumed known.

By writing down the likelihood function and maximizing it for an analogous case, Johnston (1963) demonstrates a maximum likelihood estimator over the parameter space ($\gamma_0, \gamma_1, \gamma_2, \beta_{it}$ for $i=1,2,\ldots,N_t,\phi$). This has the undesirable characteristic that the parameter space grows with the sample size.[9] It turns out in our problem that ϕ is assumed known. If this ϕ satisfies (B.15), when in (B.15) we use the sample co-moment estimates for the population parameters, then Johnston's M.L. procedure coincides with the solution to (B.11) through (B.13). Whereas our estimators are linear in the returns and can be interpreted as portfolios, the expanded parameter space estimator in Johnston is non-linear and has no such analog to theory. Thus conditional on ϕ^2 coinciding with the residual variation in the sample, using our estimates, the estimator in (49) is a maximum likelihood estimator over the parameter space ($\gamma_0, \gamma_1, \gamma_2$).

[9] See Kendall and Stuart (1973, especially pp. 62 and 402).

References

Amemiya, T., 1972, Theory of econometrics: Lecture notes, Unpublished manuscript (Department of Economics, Stanford University, Stanford, CA).

Bailey, M.J., 1969, Capital gains and income taxation, in: A.C. Harberger and M.J. Bailey, The taxation of income from capital (Brookings Institution, Washington, DC) 11–49.

Black, F., 1972, Capital market equilibrium with restricted borrowing, Journal of Business 45, 444–454.

Black, F., M. Jensen and M. Scholes, 1972, The capital asset pricing model: Some empirical tests, in: M. Jensen, ed., Studies in the theory of capital markets (Praeger, New York) 79–121.

Black, F. and M. Scholes, 1974, The effects of dividend yield and dividend policy on common stock prices and returns, Journal of Financial Economics 1, 1–22.

Blume, M. and I. Friend, 1973, A new look at the capital asset pricing model, Journal of Finance 28, 19–33.

Brennan, M.J., 1973, Taxes, market valuation and corporate financial policy, National Tax Journal 23, 417–427.

Brennan, M.J., 1970, Investor taxes, market equilibrium and corporation finance, Unpublished Ph.D. Dissertation (Massachusetts Institute of Technology, Cambridge, MA).

Cottle, S., D.L. Dodd and B. Graham, 1962, Security analysis: Principles and techniques (McGraw-Hill, New York).

Elton, E. and Gruber, 1970, Marginal stockholder tax rates and the clientele effect, Review of Economics and Statistics 52, 68–74.

Fama, E.F. and J.D. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, Journal of Political Economy 71, 607–636.

Fama, E.F. and M.H. Miller, 1972, The theory of finance (Holt, Rinehart and Winston, New York).

Fisher, F.M., 1977, The effect of simple specification error on the coefficients of 'unaffected' variables, Working Paper no. 194 (Department of Economics, Massachusetts Institute of Technology, Cambridge, MA).

Friend, I. and M. Blume, 1970, Measurement of portfolio performance under uncertainty, American Economic Review, 561–575.

Gonzalez-Gaverra, N.G., 1973, Inflation and capital asset market prices: Theory and tests, Unpublished Ph.D. Dissertation (Graduate School of Business, Stanford University, Stanford, CA).

Gordon, M.J., 1963, Optimal investment and financing policy, Journal of Finance 18, 264–272.

Holland, D.M., 1962, Dividends under the income tax (NBER, Princeton, NJ).

Johnston, J., 1963, Econometric methods (McGraw-Hill, New York).

Johnston, J., 1972, Econometric methods (McGraw-Hill, New York).

Kendall, M.G. and A. Stuart, 1973, The advanced theory of statistics (Hafner, New York).

Kraus, A. and R.H. Litzenberger, 1976, Skewness preference and the valuation of risk assets, Journal of Finance 31, 1085–1100.

Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, Review of Economics and Statistics 47, 13–37.

Litzenberger, R.H. and J.C. Van Horne, 1978, Elimination of the double taxation of dividends and corporate financial policy, Journal of Finance 34, 737–749.

Long, J., 1975, Efficient portfolio choice with differential taxation of dividends and capital gains, Journal of Financial Economics 5, 25–53.

Mayers, D., 1972, Non-market assets and capital market equilibrium under uncertainty, in: M.C. Jensen, ed., Studies in the theory of capital markets (Praeger, New York).

Miller, M. and M. Modigliani, 1961, Dividend policy growth, and the valuation of shares, Journal of Business 4, 411–433.

Miller, M. and M. Scholes, 1972, Rates of return in relation to risk: A re-examination of some recent findings, in: M. Jensen, ed., Studies in the theory of capital markets (Praeger, New York).

Miller, M. and M. Scholes, 1977, Dividends and taxes, Working Paper no. 8 (University of Chicago, Chicago, IL).

Mood, A., F.S. Graybill and D.C. Boes, 1974, Introduction to the theory of statistics (McGraw-Hill, New York).

Roll, R., 1977, A critique of the asset pricing theory's tests, Journal of Financial Economics 4, 129–176.

Rosenberg, B. and V. Marathe, 1978, Test of capital asset pricing hypotheses, Journal of Financial Research, forthcoming.

Rubinstein, M., 1973, A comparative statics analysis of risk premiums, Journal of Business 46.

Sharpe, W.F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance 19, 425–442.

Sharpe, W.F., 1977, The capital asset pricing model: A multi-beta interpretation, in: H. Levy and M. Sarnat, eds., Financial decision making under uncertainty (Academic Press, New York).

Stehle, R.D., 1976, The valuation of risk assets in an international capital market: Theory and tests, Unpublished Ph.D. Dissertation (Graduate School of Business, Stanford, CA).

Vasicek, O., 1971, Capital market equilibrium with no riskless borrowing, Wells Fargo Bank Memorandum (San Francisco, CA).

The Effects of New Equity Sales Upon Utility Share Prices

By RICHARD H. PETTWAY*

Public knowledge of a forthcoming sale of new equity by a utility company often precipitates a decline in the market price of that equity and continues to impact share prices after the sale has taken place. Such price changes are part of the real cost of selling the new issue. The market pressure costs of new equity capital have been the subject of much speculation in utility rate cases, but have received little detailed study. The author of this article has made such a study and here presents a quantitative analysis of price-return movements encountered by utility stocks in the market, after first defining market pressure as it applies particularly to the regulated utility environment. He concludes that investors clearly view a new sale of equity shares with disfavor and regulators, as well as company managements, should be concerned with the resultant decline in utility stock prices.

When a public utility decides to sell a new issue of equity capital and publicly discloses this information, share prices are thought to decline. Often these selling firms ask for an adjustment to their costs of equity capital for the effects of this market pressure upon share prices. The subsequent argument and debate about the magnitude of an adjustment for market pressure at rate hearings is well known.

The electric utility industry has been one of the largest issuers of new equity shares during the past twenty-five years. Therefore, it is surprising that there has not been much more research to determine the magnitude of market pressure of these numerous new equity sales in this industry. The objective of this article is to report on the results of an analysis of 368 equity sales by 73 different electric utilities from January 1, 1973, through December 31, 1980. The analysis will measure two effects of new common equity sales upon share prices: market pressure and sales effect. Specifically, this article will determine the magnitude of market pressure defined as the effect of the sale upon share prices which reduces the funds received by the issuing company at the sale date, and will determine the size of the sales effect defined as the total effect of the sale upon share prices from before the announcement until after the sale.

*The research underlying this article was partly funded by a grant from the Public Utility Research Center, University of Florida.

Richard H. Pettway is a professor of finance, in the Graduate School of Business at the University of Florida. For the past ten years he has been associated with the Public Utility Research Center at the University of Florida. He has written books, monographs, and articles and has made appearances as an expert witness before public utility commissions specializing in the financial and economic problems and solutions. **Dr. Pettway** received his BBA, MBA, and PhD degrees in finance and statistics from the University of Texas-Austin.

There have been studies into the size of market pressure defined as a temporary price decline in share values when a large block of shares is said to be "overhanging" the market. However, most of this research concentrates upon the price effects of new issues of industrial companies sold in the primary markets or of large blocks of existing stock sold in the secondary market [1, 2, 4, 5, 6, 9].** This literature defines market pressure as the amount of recovery in market prices *after* the issue has been sold. A review of this literature indicates either no market pressure existing in large block trades of outstanding shares, or only a small amount of pressure associated with primary market sales of new issues.

Under utility regulation, the concern is with a different definition of market pressure. Market pressure in the public utility industry is generally defined as the decline in prices *while* the issue is still overhanging, *before* it is sold. The main question is how much did the utility's stock decline in the secondary market associated with the sales announcement to the date of sale. This decline is a real cost of selling the new issue as the firm will receive only the reduced price at the sales date. An

**Numbers in brackets refer to the list of references at the end of the article.

110138-OPC-POD-60-207

article by Bowyer and Yawitz (BY) [3] measured the decline in share prices between the announcement date and the sales date of 278 new equity issues of public utilities from 1973 through 1976. But that research had some obvious problems which are corrected by this study.

The first problem with BY is their definition of the announcement date (AD). They defined this critical AD as the initial Securities and Exchange Commission filing date of the issue prospectus. This may not be the true AD as often public utilities make prior announcements of their new issues to state public service commissions, to investors in the *Irving Trust Calendar*, to underwriters, or to financial analysts much earlier than the SEC filing date. This study redefines the critical announcement date through a detailed questionnaire survey of electric utility companies. Further, an analysis of price changes prior to the established announcement date for each issue will be made to determine the actual impact of new equity sales upon share prices. It is very important to measure the complete decline in market prices associated with the information about the forthcoming sale of new equity shares.

Another problem with the BY study concerns its authors' use of the Dow-Jones utility index to measure differential declines in share prices and returns. The use of this index is flawed for at least four reasons. First, the number of companies included is small, 15 firms, and only 11 are electric companies; whereas four are gas transmission and distribution companies. The inclusion of the gas companies raises serious questions concerning the similarities of risks between electric utilities tested and the companies which make up their comparison index. Second, their index does not capture the dividend portion of the return and thus only measures the changes in prices without adjusting for dividends paid. In the electric power industry, the dividend yields tend to be a high portion of the total return and the omission of dividends could impart a bias to the index. Third, if there is evidence of market pressure in new sales of equity shares by utilities as BY found, then it is certain that this market pressure is contained also in share prices of Dow-Jones utility index firms when they sold new equity shares. The effect of using an index which contains market pressure to measure the size of market pressure of a particular firm which sold new equity naturally will understate the true amount of market pressure which is present. Fourth, if utilities are impacted differently from unregulated firms, there may be an additional "industrial effect" which will not be observed by looking only at other utilities rather than a broadly based comparison index of share prices and returns.

Finally, there are some technical problems with the way that BY measured the decline in stock returns or market pressure. These problems concern the use of average residual returns versus a more correct measure (geometric residual returns) and the way BY handled underwriting costs.

Data

A questionnaire survey was conducted of the 93 New York Stock Exchange-listed, investor-owned electric utilities from which 73 usable company replies were obtained for a response rate of over 78 per cent. Each company provided all identifiable costs and critical dates for each new equity capital sale made by the firm from January 1, 1973, through December 31, 1980. The survey results contain data on 368 actual equity sales over the eight-year survey period. The data represent more than five new equity sales per company on average over the study period. The size of these equity sales ranged from \$4.7 million to \$198 million with a mode sale value in a range between \$30 and \$49.9 million per issue. The frequency of the issues over the eight years of the survey shows that 1975 was the most popular year followed by 1976 and 1980. Yet, the individual year variation was not dramatic as the range over the eight years was from a low of 37 issues in 1974 to a high of 64 issues in 1975. Eighty-two per cent of the sales were through negotiated underwriting, 16 per cent through competitive bidding, and 2 per cent through rights offerings. See [7] for a thorough review of the data and details on the flotation costs of these issues.

Data on realized share returns including dividends for each company were obtained on a daily basis for a period which began sixty-five trading days before the announcement date and ended thirty trading days after the sale date (SD). Thus, company returns were obtained from a fixed period prior to the AD through a fixed period after the SD for each issue. It is best to think of these data sets as 368 separate arrays of returns. Because the interim time period between the AD and the subsequent SD varied for each issue, the number of return observations in each array is different. Each collected array of returns is unique to the particular announcement and issue dates and is not impacted by other equity sales of the same company.

Methodology

In order to control for risk, to adjust for movements in general prices and returns, and to reduce estimating bias, a two-stage regression process was used to measure the effects of new equity sales upon share returns and prices. First, during the estimating period, the market regression model (1) was applied to a firm's daily equity returns over a uniform estimating period which began sixty-five trading days prior to the AD and ended fifteen days before the AD for each issue. The market regression model asserts that:

$$\tilde{R}_{i,t} = \hat{a}_i + \hat{B}_i\tilde{R}_{m,t} + \tilde{e}_{i,t} \qquad (1)$$

where $\tilde{R}_{i,t}$ is the daily return including dividends of the issuing company for equity issue i — i.e., one to 368 — at time t; where daily returns of the issuing company concerning issue i are defined as $(P_{i,t} + D_{i,t} - P_{i,t-1}) / (P_{i,t-1})$; P is the price and D is the dividend per share; $\tilde{R}_{m,t}$ is the daily return at time t on a market portfolio for comparison; $\hat{a}_i$ and $\hat{B}_i$ are the estimated parameters of the market model; and $\tilde{e}_{i,t}$ is the error term of the model.

110138-OPC-POD-60-208

In order to make comparisons, an electric utility portfolio index of returns was created over the period January 1, 1973, through December 31, 1980, containing an equal investment in each of 73 electric companies which sold equity during the period. It is a daily returns index including dividends and provides the average return for each day on a portfolio consisting of an equal dollar investment in each of the 73 electric utilities.

Thus, the first stage uses an estimating period of fifty trading days, approximately two and one-half months, to determine the parameters of the market regression model. The second stage then applies these estimated parameters to the returns series during the subsequent test period after the estimating period in each array in order to calculate the expected returns for each company on each issue i using:

$$\hat{R}_{i,t} = \hat{a}_i + \hat{B}_i \hat{R}_{m,t} \tag{2}$$

where $\hat{R}_{i,t}$ is the expected return for the issuing company associated with issue i at time t. Then residual returns during the test period are obtained by comparing the actual versus the predicted returns using:

$$\tilde{R}_{i,t} - \hat{R}_{i,t} = \tilde{u}_{i,t} \tag{3}$$

where $\tilde{u}_{i,t}$ is the daily residual return of the issuing company for issue i at time t.

In order to display these residual returns properly, a decision must be made of how to combine the individual company residuals centered on a common date during the test period. The method of combining residuals used by Bowyer and Yawitz is called cumulative average residual or CAR. This method would find the average residual return of all issues on a specific day relative to the common AD or SD and would accumulate these averages over the period in an additive way. A different way of combining residual returns, average geometric residual return (AGRR), was chosen for this study. It is a theoretically better measure of residual returns over time than CAR. AGRR does not use the average residual returns on a specific date but takes the individual issue residual ($\tilde{u}_{i,t}$) from (3) and converts it into a price relative for each t and then forms a geometric return series by multiplying successive price relatives from fourteen days prior to AD to the end of the residual data for each company using formula (4). Thus, a geometric return series which precisely measures the change in investment worth for each individual issue is created. At any point in time relative to the common dates, AD and SD, the AGRR was determined as the numeric average of the geometric returns up to that point in time of all issues using formula (5).

$$GRR_{i,T} = \prod_{t=1}^{T} (1 + u_{i,t}) \tag{4}$$

$$AGRR_T = \sum_{i=1}^{N} GRR_{i,T}/N \tag{5}$$

where i is the issue number, t is time, T is the specific point in time (T=1, 2, 3, . . . total number of observations in the test period which was from fourteen days before the AD until thirty trading days after the SD), and N is the number of issues. For further details concerning the specifics of the methodology employed see [8].

In observing the pattern of these residuals over the test period, it is important to be able to use common definitions to describe their movements. "Market pressure" is defined as the decline of share prices and average geometric residual returns from fourteen days before the AD until the SD. "Sales effect" is defined as the change in share prices and AGRRs from fourteen days before the AD until thirty trading days after the SD. This sales effect would be the net change over the entire test period from before the announcement until well after the sale.

Price-Return Movements

Because the number of days between the AD and the SD are not identical for each issue, arrays of residual returns had to be centered on two separate common dates. The first common date is the AD and then data are centered on the common SD. To begin measuring any price effects of these new equity sales, the study first observed movements in residual returns when the data are centered on the common AD.

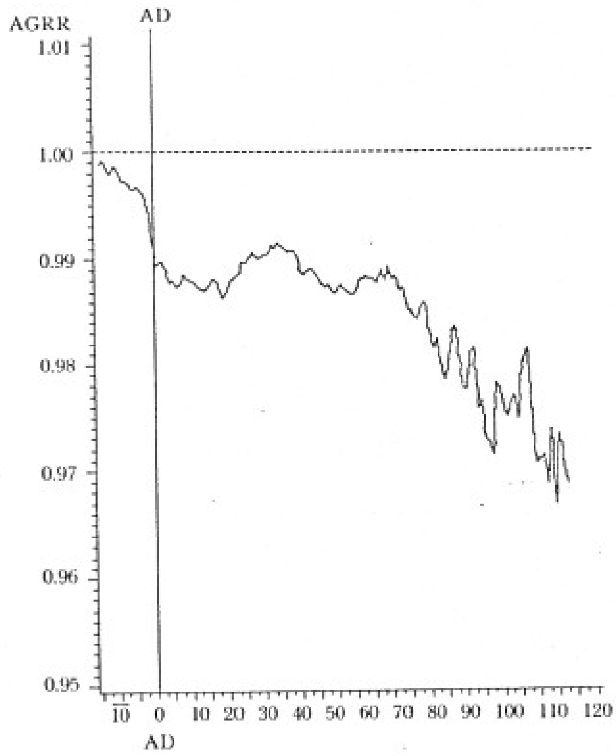
Common Announcement Date

Figure 1 illustrates the AGRRs derived from the use of the electric utility market index of returns for comparison.† The derived residuals are accumulated for 128 days starting fourteen days before the announcement date. All issues are centered on the AD. The trend of the AGRRs are clearly downward and below one during the entire span of 128 days. The downward trend is most noticeable immediately before and around the AD and is then followed by a period of relative stability. During this initial decline, share prices had fallen between one per cent and 1.4 per cent. The downward trend resumes again beginning about sixty-seven days after the AD. The latter downward trend may be associated with the SD, but since these data are centered on the AD, the SD did not occur at a common point in time in the data. Further, because SD is not a common point in the data, the amount of market pressure cannot be measured from the data in this format.

Panel 1 of the accompanying table contains statistical summaries of changes in AGRRs over the entire period shown in Figure 1. It is clear from the data that the change over the 128-day period centered on the AD was a negative 3.019 per cent, indicating a sales effect of this

†If there were no effects of new equity sales upon electric utilities which sold new shares, then the AGRRs shown on Figure 1 would be very close to one over time. A detrimental effect and a relative decline in share prices would be represented as a decline in AGRRs below one. A favorable effect would be represented as an increase in AGRRs. Also notice that the x-axis displays time with negative numbers as days before the AD and positive numbers as days after the AD. The AD, or centering date, is designated as zero.

Figure 1
AGRR Centered on Announcement Date
(Utility Index)



magnitude. Thus, comparing the returns over the same time period of an electric utility which sold new equity shares with returns of a portfolio of electric companies which also sold equity during the eight-year study period, there appears to have been a substantial and significant decline or sales effect of −3 per cent. There appear to be two periods of rapid declines, one just before and around the AD and another which appears to begin about sixty-seven days after the AD. Measuring the initial decline during a period from fourteen days before the AD to fourteen days after the AD, the specific decline was −1.2 per cent. This first major decline which begins before the AD suggests that the market was either anticipating the new equity sale or obtaining information about the new equity sale just prior to the public announcement.

Because of the decline in these residuals, it is clear that the market considered the potential new equity sale as detrimental to the future prospects of the current equity holders of the selling firm. Since the decline begins before the AD, this article measures more precisely the *total* decline in share prices than did the work of Bowyer and Yawitz.

Common Sales Date

Figure 2 shows the AGRRs using the electric utility returns index for comparison with all issues centered on the SD. This plot is clearly one whose trend is also downward across the entire time period, although it appears not to begin its major decline until eighty-five to ninety days prior to the SD.

In Panel 2 of the table are found the summary statistics describing the magnitudes of the AGRRs shown on Figure 2. The changes or sales effect during the period from fourteen days before the AD to after the SD over 147 days was −2.041 per cent.

Panel 3 of the table contains the magnitudes of AGRRs shown on Figure 2 but stopping at the SD. This decline in relative share prices and returns, called market pressure, is caused by the equity sale and is the discount required to sell the new issue. These costs of new equity issues were 1.893 per cent on average. Thus, market prices of shares of electric utilities which sold new equity declined by about 1.9 per cent from before the AD until the SD over 104 days. This is the decline in price that the firm did not receive when it sold new equity shares at the SD and is the market pressure of the new equity issue.

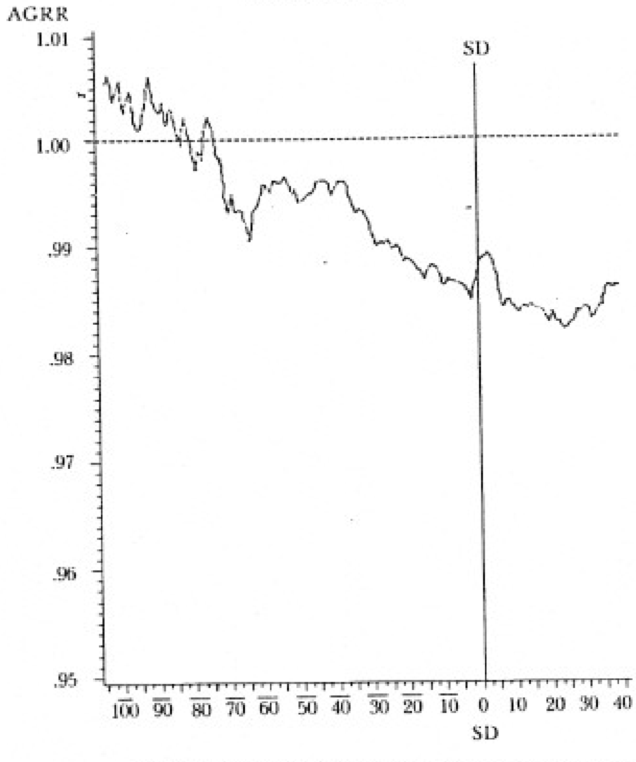
Effects of New Equity Sales of Utilities upon Share Prices
Changes in the Average Geometric Residual Returns

368 New Equity Issues of 73 Electric Utilities from
January 1, 1973, through December 31, 1980

Using the Utility Index

Measurements	*Panel 1* Centered on AD (Sales Effect)	*Panel 2* Centered on SD (Sales Effect)	*Panel 3* Centered and Ending on SD (Market Pressure)
Change over the Period	−3.019%	−2.041%	−1.893%
Length of Period (Days)	128	147	104
Change from −14 AD to +14 AD	−1.170%		
Length of Period (Days)	29		

Figure 2
AGRR Centered on Sale Date
(Utility Index)



110138-OPC-POD-60-210

Summary and Conclusions

When electric utilities sold new equity shares between January 1, 1973, and December 31, 1980, the share prices of these companies were depressed downward because of the sale. This downward movement or market pressure measured from before the announcement date to the sales date of the new issue was −1.9 per cent when compared with returns of other electric utilities which sold new equity regularly. Further, a sales effect ranging from −3 per cent to −2 per cent was found over the period from before the announcement date until after the sales date depending upon whether the data were centered on the AD or on the SD.

These averages are conservative and the minimum estimated average declines as they were derived from using a return index of comparison (electric utility) which itself contains the effects of market pressure. Further, the use of another index of return for comparison which was composed of regulated and unregulated firms would substantially raise these average costs. (In fact, if the comparison were to be made against the return of all equities listed on the New York and American stock exchanges over the same time period, the average estimate for market pressure would rise to −3 per cent and the average estimates for sales effect would rise to −4.4 per cent centered on the AD to −3.6 per cent centered on the SD. See [8] for details.)

The sizeable sales effect over the entire period from before the announcement date to after the sales date using the portfolio of electric companies for comparison provides direct evidence that share prices of electric utilities which sell new equity continue to decline after the sale has taken place. This condition may be explained as the impact of other factors than market pressure alone upon share prices. Perhaps some of these factors are due to the investors' perceptions of increased dilution problems caused by regulatory lag and regulatory risk associated with these public utilities not being allowed a rate of return on new equity equal to the investors' required rate of return over the eight-year survey period.

Even though the exact causes are not known precisely, it is definitely clear that investors view the new sale of equity shares with disfavor and that the new equity sale results in a substantial decline in equity prices. Public utility regulators should be concerned with these impacts of new equity sales upon share prices and returns and attempt to make proper adjustments in the allowed rate of return to offset or eliminate these effects in the future.

References

1. "Market Pressure: The Sales of New Common Equity and Rate of Return Regulation," by Raymond Armknecht, Fred Grygiel, and Patrick Hess, "Proceedings of the Business and Economic Statistics Section of the American Statistical Association," 1974, pp. 80-91.

2. "The Effect of the Size of Public Offerings of Common Stocks upon Preoffering Stock Prices," by Lee Bodenhamer, unpublished dissertation, Harvard University, Graduate School of Business, May, 1968.

3. "The Effect of New Equity Issues on Utility Stock Prices," by John W. Bowyer, Jr., and Jess B. Yawitz, 105 PUBLIC UTILITIES FORTNIGHTLY 25, May 22, 1980.

4. "On the Existence, Measurement, and Economic Significance of Market Pressure in the Pricing of New Equity Shares," by Robert E. Evans, unpublished dissertation, University of Wisconsin-Madison, 1978.

5. "Price Impacts of Block Trading on the New York Stock Exchange," by Alan Kraus and Hans Stoll, *Journal of Finance*, June, 1972, pp. 569-588.

6. "New Issue Stock Price Behavior," by J. G. McDonald and A. K. Fisher, *Journal of Finance*, March, 1972, pp. 97-102.

7. "A Note on the Flotation Costs of New Equity Capital Issues of Electric Companies," by Richard H. Pettway, 109 PUBLIC UTILITIES FORTNIGHTLY 68, March 18, 1982.

8. "Impacts of New Equity Sales upon Electric Utility Share Prices," by Richard H. Pettway and Robert C. Radcliffe, working paper series, Public Utility Research Center, Graduate School of Business, University of Florida, Gainesville, Florida, May, 1983.

9. "The Market for Securities: Substitution versus Price Pressure and the Effects of Information on Share Prices," by Myron S. Scholes, *Journal of Business*, April, 1972, pp. 179-211.

Utilities Raise Their Capital Appropriations

The nation's investor-owned utilities appropriated $7.2 billion (seasonally adjusted) for new plant and equipment in the final quarter of 1983, up 25 per cent over the unusually low figure recorded in the third quarter, the Conference Board reported in April. Both the gas and electric utilities shared in this fourth-quarter gain. (Capital appropriations are authorizations to spend money in the future for new plant and equipment. Appropriations are the first step in the capital investment process, preceding the ordering of equipment, the letting of construction contracts, and finally the actual expenditures. Appropriations are considered to be a leading indicator for capital spending.)

Electric utility appropriations rose to $5.8 billion in the fourth quarter, their first quarterly increase since the third quarter of 1982. Cancellations of previously approved projects were widespread, however, amounting to $2.7 billion in the final quarter of 1983.

Gas utility appropriations climbed to $1.4 billion in the fourth quarter, a 68 per cent jump over the third quarter. It was the highest quarterly total recorded last year. For the full year, however, the gas utilities appropriated only $4.4 billion, down by a third from 1982, and canceled a record $1.3 billion worth of earlier-approved projects.

Actual capital spending by the investor-owned utilities fell to $8.3 billion in the fourth quarter, an 8 per cent dip from the third quarter. The electric utilities accounted for all of the fourth-quarter decline. For 1983 as a whole, the electric utilities spent a record $32.2 billion on new plant and equipment, up 3 per cent over 1982. Gas utility expenditures amounted to $3.5 billion in 1983, down 30 per cent from 1982.

Journal of Financial Economics 5 (1977) 273–307 © North-Holland Publishing Company

ALTERNATIVE METHODS FOR RAISING CAPITAL

Rights Versus Underwritten Offerings

Clifford W SMITH, Jr *

Graduate School of Management, University of Rochester, Rochester, NY 14627 USA

Received February 1977, revised version received January 1978

This paper provides an analysis of the choice of method for raising additional equity capital by listed firms Examination of expenses reported to the SEC indicates that rights offerings involve significantly lower costs, yet underwriters are employed in over 90 percent of the offerings The underwriting industry, finance textbooks, and corporate proxy statements offer several justifications for the use of underwriters However estimates of the magnitudes of these arguments indicate that they are insufficient to justify the additional costs of the use of underwriters The use of underwriters thus appears to be inconsistent with rational, wealth-maximizing behavior by the owners of the firm The paper concludes with an examination of alternate explanations of the observed choice of financing method

1. Introduction and summary

In this paper I examine an apparent paradox Based on a comparison of costs, simple finance theory suggests that listed firms should use rights offerings to raise additional equity capital, rather than employing underwriters Yet the majority of firms choose underwritten offerings, rather than rights offerings

In an underwritten offering, underwriters contract to purchase shares from the issuing firm at a price usually set within 24 hours of the offering, and then resell the shares to the public In a rights offering the shareholder receives a right from the firm giving him the option to purchase new shares for each share owned In section 2, I show that with the proper specification of the subscription price, the proceeds of a rights offering are identical to the proceeds of an underwritten offering

Not identical, however, are costs In section 3, I examine the out-of-pocket costs of underwritten and rights offerings reported to the Securities and Exchange

*I would like to thank the participants at the Public Utilities Economics and Finance Seminar, sponsored by AT & T at the Graduate School of Management, University of California, Los Angeles, and the participants at the Finance Workshop, Graduate School of Management, University of Rochester, especially M Jensen, J Long, J Maguire, W Mikkelson, T Miller, R Ruback, L Wakeman and J Warner This research is supported by the Managerial Economics Research Center, Graduate School of Management, University of Rochester

Commission for issues registered under the Securities Act of 1933 between January 1971 and December 1975. Rights offerings are significantly less expensive. I also examine additional out-of-pocket expenses associated with both types of offerings. These include extras (options sold to underwriters), unreported expenses such as employee compensation, and the costs of rights offerings imposed directly on the owners of the firm. With these costs considered, I find rights offerings still are less expensive than underwritten offerings.

It has been suggested that selling efforts by underwriters raise stock prices while rights offerings lower them. In section 4 I study price behavior around the date of the offering. I find no empirical support for the hypothesis that abnormal positive returns are associated with underwritten offerings. Moreover, underwriters appear to set the offer price below the market value of the stock by at least 0.5 percent. While stock prices fall when rights are issued, the fall equals the market value of the rights received by the shareholder. Examination of the total rate of return to shareholders around the offer date indicates no abnormal returns; thus the wealth of the firm's owners is not reduced by a rights offering.

Section 5 provides an examination of other benefits presumed to accrue from the use of underwriters. Finance texts, corporate proxy statements, and the underwriting industry itself claim the existence of advantages in timing, insurance, distribution of ownership and from future consulting advice. My estimates of the magnitudes of the costs and benefits associated with these arguments are not sufficient to outweigh the lower costs of rights offerings as a means of raising capital. I can find no differential legal liability associated with the use of rights offerings which might explain the observed use of underwriters. Furthermore, there is no apparent difference in the sets of firms employing the alternative methods which could attribute the reported cost differences to selection bias.

In section 6, I offer a two-part hypothesis which is consistent with the observed frequency of employment of underwriters, with their higher costs, by the majority of listed firms. First, since managers' and directors' interests are different from those of shareholders in general, their financing decisions are not always in the best interests of the owners; benefits flow to management from the use of underwriters although not to shareholders. Second, I hypothesize that the cost to shareholders of monitoring their directors and managers is greater than the cost imposed by the choice of the more expensive financing method.

In section 7 I briefly present my conclusions.

A detailed description of the institutional arrangements for rights offerings and underwritten offerings is not easily available; I have provided one in Appendix 1. The reader unfamiliar with this institutional material will find it valuable to read this appendix before the body of the paper.

Appendix 2 presents a Black–Scholes (1973) option pricing analysis of rights issues and underwriting contracts, given here since general equilibrium analyses of these contracts have not been published.

2. Comparison of proceeds from rights and underwritten offerings

In a firm commitment underwritten offering, the underwriting syndicate purchases the new shares from the firm at an agreed upon price, and offers the shares for sale to the public at the offer price If the shares cannot be sold at the offer price, the underwriting syndicate breaks and the shares are sold for whatever price they will bring The underwriters bear the risk associated with adverse price movements, the proceeds to the firm are guaranteed Of course the difference between the offer price and the proceeds to the firm are expected to compensate the underwriter for bearing this risk

In a rights offering, each shareholder receives one right for each share owned This right is an option issued by the firm to purchase new shares The right states the relevant terms of the option, specifying the number of rights required to purchase each new share, the subscription price for each new share, and the expiration date of the option Since issuing rights is costly, it is in the firm's interest to insure the success of the offering A lower subscription price for the rights provides this insurance, a lower subscription price raises the market value of the right and reduces the probability that at the expiration date of the rights offering the stock price will be below the subscription price There is a corresponding fall in the market value of the stock, but this fall is like a stock split It does not affect the wealth of the owners of the firm [1]

If the shareholder does not exercise his rights, or does not sell his rights to someone who will exercise the rights, his wealth is reduced by the market value of the rights Thus the firm can make the probability of failure of the rights offering arbitrarily small by setting the subscription price low enough

Thus, since rights offerings and underwritten offerings can be specified so that the amount of capital raised by each is essentially equivalent, the decision as to which method to employ depends on the costs, the firm should employ that method which has lower net costs

3. Out-of-pocket expenses of rights and underwritten issues

"Expenses involved in a preemptive common stock rights offering are significantly greater than expenses involved in a direct offering of common stock

[1]The adjustment for the 'split effect' of a rights offering can be calculated as follows The ex-rights price of the shares, P_x, equals the with-rights price, P_w, minus the value of the right, R

$$P_x = P_w - R.$$

Ignoring the 'option value' of the right, the market value of a right is the difference between the ex-rights price and the subscription price, P_s, divided by the number of rights required to purchase one share, n

$$R = (P_x - P_s)/n$$

Substituting the second expression into the first and simplifying yields

$$P_x = (nP_w + P_s)/(n+1)$$

to the public due to additional printing and mailing costs, expenses associated with the handling of rights and the processing of subscriptions, higher underwriters' commissions and the longer time required for the consummation of financing." [2]

3.1. Reported out-of-pocket expenses

To examine the out-of-pocket expenses referred to in the quotation above (from Commonwealth Edison's 1976 proxy statement) I obtained a tape from the Securities and Exchange Commission covering the reported costs of all issues registered under the Securities Act of 1933 between January, 1971 and December, 1975. The tape contains data covering the following costs: (1) compensation received by investment bankers for underwriting services, (2) legal fees, (3) accounting fees, (4) engineering fees, (5) trustee's fees, (6) listing fees, (7) printing and engraving expenses, (8) Securities and Exchange Commission registration fees, (9) Federal Revenue Stamps, and (10) state taxes.

To restrict my analysis to equity issues by listed firms, I established the following criteria for inclusion: (1) the offering is of common stock and contains no other classes of securities; (2) the company's stock is listed on the New York Stock Exchange, American Stock Exchange, or a regional stock exchange prior to the offering; and (3) any associated secondary distribution is less than 10 percent of the gross proceeds of the issue. Table 1 is based on the issues meeting these criteria.

The data summarized in table 1 contradict Commonwealth Edison's Proxy Statement. My information, consistent with findings of previous SEC studies,[3] indicates that costs are *highest* for underwritten public offerings, and *lowest* for pure rights offerings. Furthermore, the difference in costs is striking. For a $15 million issue, the reported cost difference between an underwritten public offering and a pure rights offering is 4.83 percent, or $720,000; and for a $100 million issue the cost difference is 3.82 percent, or $3,820,000.[4] Yet underwriters were employed in over 93 percent of the issues examined.

3.2. Extras

Systematic understatement of the costs of underwriting presented in table 1 occurs because extras are omitted. Extras refer to the warrants which are associated with some underwritten issues and are used as partial payment to the underwriter. The warrants are options which are usually convertible into the

[2]Commonwealth Edison Proxy Statement, 1976.

[3]See SEC (1940, 1941, 1944, 1949, 1951, 1957, 1970, 1974).

[4]One empirical regularity in the data presented in table 1 should be noted. To a first approximation, the differences in costs among financing methods are explained by the differences in underwriter compensation. Compare 'Other Expenses' for Underwriting and Rights with Standby Underwriting with 'Total Costs' for Rights.

Table I

Costs of flotation as a percentage of proceeds for 578 common stock issues registered under the Securities Act of 1933 during 1971–1975 The issues are subdivided by size of issue and method of financing underwriting, rights with standby underwriting, and pure rights offering[a]

	Underwriting				Rights with standby underwriting				Rights	
Size of issue ($ millions)	Number	Compensation as a percent of proceeds	Other expenses as a percent of proceeds	Total cost as a percent of proceeds	Number	Compensation as a percent of proceeds	Other expenses as a percent of proceeds	Total cost as a percent of proceeds	Number	Total cost as a percent of proceeds
Under 0 50	0	–	–	–	0	–	–	–	3	8 99
0 50 to 0 99	6	6 96	6 78	13 74	2	3 43	4 80	8 24	2	4 59
1 00 to 1 99	18	10 40	4 89	15 29	5	6 36	4 15	10 51	5	4 90
2 00 to 4 99	61	6 59	2 87	9 47	9	5 20	2 85	8 06	7	2 85
5 00 to 9 99	66	5 50	1 53	7 03	4	3 92	2 18	6 10	6	1 39
10 00 to 19 99	91	4 84	0 71	5 55	10	4 14	1 21	5 35	3	0 72
20 00 to 49 99	156	4 30	0 37	4 67	12	3 84	0 90	4 74	1	0 52
50 00 to 99 99	70	3 97	0 21	4 18	9	3 96	0 74	4 70	2	0 21
100 00 to 500 00	16	3 81	0 14	3 95	5	3 50	0 50	4 00	9	0 13
Total/Average	484	5 02	1 15	6 17	56	4 32	1 73	6 05	38	2 45

[a]Issues are included only if the company's stock was listed on the NYSE, AMEX or regional exchanges prior to the offering, any associated secondary distribution represents less than ten percent of the total proceeds of the issue, and the offering contains no other types of securities The costs reported are (1) compensation received by investment bankers for underwriting services rendered, (2) legal fees, (3) accounting fees, (4) engineering fees, (5) trustees' fees, (6) listing fees (7) printing and engraving expenses (8) Securities and Exchange Commission registration fees (9) Federal Revenue Stamps, and (10) state taxes

stock of the firm at prices ranging from well below to considerably above the offering price When the underwriters acquire these warrants at a price below their market value, this represents a form of compensation to the underwriter, and it is not included in table 1

Although extras have historically been most often associated with new issues, their use in the compensation of underwriters of seasoned firms is not unusual For the years 1971–1972, the SEC (1974) reported that of the 1,599 issues which were underwritten, 530, or 33 1 percent, included extras However, since extras were included primarily with the smaller offerings, the total dollar volume of issues with extra compensation was only 7 percent of the gross proceeds from all underwritten offerings

The average exercise price of the warrants granted as a percentage of the offering price was 11 72 percent A lower bound on the value of the option is the difference between the subscription price of the offering and the exercise price of the extras, here that is 88 28 percent of the subscription price [5] Since these warrants are typically purchased by the managing investment banker at a minimal price, usually one to ten cents, the options appear to be significantly underpriced The SEC also found that the average ratio of shares granted the underwriters through extras to the number of shares offered in the underwriting was 7 99 percent To assess the impact on the figures reported in table 1, assume that the value of the warrant is 80 percent of the offering price, that the underwriter pays 5 percent of the offering price for the extras, and that the ratio of warrants received as extras to shares offered through the underwriting is 0 07, then the compensation represented by the extras would be 4 95 percent of the total proceeds These numbers suggest that for the issues employing extras, the figures in table 1 understate the underwriters' compensation on the order of 50 to 100 percent

3 3 Unreported out-of-pocket expenses

Such items as the opportunity cost of the time of the firm's employees and postage expenses[6] are not included in the summary of costs reported in table 1 However, unreported employee expenses are unlikely to explain the deviations reported in table 1 For a $15 million issue, the $720,000 difference would not be explained if 20 employees with an average salary of $30 thousand worked

[5]This is a conservative estimate of the value Merton (1973) has demonstrated that the lower bound on the value of an option is the difference between the stock price and the discounted exercise price

[6]Although postage expenses are not reported to the SEC, estimates were obtained from summaries of expenses reported to the New York State Public Utilities Commission for a sample of firms For the sample, the maximum postage expense as a percentage of total proceeds was one-tenth of one percent Even if this were understated by a factor of ten, it would be of insufficient magnitude to explain even the smallest reported difference in costs Moreover, the marginal postage expense could be reduced to zero by mailing the rights with other required mailings, such as dividend checks or quarterly reports

full time on a rights offering for a year For a $300 million issue the difference in reported costs of underwriting versus a rights issue exceeds $11 million, it would require over 350 man-years to explain this difference

It should be noted that expenses allocated to raising capital do *not* reduce the tax liability of the firm [7] These expenses are deducted from the capital account without affecting the income statement Thus, the use of internal resources can lower the tax liability of the firm if it is more expensive for the Internal Revenue Service to monitor the allocation of internal resources between capital raising activities and other activities In the above examples, if the firm's marginal tax rate is 50 percent, and if they were able to deduct all their wages for tax purposes, the required number of man-years to explain the reported cost differential would be doubled

There are strong reasons to believe that table 1 also omits significant unreported costs of the issuing firm's employees' time for underwritten offerings There are important parameters (e g, the offering price and the fee structure) which must be negotiated between the underwriter and the representatives of the firm, these parameters have wealth implications for the owners of the firm as well as the underwriter Such negotiation can be lengthy and usually directly involves top management These unreported costs of underwriting must be significantly greater than the costs of setting a subscription price for a rights issue, since the subscription price has no wealth implications for the owners of the firm as long as it is low enough to ensure that the rights will be exercised

Moreover, with an underwritten issue the firm has the same tax incentives to substitute internal for external resources if it is more expensive for the IRS to monitor the allocation of costs of internally acquired resources to capital raising activities than of those which are externally acquired Thus, it is not clear that rights offerings employ fewer unreported internal resources than do underwritten offerings

3 4 *Costs imposed directly on shareholders*

If a shareholder chooses to sell his rights, he incurs transactions costs and tax liabilities These costs, although not borne by the firm, are relevant because they affect the wealth of the owners [8]

[7] If the firm sells bonds rather than stock, the costs of selling the issue can be amortized over the life of the issue In no case, however, may these costs be expensed either for tax or reporting purposes

[8] There is a limited benefit from issuing rights to the owners of the firm under Regulation T, the Federal Reserve regulation restricting margin credit For an owner who wishes to borrow to acquire additional stock, Reg T provides for the establishment of a 'Special Subscription Account' which lowers the effective margin requirement by permitting a customer to purchase on an installment basis a margin security acquired through the exercise of subscription rights expiring within 90 days Under this provision, 75 percent of the market value of the acquired stock can be borrowed initially Quarterly installments are required over a 12 month period to bring the position up to proper margin

To determine the impact of the selling costs, let us assume generally extreme values for the relevant parameters For small dollar transactions (less than $1,000), the brokerage fee can be as much as 10 percent And for rights, the bid-ask spread can be as high as 10 percent, this represents another selling cost If half the bid-ask spread is taken as an implicit selling cost the total cost can be as much as 15 percent of the value of the rights To make the figures comparable to those in table 1, calculate transactions costs as a fraction of the proceeds of the offering to the firm The 15 percent must be multiplied by the ratio of the value of the rights to the total proceeds For the offerings in the sample, this ratio was approximately 10 percent If all individuals sold their rights, transactions costs would be 1 50 percent of the proceeds, a figure less than the difference in transactions costs for any reported issue size [9] But rights offerings are generally 50 percent subscribed by existing shareholders who do not bear these transactions costs [10] Therefore this cost appears to be less than one percent

Selling rights also has tax consequences for the shareholder For tax purposes, the cost basis of the stock must be allocated between the stock and the rights when the rights are received based on the market values of the rights and stock at that time [11] The acquisition date of the rights for tax purposes is the date on which the stock issuing the rights is acquired If the stock has risen in value since it was acquired, a relevant cost of employing a rights offering is the difference between the shareholder tax liability incurred now and the present value of the taxes which would have been paid had the rights issue not occurred [12]

To determine the impact of this cost again postulate generally extreme values for the relevant parameters Assume (1) that the marginal tax rate for the average shareholder is 50 percent (note this would be an unattainably high rate if the capital gain were long term), (2) that in the absence of the rights offering the taxes could have been postponed forever (3) that the allocated cash basis for the rights is 50 percent of the current rights price (4) that the ratio of the value of the rights to the proceeds of the issue is 10 percent, and (5) that only 20 percent of the current stockholders subscribe to the rights offering In this

[9] Note that since the expenses associated with raising equity capital are not tax deductible, these figures are comparable without further adjustment

[10] Estimates vary but ballpark figures on how investors react [to rights offerings] are as follows 50% exercise their rights 40% sell out for cash, and 10% do nothing ['Vanishing Rights' (May 2, 1977) Barron's p 25]

[11] If the fair market value of the rights is less than fifteen percent of the fair market value of the stock, the shareholder can choose to set the basis of the rights at zero leaving unaffected the basis of the stock The shareholder might choose this alternative if the cost of the bookkeeping exceeded the present value of the tax saving or if he anticipated being in a higher tax bracket when his remaining holdings were sold

[12] See Bailey (1969) for a discussion of the effective rate of capital gains tax, discounted to reflect the liability deferral

case, the cost would be 2 percent of the capital raised by the firm This is less than any reported cost differential in table 1 [15]

One other argument involving shareholder-borne costs has been offered by Weston and Brigham (1975) They argue that in a rights offering some stockholders may neither exercise nor sell, and by allowing their rights to expire unexercised they incur a loss [16] However, if an oversubscription privilege is employed with the offering, current owners in the aggregate receive full market value for the shares sold Admittedly, the oversubscription privilege affects the distribution of wealth among the owners, but it does not impose costs on owners as a whole

4. Security price behavior associated with rights and underwritten offering

4 1 Rights offerings lower the stock price

"A rights offering, under market conditions then existing, could well have a long-term depressing effect on the market price of the stock " [17]

Given the investment policy of the firm, a rights offering *will* lower the price of the stock in both the short run and in the long run as AT&T s Proxy Statement suggests But this is irrelevant to the choice of financing methods because the drop in price is *not* a reduction in the wealth of the owners and thus cannot be considered a cost of a rights issue

The fall in the stock price when rights are issued can be illustrated by the following argument Rights give the shareholders the option to purchase new shares at less than market prices Other things equal, the total market value of the firm after a rights offering, V, will then be the previous value, V' plus the subscription payments, S

$$V = V' + S \tag{1}$$

The per share price before the offering is V'/n, where n is the number of old shares If m new shares are sold, the per share price after the offering, $(V'+S)/(n+m)$ must be less than the price per share before the offering [18]

[15] If taxes were important, firms would avoid rights offerings when share prices had risen However the evidence presented in table 2 shows that, on average, firms have had abnormal positive price changes during the 12 months before an offering

[16] Stockbrokers holding securities for safekeeping do not allow the warrants to expire unexercised If no instructions are received, the broker will sell the rights immediately before expiration

[17] American Telephone and Telegraph Co , Notice of 1976 Annual Meeting and Proxy Statement

[18] Also note that arbitrage profits must not be available When a stock trades ex rights, a right is issued for each share outstanding At the ex rights date, the expected change in the stock price must equal the expected value of the right, or profit opportunities would exist if the sum of the ex rights value of the stock plus the value of the right at the ex rights date were

The fall in the stock price on the ex rights day is similar to the expected fall in the stock price at the ex dividend date. The two cases differ only in what is distributed – in the latter instance cash, in the former rights. Thus, the fall in the stock price simply reflects the fact that the shareholders have been given a valuable asset, the right.

The argument that the fall in the stock price is a relevant cost of a rights offering also appears in two related forms: (1) if an underwriter is used, the firm can raise a greater amount of capital with the same number of shares; (2) a rights offering lowers the earnings per share of the firm.[19] Both statements are true but if the fall in the stock price equals the market value of the rights, then the impact of the additional shares issued through the rights offering is the same as that of a stock split and the wealth of the owners of the firm is unaffected.

To examine whether, after correcting for the expected normal fall in the stock price, there were also abnormal price changes,[20] I studied the 853 rights offerings on the CRSP master file between 1926 and 1975. Following Fama, Fisher, Jensen and Roll (1967), I estimated the regression,

$$R_{jt} = \alpha_j + \beta_j R_{mt} + \varepsilon_{jt}, \tag{2}$$

where R_{jt} is the return to security j in month t, adjusted for capital structure changes (including rights offerings) and R_{mt} is the return to the market portfolio in month t. I estimated (2) for each of the 853 offerings, using data from the CRSP monthly return file, excluding the 25 months around the date of the offering. Setting $t = 0$ for the month of the rights offering, I used the estimated α_j and β_j to calculate the ε_{jt} for each security for the 25 months around the offering. I then calculated the average residual over all firms for each month in the interval -12 to $+12$. The average residuals were then cumulated from month -12 to the event month. The results are presented in table 2 and figure 1.

In the months subsequent to 'event month minus two' the average residuals

systematically different from the value of the stock immediately before the ex rights date, then profits could be made by taking an appropriate position in the stock upon the announcement of the rights issue.

[19]'Thus, if the amendment [to remove the preemptive right from the corporate charter] is adopted, the company will be able to obtain the amount of capital needed through the issuance of fewer shares. Over a period of time this will result in slightly less dilution, higher equity value per share and better earnings per share.' [Commonwealth Edison Proxy Statement, 1976.]

[20]E.g., Commonwealth Edison suggests, 'Selling pressures often unduly depress both stock and rights values during the two or three week offering period which is a practical necessity when stock is sold with preemptive rights. Because the majority of stockholders do not exercise their rights but offer them for sale, the market value of the rights is driven far too low. Outsiders are then able to benefit by selling large amounts of stock during the offering period while buying rights for almost nothing and then exercising their rights to purchase stock at a discount to cover their sales. As a result, rights offerings tend to cost the company more than the rights themselves are worth to the stockholders who get them.'

are all insignificantly different from zero[21] and there is no significant sign pattern in the time series of average residuals. The cumulative average residuals in table 2 are also at approximately the same level three months before the

Table 2

Summary of average residual and cumulative average residual analysis of 853 rights offerings between 1926 and 1975 for the 25 event months [−12 to +12] surrounding the offer date.

Event month	Average residual	Cumulative average
−12	0.00721	0.00721
−11	0.01004	0.01725
−10	0.00255	0.01980
− 9	0.00629	0.02609
− 8	0.00388	0.02997
− 7	0.01062[a]	0.04059
− 6	0.00750	0.04809
− 5	0.00622	0.05431
− 4	0.01334[a]	0.06765
− 3	0.00662	0.07427
− 2	0.01624[a]	0.09051
− 1	−0.00649	0.08401
0	−0.00739	0.07663
+ 1	0.00779	0.08441
+ 2	0.00412	0.08853
+ 3	0.00405	0.09258
+ 4	−0.00110	0.09149
+ 5	−0.00047	0.09102
+ 6	0.00053	0.09155
+ 7	−0.00338	0.08817
+ 8	−0.00387	0.08430
+ 9	0.00256	0.08686
+10	−0.00264	0.08422
+11	−0.00013	0.08408
+12	−0.00476	0.07933

[a]Greater than 2σ. (Computation of the standard deviation is described in footnote 21.)

offering, on the date of the offering and 12 months after the offering. The significant positive residuals prior to the offer date are to be expected because of selection bias; firms which raise capital tend to have been doing well.

[21]As an estimate of the dispersion of an average residual, the approximation

$$\sigma^2 = (\sigma^2_M/r^2)(1-r^2)/N$$

was employed where σ^2_M is the variance of the market return, r^2 is the squared correlation coefficient between the return to an asset and the market return, and N is the number of securities in the sample. If σ_M is 0.089 [from Black Jensen Scholes (1972)], $r^2 = 0.25$, and $N = 853$ then $\sigma^2 = 0.000028$ and $\sigma = 0.00528$.

The results presented in table 2 are consistent with previous studies of this question Nelson (1965) examined all the rights offerings by firms listed on the New York Stock Exchange between January 1, 1946 and December 31, 1957. He found after the price series is adjusted for the 'split effect' in the rights offerings and general market movements are removed, prices six months after a rights offering are not significantly different from prices six months before the offering [22] Scholes (1972) found that the price of shares generally rose in value before the issue, fell 0 3 percent during the month of the issue, but experienced no abnormal gains or losses after the issue

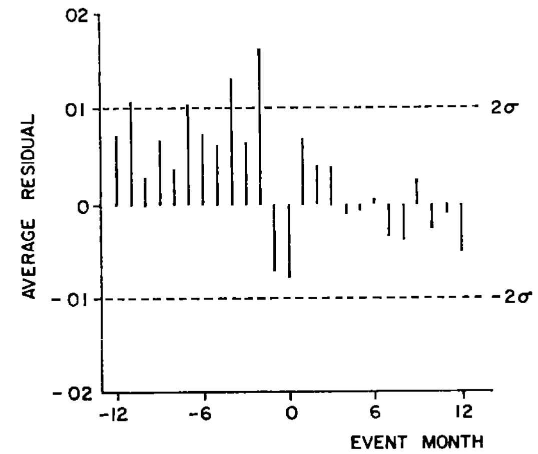


Fig 1 Plot of average residuals for 853 rights offerings between 1926 and 1975 for the 25 event months [− 12 to + 12] surrounding the offer date

4 2 Underwriters increase the stock price

Some argue that underwriters cause an increase in the stock price (1) by increasing 'public confidence' through external certification of the legal, accounting, and engineering analyses and (2) by the selling efforts of the underwriting syndicate.[23]

To examine the behavior of stock prices around the offer date of underwritten offerings and rights offerings, I obtained the returns for those securities which were included both in the sample of 578 firms covered in table 1 and on the CRSP daily return file There were 344 underwritten offerings and 52 rights offerings in this sample I set the offer date equal to day zero for all offerings and formed a portfolio of underwritten offerings and a portfolio of rights offerings I weighted securities in the portfolio of underwritten offerings so that

[22]The 'split effect' adjustment used by Nelson is derived in footnote 1
[23]See e g Brigham (1977, pp 473–474)

the two portfolios had equal betas. Then I calculated the difference in the portfolio returns for the 130 days before and 130 days after the offerings. The difference in average returns between two portfolios with equal risk will measure abnormal returns from either underwritten offerings or rights offerings. Table 3 presents the results for the period 20 days before the offering to 20 days after the offering; and figure 2 graphically presents the results for the period 40 days before to 40 days after the offering.

The average difference in returns to the two portfolios over the 260 days around the offer date is +0.00006, with a sample standard deviation of 0.00265. Therefore rights offerings have marginally higher returns during the 40 days around the offer date, but there is no obvious abnormal price behavior around the offer date for either underwritten offerings or rights offerings.

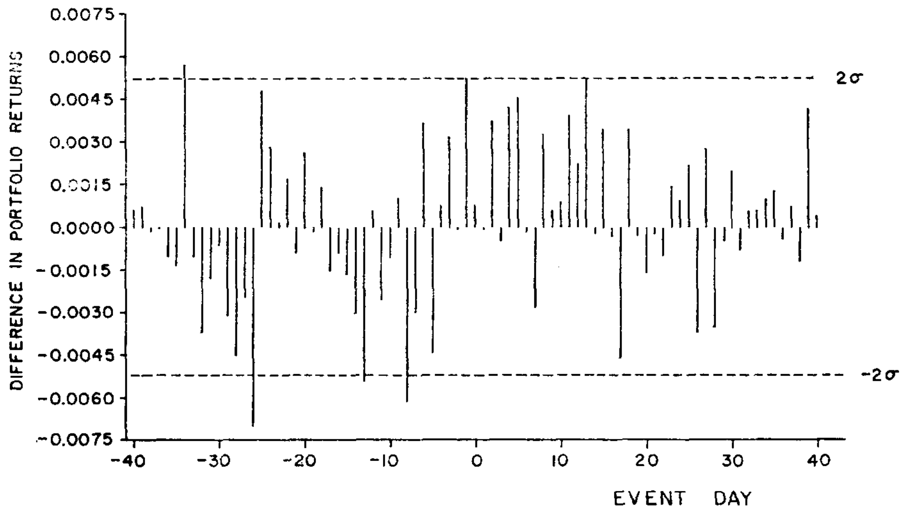


Fig. 2. Differences in daily returns between a portfolio of 52 rights offerings and a portfolio of 344 underwritten offerings for the 81 event days [−40 to +40] surrounding the offer date. (Portfolio weights are adjusted so that the two portfolios have the same beta.)

That underwriters are unable to generate abnormal positive price behavior should not be surprising. The firm always has the option of disclosing more information than is required by the Securities and Exchange Commission. The firm will expend resources on certification by external legal, accounting, and engineering firms until the net increase in the value of the firm is zero. Since the firm can contract for external certification of any disclosure, the benefit of whatever 'expert' valuation by the investment banker associated with an underwriting is limited to the difference in costs between certification through the underwriting process and independent certification.

But if underwriters are employed they influence the firm's decision about the

Table 3

Differences in daily returns between a portfolio of 52 rights offerings and a portfolio of 344 underwritten offerings between January 1971 and December 1975 for the 41 event days [−20 to +20] surrounding the offer date (Portfolio weights are adjusted so that the two portfolios have the same beta)

Event day	Rights average return	Underwritten average return	Difference (rights-und)	Cumulative difference
−20	−0 000361	−0 003007	0 002646	0 002646
−19	−0 001642	−0 001523	−0 000120	0 002526
−18	0 000072	−0 001361	0 001433	0 003959
−17	−0 001325	0 000175	−0 001500	0 002458
−16	−0 001134	−0 000231	−0 000902	0 001556
−15	−0 002865	−0 001229	−0 001636	−0 000080
−14	−0 002245	0 000732	−0 002977	−0 003057
−13	−0 004471	0 000949	−0 005420	−0 008477
−12	0 001722	0 001110	0 000611	−0 007866
−11	−0 002834	−0 000264	−0 002570	−0 010436
−10	−0 001226	−0 000125	−0 001102	−0 011538
− 9	0 001961	0 000960	0 001000	−0 010537
− 8	−0 004966	0 001151	−0 006117	−0 016654
− 7	0 001031	0 001327	−0 000296	−0 016950
− 6	0 002433	−0 001257	0 003690	−0 013260
−	−0 002373	0 002069	−0 004442	−0 017702
4	0 002180	0 001384	0 000797	−0 016905
− 3	0 001978	−0 001284	0 003262	−0 013642
− 2	−0 000570	−0 000557	−0 000013	−0 013656
− 1	0 004425	−0 000803	0 005228	−0 008428
0	0 001413	0 000583	0 000829	−0 007598
1	−0 000000	0 000054	−0 000054	−0 007653
2	0 003127	−0 000605	0 003732	−0 003921
3	−0 001182	−0 000700	−0 000482	−0 004403
4	0 003059	−0 001195	0 004254	−0 000149
5	0 005288	0 000710	0 004577	0 004428
6	0 000311	0 000477	−0 000166	0 004262
7	−0 002551	0 000206	−0 002757	0 001505
8	0 004396	0 001072	0 003324	0 004829
9	0 000851	0 000221	0 000630	0 005458
10	0 001601	0 000720	0 000881	0 006339
11	0 004703	0 000768	0 003934	0 010273
12	0 002369	0 000099	0 002271	0 012544
13	0 004764	−0 000502	0 005267	0 017811
14	−0 000734	−0 000495	−0 000239	0 017572
15	0 002944	−0 000527	0 003471	0 021043
16	−0 001089	−0 000790	−0 000299	0 020744
17	−0 001809	0 003065	−0 004874	0 015870
18	0 001228	−0 002196	0 003424	0 019294
19	0 000169	0 000458	−0 000289	0 019004
20	−0 000823	0 000711	−0 001534	0 017471

level of disclosure The underwriters will request that level of disclosure for which the marginal private costs and benefits to the underwriter are equal Given the legal liability of underwriters under the 1933 Act, the incentives of the firm and underwriter can differ Any divergence from the level of disclosure which maximizes the market value of the firm imposes a cost on the shareholders, and underwriters do ask for 'comfort letters' from accountants, frequently requiring expensive auditing procedures not produced without underwriters Thus, I conclude that the disclosure incentives of the underwriters lead to an over-investment in information production However, the costs of this over-investment should be reflected in the figures in table 1

4 3 Do underwriters underprice the securities?

In Ibbotson's (1975) study of unseasoned new issues he found that the offer price on average is set 11 4 percent below the market value of the shares If seasoned new issues are also underpriced, the difference between market value and offer price would represent another cost of employing underwriters

There are reasons to believe that underwriters underprice the seasoned new issues For a firm commitment underwriting agreement the Rules of Fair Practice of the National Association of Securities Dealers[24] require that once the offer price is set, the underwriter cannot sell the shares at a higher price. If the offer price is set above the market value of the shares excess supply results If the offer price presents a binding constraint to the underwriter, the limit order placed with the specialist by the managing underwriter results in the purchase of additional shares at the offer price If continued this purchasing would cause the underwriting syndicate to break Since very few underwriting syndicates break,[25] the implication must be either that the offer price is generally set below the market value of the shares, or that the offer price constraint can be circumvented

There are two ways in which the offer price could be circumvented First, for hot issues (i e , underpriced issues for which there is significant excess demand) the underwriters allocate the shares to preferred customers One way to achieve preferred customer status is to purchase issues for which there is an excess supply Second, underwriters employ 'swaps' In a swap, the underwriter buys another security from a customer while selling the underwritten security at the offer price Through this tie-in sale, the underwriter can shift the profit or loss These two tying arrangements allow the underwriter to minimize the impact of the regulation

[24]Although the rules of fair practice were established by the NASD, and not Congress or the SEC, there is little difference in the impact These rules are a response to the SEC's self regulatory position If the SEC found them unsatisfactory the SEC could establish superseding regulation

[25]See *History of Corporate Finance for the Decade* (1972)

To see if seasoned new issues are underpriced I calculated the return from the closing price the day prior to the offer date to the offer price, and the return from the offer price to the close on the offer date For the 328 firms with the requisite data, the average return from the close to the offer price is $-0\ 0054$ and the average return from the offer price to the close on the offer date is $+0\ 0082$ For the 260 days around the offer date the average daily return is 0 0005 with a sample standard deviation in the time series of average returns of 0 0013 Therefore, both figures, although much smaller than the 11 4 percent found by Ibbotson, are *significantly different from the average daily return* [26] Thus the underpricing imposes an additional cost on the owners of the firm of between 0 5 and 0 8 percent of the proceeds of the issue, a cost which is not reflected in table 1

5. Miscellaneous arguments favoring underwritten offerings

5 1 Insurance

It is frequently argued that employing an underwriter provides an 'insurance policy', reducing uncertainty of the offering's success [27] In effect, the firm

[26]One difference between Ibbotson's unseasoned issues and the seasoned issues examined here is that the unseasoned shares trade on the OTC market One hypothesis which has been suggested to explain the differences in the results is that the underpricing is a method of compensating the underwriter for maintaining a secondary market in the security Although the argument can explain why underwriter's compensation (including underpricing costs) for unseasoned issues is higher than for seasoned issues it does not explain the differential underpricing

[27]Another type of '*insurance*' *might be relevant If material errors are found in the regis*tration statement of a public issue, parties who allege damage can bring suit The suit typically names as co-defendants the firm, the board of directors of the firm, the firm s accountants, and the firm s underwriter If the underwriter assumes a large share of the liability for the error, sheltering the firm from suit, then the underwriter will receive a normal compensation for bearing that risk

Direct evidence on the hypothesis that underwriters reduce the firm's liability in case of a suit is expensive to obtain, economic studies of securities fraud suits have not been published However indirect evidence suggests that this factor cannot be of a sufficiently large magnitude to make this an important factor in the choice of underwritten issues over rights issues First, damage must be demonstrated – i e in addition to finding a material misstatement in the registration statement, the share price must have fallen after the offering Second, the underwriters explicitly seek to limit their liability as much as is legally feasible '[Issuer-Underwriter Indemnification] agreements are universally used in today s underwriting These agreements, although varying in specific language provide essentially for indemnification of the 'passively' guilty party by the party whose omissions or misstatements were the source of the liability ' (See 'The Expanding Liability of Security Underwriters', *Duke Law Journal*, Dec 1969, pp 1191–1246) Thus underwriters contracts seek to minimize their exposure in this area Third if the courts imposed a significant share of the responsibility for material errors on the underwriter, it would be expected that accounting firms would recognize this by offering lower rates for securities work to firms employing underwriters This does not seem to be the case At least when this issue was raised with several partners of eight big accounting firms, this effect was denied The judicial procedure tends to make the liability of each of the groups of defendants in this type of suit virtually independent.

purchases an option to sell the shares to the underwriter at the offer price (See Appendix 2) Note four things about this option First, in an underwritten issue, the offer price is not set generally until within 24 hours of the offering when the final agreement is signed, and hence the net proceeds are not determined until that time Second, as shown in section 4 3, the offer price on average is set below the market value of the stock Thus, the firm purchases a one-day option to sell shares at a discount of ½ percent below their market value Third, subject to certain conditions specified in the letter of intent, the underwriter has the option of backing out of the tentative agreement until the date the final agreement is signed Thus, the 'insurance policy' is of limited value because its effective duration is short Fourth, as argued above, the subscription price for a rights offering can be set low enough so that the probability of failure of the rights offering becomes arbitrarily close to zero So an alternate source of 'self-insurance' is available through the rights offering For these reasons, the possible value of the 'insurance policy' associated with underwritten issues must be small

5 2 Timing

Commonwealth Edison claims that the proceeds of an underwritten issue are available to the firm sooner than in a rights issue [28] But timing benefits provided by underwriters must be small First, the settlement date for an underwritten issue is generally seven days after the offer date, while the settlement date for a rights offering is generally seven days after the expiration of the offering Since the offering generally lasts about 18 days, any reasonable estimate of the cost in terms of the lost interest which would be imposed on the firm by waiting that short period of time would have to be small Second, since it is not expected that the rights will be exercised prior to their expiration,[29] the owners of the firm have the use of the funds during the period of the offering Thus, the time period which entails an opportunity cost of the funds is reduced to a seven- to ten-day period both for rights and underwritten offerings Third, if the services provided by the underwriter and transfer agents are competitively supplied, the fees charged will reflect the opportunity cost of the funds at their disposal This would imply that the timing cost is impounded in the figures in table 1 And fourth, unless there is an unforeseen urgency associated with obtaining the funds, the firm can simply initiate the rights procedure at an earlier date

Moreover, under certain circumstances, the registration procedure with the SEC is simpler when a rights issue is employed It is my belief that with a rights offering, the SEC is more likely to presume a regular dialogue between the firm and its owners and thus impose less restrictive disclosure requirements There-

[28]Commonwealth Edison Proxy Statement, 1976
[29]See Merton (1973) or Smith (1976)

fore, the time until the registration becomes effective can be expected to be shorter with a rights offering than with an underwritten offering. This shorter registration time reduces the total time from the point where the decision is made to raise additional capital to the receipt of the proceeds.

5 3 *Distribution of ownership*

Weston and Brigham (1975) argue that underwriters provide a wider distribution of the securities sold, 'lessening any possible control problem' Since change in control may result in a change in management, this is likely to be a relevant issue for the current management. Yet it is not clear that possible control problems should be a concern of the owners I know of no reason to believe that one group of owners is any better (i e , will price the firm any higher) than another group

Furthermore, it is not obvious that underwriters will achieve a wider distribution of ownership than will a rights offering For most rights offerings of listed firms, the consensus among investment bankers is that the subscription rate of the current owners of the firm ranges from 20 to 50 percent It is difficult to estimate what percentage of an underwritten issue is purchased by the current owners of the firm, but there is no reason to believe it is zero Further, underwritten issues seem to attract more institutional interest, resulting in large block purchases and therefore more concentration of ownership

These factors preclude any general conclusions about the effect of financing method on ownership distribution With this uncertainty it is not clear that management, even if concerned with control issues, should prefer the use of an underwriter

5 4 *Consulting advice*

Van Horne (1974) suggests that 'advice from investment bankers may be of a continuing nature, with the company consulting a certain investment banker or group of bankers regularly' It is more expensive for the firm to compensate the investment banker for future consulting services by including in the underwriting fee a payment for the present value of the expected advice Costs incurred in raising capital are not tax deductible, they directly reduce the capital account and do not enter the income statement Thus, compared to separate billing for services rendered, paying for future consulting through a higher underwriting fee doubles its cost for a firm with a marginal tax rate of 50 percent

5 5 *Expected legal costs*

If there were a law, regulation, or merely an unresolved judicial principle which might impose additional liability on a firm using rights offerings, then the expected legal costs of using rights could explain the observed use of underwriters But I can find no differential legal liability associated with the use of rights offerings

5 6 Selection bias

If the firms which employ rights offerings were systematically different from the firms which employ underwritten offerings, then the observed cost differences could be attributable to selection bias It could be that if the firms which employed underwriters had used rights, their expenses would have been greater

There is a significant difference in the betas of the firms in the two groups I calculated the betas for those firms in the sample which were listed on the New York Stock Exchange and included on the daily CRSP tape The average beta for the 344 underwritten offerings is 0 731 with a standard deviation of 0 560, and the average beta for the 52 rights offerings is 0 493 with a standard deviation of 0 330 But I can find no other systematic difference between the two populations

Examination of the data shows similar distributions of firms across industries, 80 8 percent of the firms employing rights and 73 2 percent of the firms employing underwritten offerings were utilities (electric, gas, or telephone companies) I attempted to predict the choice of underwritten versus rights offering based on the following variables (1) the percentage of the firm which is sold through the offering, (2) the market value of the firm, and (3) the variance of the returns on the stock The r^2 for the regression is 0 016 None of the t statistics for the variables appears to be significant

Although differences exist between the two sets of firms, the nature and magnitude of the differences seem insufficient to account for the observed cost differences

6. A monitoring cost hypothesis

6 1 Why not monitor the choice of financing method?

My examination of alternative financing methods suggests that rights offerings are significantly less expensive than underwritten offerings Yet underwriters are employed in over 90 percent of the offerings studied One hypothesis consistent with the evidence is (1) managers and members of the board of directors receive benefits from the use of underwriters which do not accrue to the other owners of the firm, and (2) the expenses which would be imposed on the owners of the firm by monitoring the managers and directors in the choice of financing method are greater than the costs without monitoring

Managers or members of the board of directors may recommend that offerings be underwritten because their welfare increases as a by-product of the use of underwriters in several ways [30] First, firms frequently include an investment banker as a member of the board of directors It is in his interest to lobby for the use of underwriters, particularly the use of his investment banking firm as managing underwriter Second, there is the possibility of 'bribery' This may be simply consumption for the managers and directors through 'wining and dining' by the underwriters But there is a more important possibility In an underwritten issue, if the offer price is set below the market value of the shares, the issue will be oversubscribed To handle this excess demand, underwriters ration the shares In the rationing process the underwriters presumably favor their preferred customers, and preferred customer status could be given to key management people or members of the board of directors of firms employing the underwriter This form of payment would be virtually impossible to detect, since the shares the officer of Company A would favorably acquire are those of Company B and would therefore call for no disclosure [31]

Further possible benefits to managers include the reduction of possible control problems, if underwritten offerings produce a wider distribution of ownership than rights offerings Finally, managers whose compensation is a function of reported profits will prefer an underwriter's fee which includes a payment for future consulting advice, the manager's compensation will be higher because payment through underwriting does not affect reported profits while separate billing for consulting does

Jensen and Meckling (1976) show that the costs which the managers and directors can impose on the other owners of the firm are limited by the costs of monitoring their activities Thus the cost to shareholders of monitoring the method of raising capital must be greater than the costs imposed by the financing method chosen Given the dispersion of ownership in modern corporations, the benefit to any single shareholder from voting his shares is small Thus the costs that he would rationally incur in voting are small,[32] and the resources the shareholder would rationally devote to deciding whether a 'yes' or 'no' vote is more in his interest are few Moreover, voting procedures in most corporations ensure that management has a disproportionate voice in the outcome Management is often assigned votes by proxy, and in many firms management has the

[30]Certain management compensation plans, such as stock option plans, make managers' compensation a function of the price of the firm's shares If the compensation plan were not adjusted to reflect the effect of the rights offering on the share price, management could be expected to provide a strong lobby in favor of employing underwriters In fact, however, employee stock option plans have general clauses calling for adjustment of the terms of the plan to reflect relevant capital structure changes Furthermore, most plans include specific reference to rights issues Thus, agency costs resulting from compensation plans do not seem to offer an explanation of the observed behavior

[31]This argument is similar to that of Manne (1966), especially Chapter V

[32]See Downs (1957) Basically, if a person owns 100 shares in a firm, his vote only matters if the vote is tied or his 'side' would have lost by 100 votes or less The probability is low that out of 50 million votes, the issue will split that way Thus the expected benefit (benefit times probability) of voting is very small

power to vote unreturned proxies They are also permitted to vote proxies on specific questions when the stockholder does not specify a choice These factors raise the cost of monitoring management

6 2 The preemptive right as a monitoring tool

There appears to be a low cost method of monitoring the use of underwriters the preemptive right The preemptive right is a provision which can be included in a firm s charter requiring the firm to offer any new common stock first to its existing shareholders But the inclusion of the preemptive right does not solve the problem firms can still employ underwriters through a standby under-

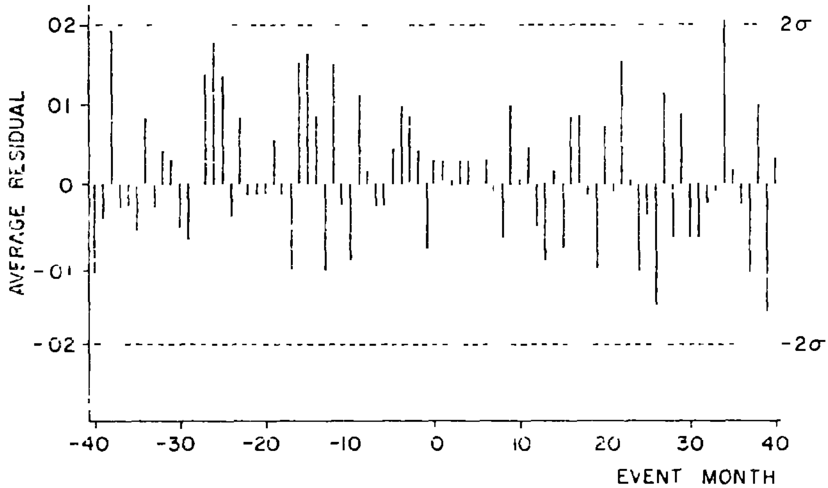


Fig 3 Plot of average residuals from 89 firms which removed the preemptive right from their corporate charter for the 81 event months [−40 to +40] surrounding the month of removal

writing agreement Since the figures in table 1 suggest a negligible difference in costs between a firm commitment underwritten offering and a rights offering with a standby underwriting agreement what becomes important is not a requirement to use rights, but a prohibition against using underwriters

To test the hypothesis that the impact of removing the preemptive right from the corporate charter is negligible, I collected a sample of 89 firms listed on the New York Stock Exchange which have removed the preemptive right The results of this study are presented in table 4 and figure 3 The average residual in the month of removal is 0 277 percent, and the mean average residual for the six prior months is 0 309 percent There is no apparent impact

I believe the results in table 4 provide a plausible explanation for why the intellectual level of the argument involving the preemptive right is so low on both sides of the question For example, the above quotes from Commonwealth

Table 4

Summary of residual analysis of 89 firms which removed the preemptive right from their corporate charter for the 81 event months [−40 to +40] surrounding the month of removal

Event month	Average residual	Cumulative average residual	Event month	Average residual	Cumulative average residual
−40	−0 00995	−0 00995	1	0 00363	0 11718
−39	−0 00382	−0 01376	2	0 00028	0 11745
−38	0 01999	0 00623	3	0 00293	0 12038
−37	−0 00258	0 00365	4	0 00276	0 12315
−36	−0 00160	0 00205	5	0 00101	0 12415
−35	−0 00414	−0 00209	6	0 00336	0 12751
−34	0 00842	0 00633	7	−0 00017	0 12734
−33	−0 00238	0 00395	8	−0 00537	0 12196
−32	0 00483	0 00878	9	0 00963	0 13159
−31	0 00375	0 01254	10	0 00002	0 13162
−30	−0 00419	0 00834	11	0 00406	0 13568
−29	−0 00632	0 00202	12	−0 00446	0 13122
−28	0 00082	0 00284	13	−0 00855	0 12266
−27	0 01337	0 01621	14	0 00210	0 12476
−26	0 01839	0 03460	15	−0 00696	0 11780
−25	0 01440	0 04900	16	0 00903	0 12683
−24	−0 00397	0 04503	17	0 00752	0 13435
−23	0 00800	0 05303	18	−0 00096	0 13339
−22	−0 00102	0 05201	19	−0 00942	0 12397
−21	−0 00007	0 05195	20	0 00701	0 13097
−20	−0 00072	0 05123	21	−0 00021	0 13077
−19	0 00602	0 05725	22	0 01591	0 14668
−18	−0 00067	0 05658	23	0 00090	0 14758
−17	−0 01032	0 04626	24	−0 01043	0 13715
−16	0 01575	0 06201	25	−0 00281	0 13434
−15	0 01608	0 07809	26	−0 01389	0 12046
−14	0 00828	0 08637	27	0 01069	0 13115
−13	−0 00943	0 07694	28	−0 00566	0 12548
−12	0 01496	0 09190	29	0 00901	0 13449
−11	−0 00183	0 09007	30	−0 00592	0 12857
−10	−0 00833	0 08174	31	−0 00624	0 12233
− 9	0 01103	0 09277	32	−0 00240	0 11993
− 8	0 00138	0 09415	33	−0 00071	0 11922
− 7	−0 00185	0 09230	34	0 02059	0 13981
− 6	−0 00170	0 09060	35	0 00183	0 14165
− 5	0 00508	0 09568	36	−0 00263	0 13901
− 4	0 00998	0 10566	37	−0 01103	0 12799
− 3	0 00816	0 11382	38	0 00971	0 13770
− 2	0 00477	0 11859	39	−0 01524	0 12246
− 1	−0 00782	0 11078	40	0 00300	0 12546
0	0 00277	0 11355			

Edison's Proxy Statement are demonstrably false, and the quote from AT&T's Proxy Statement is irrelevant The primary lobbying effort in favor of the preemptive right is from Lewis D Gilbert, John J Gilbert and Wilma Soss who regularly introduce proposals to reincorporate the preemptive right into the corporate charter of corporations which have removed it However, their reason for the use of rights is so that shareholders can maintain their proportionate interest in the firm For large firms this 'benefit' has negligible value [33]

6 3 Other considerations

It should be emphasized that the monitoring cost hypothesis is consistent with both observed institutional arrangements and rational, wealth-maximizing behavior by the stockholders Rational behavior implies that actions will be taken if the benefits exceed the costs I have pointed out certain costs associated with the voting mechanism within corporations inclusion of an investment banker on the board of directors, and certain management compensation plans These practices, while costly, would still be in the stockholders' best interests if there are offsetting benefits

Furthermore, the monitoring cost hypothesis does not imply that there are rents which accrue to the underwriting industry There are two available 'technologies' with which additional equity capital can be raised If the underwriting industry is competitive, the underwriting fees reported in table 1 would reflect a normal return to the resources required in employing that technology

However, the monitoring cost hypothesis does present some problems I do not observe the costs of monitoring management Hence the hypothesis is not directly tested Furthermore, while the incentives set up through the voting mechanism suggest that it is plausible that monitoring costs are large enough to explain the observed use of underwriters, competition in the market for management should reduce the required monitoring expenditures If the use of rights offerings is in the best interests of stockholders, then it will pay potential managers to incur bonding costs to guarantee not to use underwriters

7. Conclusions

In my examination of the choice of method for raising additional equity capital by listed firms I demonstrate that properly constructed rights offerings provide proceeds which are equivalent to those of an underwritten offering Furthermore, estimates of expenses from reports filed with the Securities and

[33]For a firm with 50 million shares outstanding, a ten percent increase in the number of outstanding shares would change the percentage ownership for someone with 100 shares only in the sixth decimal place With so many inexpensive alternate ways for a stockholder to maintain his proportionate interest in the firm the proportionate interest argument lacks importance

Exchange Commission indicate that rights offerings involve lower out-of-pocket costs than underwritten offerings Yet underwriters are employed in over 90 percent of the issues Examination of the arguments to justify the use of underwriters advanced by the underwriting industry, finance textbooks, corporate officers, and securities lawyers suggest that none of the arguments are capable of explaining the observed choice of financing method in terms of rational, wealth-maximizing behavior by the stockholders of the firm

The one hypothesis I find which is consistent with the available evidence relates to the costs of monitoring management Although direct expenses imposed on shareholders are higher per dollar raised through the use of underwriters, I hypothesize that management derives benefits from their use From the shareholders' standpoint, the firm's use of underwriters is optimal because the cost of monitoring management exceeds the savings in out-of-pocket expenses from using rights If this hypothesis is correct, then the present value of the stream of differences in costs reported in this paper provides a lower bound on the costs of getting shareholders together to monitor and control management on the method of raising capital Thus, the present value of the differences in costs establishes a lower bound on the expected costs of control mechanisms such as proxy fights, tender offers, and takeover bids

The monitoring cost hypothesis does present some problems I do not observe directly the costs of monitoring management While it is possible that the monitoring costs are large enough to explain the observed choice of underwriters, consideration of competition in the market for management reduces the plausibility of this hypothesis But if the monitoring cost hypothesis is rejected, then the observed choice of financing method cannot be explained in terms of rational, wealth-maximizing behavior by the owners of the firm, unless it can be shown that I have either ignored or misestimated a relevant cost of using rights or benefit from using underwriters

Appendix 1: A description of the institutional arrangements for rights and underwritten offerings

A description of the procedures followed in the various types of offerings specified in sufficient detail to answer the questions addressed in this study is not available This appendix provides that information Some of this material comes from written sources [34] However, much of the material comes from conversations with underwriters, corporate financial officers, and SEC officials.

Underwritten offerings

The firm typically selects an underwriter in one of two ways – either by competitive bidding or by negotiated underwriting In competitive bidding, the firm

[34] See Weston and Brigham (1975), SEC (1974), and Pessin (1976)

files appropriate papers with the SEC, then specifies the terms of the issue and has potential underwriters submit sealed bids Government regulation requires the use of this procedure by electric utility holding companies the primary users of competitive bidding In a negotiated underwriting bid, the important variables in the underwriting contract are determined by direct negotiation between firm and underwriter

Negotiated underwriting begins with a series of pre-underwriting conferences, when decisions as to the amount of capital, type of security, and other terms of the offering are discussed Several general forms of the underwriting agreement can be employed [35] The first is a 'firm commitment' underwriting agreement, under which the underwriter agrees to purchase the whole issue from the firm at a particular price for resale to the public Almost all large underwriters employ this form In the second form, a 'best efforts' underwriting, the underwriter acts only as a marketing agent for the firm The underwriter does not agree to purchase the issue at a predetermined price, but sells the security for whatever price it will bring The underwriters take a predetermined spread and the firm takes the residual A variant of this agreement employs a fixed price but no guarantee on the quantity to be sold The third possibility is an 'all-or-nothing' commitment which requires the underwriter to sell the entire issue at a given price, usually within thirty days, otherwise the underwriting agreement is voided

If the corporation and underwriter agree to proceed,[36] the underwriter will begin his underwriting investigation, in which he assesses the prospects for the offering This investigation includes an audit of the firm s financial records by a public accounting firm, which aids in preparing the registration statements required by the Securities and Exchange Commission A legal opinion of the offering will be obtained from lawyers who typically participate in writing the registration statement Reports may also be obtained from the underwriter s engineering staff when applicable

Before a company can raise capital through a public offering of new stock it must comply with the Federal Law that governs such a sale – the Securities Act of 1933, and the Securities Exchange Act of 1934 The Securities and Exchange Commission, established to administer both laws, requires full disclosure of all pertinent facts about the company before it makes a public offering of new stock The firm must file a lengthy registration statement with the SEC setting forth data about its financial condition For underwritten issues,

[35]The underwriter may make a 'standby commitment' during a rights offering under which he will purchase and distribute to the public any amount of the rights issue not purchased by the present security holders This form will be discussed further below

[36]Agreements are usually subject to conditions, most allow the underwriters to void their obligation in the event of specified adverse developments For example, a negative finding in the lawyer's or auditor's reports may allow voiding the contract

the firm usually files the form S-1 or S-7 registration statement Form S-7 is less expensive, but requires certain conditions to qualify [37]

The SEC has 20 days to examine the registration statement for material omissions or misrepresentations If any error is found, a deficiency letter is sent to the corporation and the offering is delayed until the deficiency is corrected If no deficiency letter is sent, a registration statement automatically becomes effective 20 days after filing, except when the SEC notifies the firm that the commission's workload is such that it requires more time to review the registration statement [38] The firm will typically amend the registration statement to include the offer price and the offer date after the SEC has examined the rest of the statement This procedure allows the firm and underwriter to postpone the effective date of the registration statement until they agree the offering should proceed

In addition to the registration requirements under the Securities Act of 1933, firms must qualify their securities under the state securities laws, the so-called 'Blue Sky Laws', in those states where the securities are to be sold Some states are satisfied with SEC approval, others require a registration statement be filed with state securities commissioners

The underwriter usually does not handle the purchase and distribution of the issue alone, except for the smallest of security issues The investment banker usually forms a syndicate of other investment bankers and security dealers to assist the underwriting [39] During the waiting period between the filing and the offer date, no written sales literature other than the so-called 'red herring'

[37]For example, the majority of the board of directors have been members for the last three years, there have been no defaults on preferred stock or bond payments for the past 10 years, net income after taxes was at least $500,000 for the past five years, and earnings exceeded any dividend payments made over the past five years

[38]In 1960 and 1961, delays of four to six months occurred for this reason

[39]Prior to the passage of the Securities Act in 1933 most new issues were purchased by an originating house The originating house would resell the issue at a small increase in price to a so-called banking group, generally a few large houses The banking group would then sell the issue to an underwriting group, which in turn sold it to a selling syndicate – each sale occurred at a fractional increase in price The selling syndicate members, however, were liable for their proportional interest of any securities remaining unsold Late in the 1920s it became frequent practice to make the final group a so-called selling group, the members of which had no liability except for securities which they had purchased from the underwriting syndicate

The Securities Act, as amended shortly after its passage, contained a provision limiting an underwriter's liability for misstatements and omissions in the registration statement to an amount not 'in excess of the total price at which securities underwritten by him and distributed to the public were offered to the public' This Act changed the method of wholesaling securities, the use of the joint syndicate in handling registered securities disappeared Because of the provisions of the Act, it was to the advantage of the manager of the offering to have his fellow participants purchase direct from the company, since then the manager's liability under the Act became limited to the amount which the firm itself underwrote Liability for transfer taxes that would have been payable on the sale by the manager to the underwriters was thus avoided At the present time, underwriters of securities registered under the Act contract to buy directly from the issuer even though the manager of the offering signs the agreement with the issuer on behalf of each of the underwriting firms prospectus[40] and 'tombstone' advertisements[41] are permitted by the SEC However, oral selling efforts are permitted, and underwriters can and do note interest from their clients to buy at various prices These do not represent legal commitments, but are used to help the underwriter decide on the offer price for the issue Underwriters typically attempt to obtain indications of interest for approximately 10 percent more shares than will be available through the offering [42]

Before the effective date of the registration, the corporation's officers meet with the members of the underwriting group Given the personal liability provisions of the 1933 Act, this meeting is often identified as a due diligence meeting An investment banker who is dissatisfied with any of the terms or conditions discussed at this session can still withdraw from the group with no legal or financial liability Discussed at this meeting are (1) the information in the firm's registration statement, (2) the material in the prospectus, (3) the specific provisions of the formal underwriting agreement As a rule, all the provisions of the formal underwriting agreement are set except the final sales price

The 'Rules of Fair Practice' of the National Association of Security Dealers require that new issues must be offered at a fixed price and that a maximum offering price be announced two weeks in advance of the offering However, the actual offering price need not be established until immediately before the offering date In fact, the binding underwriting agreement which specifies the offer price is not normally signed until within 24 hours of the effective date of the registration

Once the underwriter files the final offering price with the SEC, the underwriters are precluded from selling the shares above this price The SEC permits the managing underwriter to place a standing order with the specialist to buy the stock at the public offer price If the underwriter buys more than 10 percent of the shares to be issued through this order, the syndicate usually breaks, permitting the stock to be sold below the offer price The syndicate can also be broken if the managing underwriter feels that the issue cannot be sold at the offer price [43] On the other hand, if all the indications of interest become orders

[40]The red herring prospectus derives its name from the required disclaimer on the front printed in red

A registration statement relating to these securities has been filed with the Securities and Exchange Commission but has not yet become effective Information contained herein is subject to completion or amendment These securities may not be sold nor may offers to buy be accepted prior to the time the registration statement becomes effective This prospectus shall not constitute an offer to sell or the solicitation of an offer to buy nor shall there be any sale of these securities in any state in which such offer, solicitation or sale would be unlawful prior to registration or qualification under the securities laws of any such state

[41]The very limited notice of the offering permitted is often presented in a form resembling the inscription on a tombstone – hence the name

[42]This procedure is like 'over-booking' on airplane flights

[43]Syndicates break infrequently, my impression is that this occurs less than five percent of the time See *History of Corporate Finance For the Decade* (1972)

for shares, the issue is oversold In that case the managing underwriter typically sells additional shares short and covers these short sales in the aftermarket

The final settlement with the underwriter usually takes place seven to ten days after the registration statement becomes effective At that time, the firm receives the proceeds of the sale, net of the underwriting compensation

Rights offering

Offering of stock to existing shareholders on a pro rata basis is called a rights offering Each stockholder owning shares of common stock at the issue date receives an instrument (formally called a warrant) giving the owner the option to buy new shares [44] One warrant or right is issued for each share of stock held [45] This instrument states the relevant terms of the option (1) the number of rights required to purchase one new share, (2) the exercise price (or subscription price) for the rights offering, (3) the expiration date of the rights offering

Before the offering, the firm must file a registration statement for these securities For rights offerings, the firm typically files either a form S-1 or S-16 registration S-16 is simpler, but has usage requirements similar to those of form S-7

After the SEC approves the registration statement, the firm establishes a holder of record date The stock exchange establishes the date five business days earlier as the ex rights date [46] All individuals who hold the stock on the ex rights date will appear in the company's records on the holder of record date and will receive the rights However, the rights can be traded on a 'when issued' basis Usually trading begins after the formal announcement of the rights offering To ensure that there is adequate time for the stockholders to exercise or sell their rights, the New York Stock Exchange requires that the minimum period during which rights may be exercised is 14 days Rights trade on the exchange where the stock is listed

Issuing rights is costly in terms of management s time, postage and other expenses, so it is in the best interest of the firm to ensure the success of the offering Therefore, the firm has an incentive to set the subscription price of the rights low enough to ensure that the rights will be exercised But some of

[44] In the 1880s it was customary to require a stockholder to appear in person in the office of the corporation to subscribe to the issue After the 1880s, it became customary to send out a printed slip of paper so the stockholders could sign and subscribe for the stock without actually having to appear Later, it became the practice to make these slips of paper transferable, so that they could be sold Around 1910 the engraved form of warrant was first issued

[45] The Uniform Practice Code of the National Association of Security Dealers, Inc, provides that subscription rights issued to security holders shall be traded in the market on the basis of one right accruing on each share of outstanding stock, except when otherwise designated by the National Uniform Practice Committee Thus, the price quotation will be based on a single right even though several rights may be necessary to purchase one new share

[46] This procedure is comparable to that used in setting the ex dividend date

the warrants of most offerings do expire unexercised These unexercised rights can be offered through an over-subscription privilege to subscribing shareholders on a pro rata basis Shares not distributed through the rights offering or through the over-subscription privilege can be sold by the firm either to investment bankers or directly to the public

Rights offerings with a standby underwriting agreement

A formal commitment with an underwriter to take the shares not distributed through a rights offering is called a standby underwriting agreement Several types of fee schedules are generally employed in standby underwriting agreements A single fee may be negotiated, the firm paying the underwriter to exercise any unexercised rights at the subscription price A two fee agreement employs both a 'standby fee', based on the total number of shares to be distributed through the offering and a 'take-up fee', based on the number of warrants handled The 'take-up' fee may be a flat fee or a proportioned fee [47] These agreements generally include a profit sharing arrangement on unsubscribed shares (e g, if the underwriter sells the shares for more than the subscription price, this difference in prices is split between the underwriter and the firm according to an agreed formula)

Underwriters are prohibited from trading in the rights until 24 hours after the rights offering is made [48] After that time, they can sell shares of the stock short and purchase and exercise rights to cover their short position in the stock, thus hedging the risk that they bear

Appendix 2: A contingent claims analysis of rights and underwriting contracts

The derivation of general equilibrium pricing implications of rights and underwriting contracts has not been presented Black and Scholes (1973) suggest the approach I employ to value rights, but they do not carry out the analysis or present the solution Ederington (1975) provides a model of under-

[47] A proportioned fee involves more than one price for the shares handled by the underwriter For example there may be one price for the first 15% of the issue, a higher price for from 15% to 30% of the issue, and a still higher price for any of the issue over 30% which is unexercised through the rights offering and must be purchased by the underwriter

[48] Through the late 1940s underwriters were prohibited from trading in the rights during the offering This arrangement increased the underwriter s risk because the 14-day time period allowed large adverse price movements in the stock The NYSE instituted a study in 1947 after the failure of three rights offerings They found than on 43 rights offerings which had been successful the total underwriting profit was approximately \$2 4 million, while on the three unsuccessful offerings, their losses were in excess of \$3 million Underwriters were reportedly refusing to sign standby agreements unless the offering period were as short as five days Since this violated NYSE rules no NYSE listed firms used rights issues with standby underwriting agreements In response to this impasse, the NYSE now allows underwriters to trade in the rights 24 hours after the rights offering is made

writer behavior, but his model assumes underwriters maximize expected profits, and thus does not represent a general equilibrium solution in a market where the agents are risk averse The option pricing framework employed here will yield a solution which is consistent with general equilibrium, no matter what the risk preferences of the agents in the market.

I employ the contingent claims pricing techniques to derive a specification of the equilibrium value of these contracts For valuing both contracts I assume

(1) There are homogeneous expectations about the dynamics of firm asset values and of security prices The distribution of firm values at the end of any finite time interval is log normal The variance rate, σ^2, is constant
(2) Capital markets are perfect There are no transactions costs or taxes and all traders have free and costless access to all available information Borrowing and perfect short sales of assets are allowed Traders are price takers in the capital markets
(3) There is a known constant instantaneously riskless rate of interest, r, which is the same for borrowers and lenders
(4) Trading takes place continuously, price changes are continuous and assets are infinitely divisible
(5) The firm pays no dividends

Rights offerings

To derive the equilibrium value of the rights offering I make the following assumptions about the specification of the rights offering

The total proceeds to the firm if the rights are exercised is X (the exercise price per share times the total number of shares sold through the rights issue) The rights expire after T time periods If the rights are exercised, the shares sold through the offering will be a fraction, γ, of the total number of shares outstanding ($\gamma \equiv Q_R/(Q_S+Q_R)$, where Q_R is the number of shares sold through the rights offering and Q_S is the existing number of shares) Any assets acquired with the proceeds of the rights offering are acquired at competitive prices [49]

Given the above assumption, Merton (1974) has demonstrated that any contingent claim, whose value can be written solely as a function of asset value and time must satisfy the partial differential equation

$$\frac{\partial f}{\partial t} = \frac{1}{2}\frac{\partial^2 f}{\partial V^2}\sigma^2 V^2 + rV\frac{\partial f}{\partial V} - rf, \tag{A1}$$

[49]This last assumption is necessary to avoid the problem of the dependence of the dynamic behavior of the stock price on the probability of the rights being exercised

where $f(V, t)$ is the function representing the value of the contingent claim [e.g., $R = R(V, t)$]. To solve this equation, normally two boundary conditions are required, one in the time dimension and one in the firm value dimension.

To derive the appropriate boundary condition in the time dimension, note that when the time to expiration is zero, R^*, the value of the rights at the expiration date will be either zero (in which case the rights will not be exercised) or, if the rights are valuable and are exercised, their value is their claim on the total assets of the firm, $\gamma(V^*+X)$ (where V^* is the value of the firm's assets and X is the proceeds from the exercise of the rights) minus the payment the right-holders must make, X:

$$R^* = \mathrm{Max}[0, \gamma(V^*+X)-X], \tag{A2}$$

where:

V^* is the value of the firm's assets at the expiration date of the issue.

X is the proceeds to the firm of the exercise of the rights.

γ is the fraction of new shares issued through the rights offering to the total shares of the firm (both old and new).

The most natural boundary condition in the firm value dimension is that when the value of the firm is zero, the value of the rights issue, R, is zero. However, the first assumption, that the distribution of firm values is log normal, insures that V can never be zero; therefore, this boundary condition will never be binding.

This equation can be solved by noting that no assumptions about risk preferences have been made, thus the solution must be the same for any preference structure which permits equilibrium. Therefore choose that structure which is mathematically simplest.[50] Assume that the market is composed of risk-neutral investors. In that case, the equilibrium rate of return on all assets will be equal. Specifically, the expected rate of return on the firm, and the rights will equal the riskless rate. Then the current rights price must be the discounted terminal price:

$$R = \mathrm{e}^{-rT} \int_{((1-\gamma)/\gamma)X}^{\infty} [\gamma V^*-(1-\gamma)X]L'(V^*)\,\mathrm{d}V^*, \tag{A3}$$

where $L'(V^*)$ is the log normal density function.

Eq. (A3) can be solved to yield:[51]

[50] See Cox and Ross (1976) or Smith (1976). For a mathematical derivation of this solution technique, see Friedman (1975), especially page 148.

[51] See Smith (1976, p. 16) for a theorem which can be employed to immediately solve (A3) to yield (A4).

$$R = \gamma VN\left\{\frac{\ln(\gamma V/(1-\gamma)X)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}\right\}$$
$$-e^{-rT}(1-\gamma)XN\left\{\frac{\ln(\gamma V/(1-\gamma)X)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}\right\}$$
$$= R(V, T, X, \gamma, \sigma^2, r) \tag{A4}$$

where $\partial R/\partial V$, $\partial R/\partial T$, $\partial R/\partial\gamma$, $\partial R/\partial\sigma^2$, $\partial R/\partial r > 0$ and $\partial R/\partial X < 0$

The indicated partial effects have intuitive interpretations Increasing the value of the firm, decreasing the exercise price (holding the proportion of the firm's shares offered through the rights offering constant), or increasing the proportion of the firm's shares offered through the rights offering (holding the total proceeds of the issue constant) increase the expected payoff to the rights and thus increases the current market value of the rights offering An increase in the time to expiration of the riskless rate lowers the present value of the exercise payment, and thus increases the value of the rights Finally, an increase in the variance rate gives a higher probability of a large increase in the value of the firm and increases the value of the rights

Underwriting agreements

To analyze the appropriate compensation to the underwriter for the risk he bears in the distribution of the securities make the following assumptions about the underwriting contract

Underwriters submit a bid, B, today which specifies that on the offer date, T time periods from now, the underwriter will pay B dollars and receive shares of stock representing fraction γ of the total shares of the firm He can sell the securities at the offer price and receive a total payment of Ω, or (if the share price is below the offer price) at the market price, $\gamma(V^*+B)$ If his bid is accepted, he will be notified immediately

Again, (A1) can be employed where $f(V, t)$ is the function representing the value of the underwriting contract (i e , $U - U(V, t)$) The boundary condition for this problem is

$$U^* = \mathrm{Min}[\gamma(V^*+B)-B, \Omega-B] \tag{A5}$$

This assumes that at the offer date the underwriter will pay the firm B dollars The shares which the underwriter receives represent a claim to a fraction γ of the total assets of the firm, V^*+B If the offer price is greater than the value of the shares, $\gamma(V^*+B)$, then the underwriter will be unable to sell the shares at the offer price, hence he will receive $\gamma(V^*+B)$ If, at the offer date the offer price is less than the value of the shares, the underwriter receives the offer price Therefore, the boundary condition is that at the offer date the underwriting contract is worth the minimum of the market value of the shares minus the bid, B, or the proceeds of the sale at the offer price minus the bid

Again, the above solution technique can be employed to solve (A1) subject to (A5). In a risk-neutral world, the expected value of the underwriting contract can be expressed as [52]

$$U = \int_0^{(\Omega/\gamma)-B} [\gamma(V^*+B)-B]L'(V^*)\mathrm{d}V^*$$

$$+\int_{(\Omega/\gamma)-B}^{\infty} [\Omega-B]L'(V^*)\mathrm{d}V^*. \tag{A6}$$

Note that this can be rewritten as

$$U = \int_0^{\infty} [\gamma(V^*+B)-B]L'(V^*)\mathrm{d}V^*$$

$$-\int_{(\Omega/\gamma)-B}^{\infty} \gamma\left[V^*-\left(\frac{\Omega}{\gamma}-B\right)\right]L'(V^*)\mathrm{d}V^* \tag{A7}$$

Eq (A7) can be solved for the risk-neutral case to yield

$$U = \mathrm{e}^{rT}\gamma V-(1-\gamma)B-\mathrm{e}^{rT}\gamma VN\left\{\frac{\ln(\gamma V/(\Omega-\gamma B))+(r+\sigma^2/2)T}{\sigma\sqrt{T}}\right\}$$

$$+(\Omega-B\gamma)N\left\{\frac{\ln(\gamma V/(\Omega-\gamma B)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}\right\} \tag{A8}$$

Examination of (A8) reveals that the underwriting contract is equivalent to a portfolio consisting of a long position in the firm, a cash payment, and writing a call on γ of the firm with an exercise price equal to $(\Omega-\gamma B)$

$$U = \mathrm{e}^{rT}\gamma V-(1-\gamma)B-\mathrm{e}^{rT}C(\gamma V, T, \Omega-\gamma B)$$

$$= \mathrm{e}^{rT}\gamma V-(1-\gamma)B-\mathrm{e}^{rT}\gamma C\left(V, T, \frac{\Omega}{\gamma}-B\right), \tag{A9}$$

where $C(\ )$ is the Black–Scholes call option function

If the process of preparing and submitting a bid is costless, then in a competitive equilibrium, the value of the underwriting contract must be zero [53]

[52] Since the contract calls for the payment only at t^*, to find the current value of the underwriting contract does not require discounting

[53] If this were not the case, arbitrage profits could be earned by acquiring an underwriting contract and establishing the above hedge

110138-OPC-POD-60-244

Therefore the bid which would represent a normal compensation for the risk he bears is implicitly defined by the equation [54]

$$B-e^{rT}\frac{\gamma}{1-\gamma}\left[V-C\left(V,T,\frac{\Omega}{\gamma}-B\right)\right]=0 \tag{A10}$$

The firm generally receives less than the market value of the stock[55] given the specification of the underwriting contract, if the equilibrium stock price at the offer date is above the offer price then the initial purchaser of the issue receives 'rents', he obtains the shares for less than the market value of the shares Therefore, if the offer price in the underwriting agreement represents a binding constraint to the underwriter, then in a perfect market underwriting must be a more expensive method of raising additional capital than is a rights issue Therefore, under these conditions, underwriting would not be employed

The above analysis implicitly assumes that the terms of the underwriting contract represent a binding constraint to the underwriter, i e , if the security price is above the offer price, then the offer price presents a constraint to the underwriter and a pure profit opportunity to the potential investor However, in a market without transactions costs, this could not be the case If the security price is above the offer price there will be excess demand for the issue To the extent that the underwriter can, through the rationing process, extract those profits, they will accrue to the underwriter rather than to the initial purchaser In this situation competition among underwriters would ensure that the profits were in fact garnered by the firm In that case the offer price presents no effective constraint and the competitive bid becomes simply

$$B=e^{rT}\left(\frac{\gamma}{1-\gamma}\right)V \tag{A11}$$

Therefore, if through tie-in sales or other means the offer price in an underwriting agreement can be circumvented, then underwriting is no more expensive a method of raising additional capital than a rights offering

[54] This equation implicitly defines the bid because B appears twice in the equation The explicit solution for equilibrium bid can be found by standard numerical analysis techniques

[55] A sufficient condition for the bid to be less than the market value of the shares is that $(1-\gamma)$ be less than $e^{\gamma T}$ Since T is generally a matter of days, this condition should be met

References

Bailey, M J , 1969, Capital gains and income taxation, in A C Harberger and M J Bailey, eds , The taxation of income from capital (Brookings, Washington, D C)

Black, F , M C Jensen and M S Scholes, 1972, The capital asset pricing model Some empirical tests, in M C Jensen, ed , Studies in the theory of capital markets (Praeger, New York)

Black, F and M Scholes, 1973 The pricing of options and corporate liabilities, Journal of Political Economy 81, 637–654
Brigham, E F, 1977, Financial management (Dryden, Hinesdale, Illinois)
Cox, J C and S A Ross, 1975, The pricing of options for jump processes, Rodney L White Center for Financial Research, Working Paper 2-75 (University of Pennsylvania, Philadelphia Pennsylvania)
Downs, A C, 1957 An economic theory of democracy (Harper, New York)
Ederington, Louis H 1975, Uncertainty, competition, and costs in corporate bond underwriting, Journal of Financial Economics 2 71–94
Fama E F L Fisher, M C Jensen, and R Roll, 1969, The adjustment of stock prices to new information International Economic Review 10 1–21
Friedman A 1975, Stochastic differential equations and applications Volume 1 (Academic Press, New York)
Gilbert, Lewis D and John J Gilbert, various dates Annual report of stockholder activities at annual meetings, New York
Hess Alan C, 1977 The role of investment banking in an efficient capital market, unpublished manuscript Securities and Exchange Commission, Washington, D C
History of Corporate Finance for the Decade, 1972 (Investment Dealers Digest, New York)
Ibbotson, Roger C, 1975, Price performance of common stock new issues, Journal of Financial Economics 2, 235-272
Jensen M C and W H Meckling 1976, Theory of the firm Managerial behavior, agency costs and ownership structure Journal of Financial Economics 3, 305-360
Manne, Henry G, 1966, Insider trading and the stock market (The Free Press, New York)
Merton, Robert C 1973 Theory of rational option pricing, Bell Journal of Economics and Management Science 4, 141-183
Merton, Robert C 1974 On the pricing of corporate debt The risk structure of interest rates, Journal of Finance 29
National Association of Security Dealers, NASD manual
Nelson, J R, 1965, Price effects in rights offerings Journal of Finance 20, 647–650
Pessin, A H 1973, The work of the securities industry (New York Institute of Finance, New York)
Schneider, Carl W and Joseph M Manko, 1970, Going public – Practice, procedure and consequences, Villanova Law Review 15, 283-312
Scholes, Myron S, 1972, The market for securities Substitution versus price pressure and the effects of information on share price, Journal of Business 45, 179–211
Smith, Clifford W, Jr, 1976 Option pricing A review, Journal of Financial Economics 3, 3–51
U S Securities and Exchange Commission, 1940, Cost of flotation for small issues, 1925–1929 and 1935-1938
U S Securities and Exchange Commission, 1941, Cost of flotation for registered securities 1930–1939
U S Securities and Exchange Commission, 1944, Cost of flotation of equity securities of small companies (Statistical Series Release no 744)
U S Securities and Exchange Commission, 1949, Cost of flotation, 1945–1947
U S Securities and Exchange Commission, 1951, Cost of flotation, 1945–1949
U S Securities and Exchange Commission, 1957, Cost of flotation of corporate securities, 1951-1955
U S Securities and Exchange Commission, 1970, Cost of flotation of registered equity issues, 1963–1965
U S Securities and Exchange Commission, 1974, Cost of flotation of registered issues, 1971–1972
Van Horne, J D, 1974, Financial management and policy (Englewood Cliffs, New Jersey)
Weston, J Fred and Eugene F Brigham, 1975, Managerial finance (Hinesdale, Illinois)
Zwick, Jack and Nathaniel R Norton, III, 1970, Investment banking and underwriting, in F G Zarb and G T Kerekes, eds, The stock market handbook, 54–71

Value Line Forecast for the U.S. Economy

	ACTUAL					ESTIMATED				
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
GROSS DOMESTIC PRODUCT AND ITS COMPONENTS (2005 CHAIN WEIGHTED $) BILLIONS OF DOLLARS										
Final Sales	12588	12917	13234	13341	13111	13283	13611	13998	14390	14822
Total Consumption	8819	9074	9314	9265	9154	9426	9695	9949	10188	10412
Nonresidential Fixed Investment	1347	1454	1544	1557	1291	1368	1440	1537	1676	1843
Structures	351	384	441	464	370	344	323	332	375	417
Equipment & Software	996	1070	1097	1082	916	1022	1117	1218	1315	1434
Residential Fixed Investment	775	718	585	444	343	348	358	458	523	575
Exports	1305	1422	1546	1648	1491	1642	1770	1891	2024	2175
Imports	2028	2151	2194	2152	1854	2116	2301	2446	2593	2735
Federal Government	876	895	906	972	1028	1074	1070	1032	1012	1002
State & Local Governments	1494	1507	1537	1533	1519	1523	1520	1521	1536	1559
Gross Domestic Product	12638	13399	14062	14369	14119	14657	15210	15858	16595	17434
Real GDP (2005 Chain Weighted $)	12638	12976	13229	13229	12881	13234	13536	13925	14343	14802
PRICES AND WAGES-ANNUAL RATES OF CHANGE										
GDP Deflator	3.3	3.2	2.7	2.2	0.9	1.7	1.3	1.5	1.6	1.8
CPI-All Urban Consumers	3.4	3.2	2.9	3.8	-0.3	1.2	1.8	2.1	2.2	2.3
PPI-Finished Goods	4.9	3.0	3.9	6.4	-2.5	3.2	1.5	2.3	2.3	2.5
Employment Cost Index—Total Comp.	3.1	2.9	3.1	2.8	1.5	2.1	2.0	2.3	2.4	2.6
Productivity	1.8	1.0	1.4	2.8	3.5	1.5	1.1	1.1	1.0	1.5
PRODUCTION AND OTHER KEY MEASURES										
Industrial Prod. (% Change)	3.3	2.2	1.7	-2.2	-9.3	5.2	3.1	3.6	3.8	4.0
Factory Operating Rate (%)	78.6	79.4	79.4	75.1	67.2	71.5	73.7	75.7	77.0	78.0
Nonfarm Inven. Change (2005 Chain Weighted $)	39.1	46.3	-3.7	-34.3	-116.9	61.9	37.5	45.0	50.0	55.0
Housing Starts (Mill. Units)	2.07	1.81	1.34	0.90	0.55	0.59	0.79	1.19	1.40	1.60
Existing House Sales (Mill. Units)	7.08	6.51	5.67	4.89	5.16	4.79	4.78	5.30	5.65	6.25
Total Light Vehicle Sales (Mill. Units)	17.0	16.5	16.1	13.1	10.4	11.5	12.8	14.6	15.3	16.0
National Unemployment Rate (%)	5.1	4.6	4.6	5.8	9.3	9.7	9.5	9.1	8.5	8.0
Federal Budget Surplus (Unified, FY, $Bill)	-321.0	-248.0	-162.0	-455.0	-1416	-1376	-1175	-1015	-750	-700
Price of Oil ($Bbl., U.S. Refiners' Cost)	56.56	66.12	72.18	99.75	59.40	77.70	82.00	88.00	94.00	100.00
MONEY AND INTEREST RATES										
3-Month Treasury Bill Rate (%)	3.1	4.7	4.4	1.4	0.2	0.2	0.3	1.1	3.0	3.5
Federal Funds Rate (%)	3.2	5.0	5.0	1.9	0.2	0.2	0.2	0.8	2.8	3.3
10-Year Treasury Note Rate (%)	4.3	4.8	4.6	3.7	3.3	3.2	2.6	3.4	4.5	4.8
Long-Term Treasury Bond Rate (%)	4.6	4.9	4.8	4.3	4.1	4.6	4.1	4.7	5.5	5.8
AAA Corporate Bond Rate (%)	5.2	5.6	5.6	5.6	5.3	5.0	4.8	5.6	6.0	6.5
Prime Rate (%)	6.2	8.0	8.0	5.1	3.3	3.3	3.3	4.6	6.5	7.0
INCOMES										
Personal Income (% Change)	5.6	7.1	6.1	3.8	-1.7	3.3	3.3	4.1	4.5	5.5
Real Disp. Inc. (% Change)	1.4	3.5	2.8	1.3	0.6	1.8	1.6	2.3	2.5	3.0
Personal Savings Rate (%)	0.4	0.7	0.6	1.8	5.9	5.6	4.8	3.8	3.5	3.5
After-Tax Profits ($Bill)	1207	1405	1436	1231	1062	1404	1400	1477	1550	1659
Yr-to-Yr % Change	34.5	16.4	2.2	-14.3	-13.7	32.2	-0.3	5.5	5.0	7.0
COMPOSITION OF REAL GDP-ANNUAL RATES OF CHANGE										
Gross Domestic Product	3.1	2.7	2.1	0.4	-2.6	2.7	2.3	2.9	3.0	3.2
Final Sales	3.1	2.8	2.4	1.4	-2.1	1.3	2.5	2.8	2.8	3.0
Total Consumption	3.0	3.0	2.8	0.2	-1.2	3.0	2.9	2.6	2.4	2.2
Nonresidential Fixed Investment	7.2	7.5	4.9	1.6	-17.1	6.0	5.2	6.8	9.0	10.0
Structures	1.3	8.2	12.7	11.2	-20.4	-7.0	-6.0	2.8	13.0	11.0
Equipment & Software	9.3	7.2	1.7	-3.0	-15.3	11.6	9.2	9.1	8.0	9.0
Residential Fixed Investment	6.3	-7.1	-17.9	-20.8	-22.9	1.6	3.0	27.9	14.0	10.0
Exports	7.0	9.1	8.4	6.2	-9.5	10.2	7.8	7.0	7.0	7.5
Imports	5.9	6.0	2.2	-3.5	-13.8	14.1	8.8	6.3	6.0	5.5
Federal Government	1.2	2.3	1.6	6.0	5.7	4.5	-0.3	-3.5	-2.0	-1.0
State & Local Governments	-0.1	1.3	2.3	1.1	-0.9	0.2	-0.2	0.0	1.0	1.5

110138-OPC-POD-60-247