Journal of Banking and Finance 10 (1986) 115–132. North-Holland

SYSTEMATIC RISK, TOTAL RISK AND SIZE AS DETERMINANTS OF STOCK MARKET RETURNS*

Josef LAKONISHOK

Tel Aviv University, Tel Aviv, Israel
Cornell University, Ithaca, NY 14853, USA

Alan C. SHAPIRO

University of Southern California, Los Angeles, CA 90007, USA

Received March 1984, final version received January 1985

This paper studies the historical relationship for the period 1962–1981 between stock market returns and the following variables: beta, residual standard deviation (or total variance), and size. We conclude that neither the traditional measure of risk (beta) nor the alternative risk measures (variance or residual standard deviation) can explain the cross-sectional variation in returns; only size seems to matter. When January returns are eliminated, even the size variable loses its statistical significance.

1. Introduction

Recent empirical work by both Banz (1981) and Reinganum (1981) has demonstrated that firm-size data can be used to create portfolios that earn 'abnormal' returns of up to 40 percent annually. In particular, the smaller a firm's capitalization, the greater the apparent abnormal returns.[1] These results appear to be inconsistent with the traditional single-period Sharpe–Lintner capital asset pricing model (CAPM), which posits a specific relationship between systematic risk (beta) and required asset returns.

The purpose of this paper is to test the hypothesis, suggested directly or indirectly by the work of Levy (1978), Mayshar (1979, 1981, 1983) and others,

*This paper was presented at the 'Small Firm Effect Conference' at the University of Southern California, and at the University of North Carolina, Chapel Hill, Duke University, Tel Aviv University, and the European Finance Association Meeting at INSEAD. We would like to thank the various participants at these seminars and, in particular, Hans Stoll, Fischer Black and Richard Roll for their helpful comments, without holding them responsible for any errors that remain. Alan Shapiro's work on this paper was financed by a research grant from the School of Business Administration, University of Southern California, which is gratefully acknowledged.

[1]More recent work by Keim (1983), Reinganum (1983), Roll (1983), and Lakonishok and Smidt (1984) has demonstrated the existence of 'an anomaly within an anomaly', namely that excess returns for small firms are concentrated within a short period at the turn of the year, rather than being distributed evenly throughout the year. Our empirical analysis deals explicitly with this so-called 'January effect'.

that a partial explanation for the small firm effect is that, due to transaction costs and other barriers to trade which limit investor diversification, the Sharpe–Lintner CAPM, which assumes complete diversification for all investors, is misspecified and thus an inappropriate measure of risk is being used to calculate risk-adjusted returns. In particular, we test the hypothesis that the shares of small firms, which generally are not widely held are affected more by their own variances than are the widely held shares of larger firms.[2]

The paper begins with our model specification in section 2. We discuss our methodology and data in section 3 and our empirical results in section 4. In section 5 we reformulate the model presented in section 2 in order to study the small firm effect in greater detail and analyze the results obtained from this reformulated model. We summarize our findings and present our conclusions in section 6.

2. Model specification

One possible explanation for the anomalous results reported by Banz (1981) and Reinganum (1981), which is the focus of this paper, is that other elements of risk besides data are priced.[3] Specifically, we test the proposition that, due to limited diversification, residual or unsystematic risk is priced in addition to systematic risk. Since residual risk appears to be inversely proportional to firm size [see, for example, evidence from Basu and Cheung (1982) as well as our table 2], this could account for some of the observed differences in returns. In addition, the unit compensation for residual risk may well be greater for smaller firms because of the lesser degree of diversification of small firm shareholders.

2.1. A generalized asset pricing model

Our empirical tests are based on a generalized linear asset pricing model of the form

$$\mathrm{E}(R_i)=R_f+\gamma_1\beta_1+\gamma_2 s_i+\gamma_3 \ln \phi_i|\phi_m, \quad \text{where} \tag{1}$$

$\mathrm{E}(R_i)$ = expected return on security i,
R_f = risk-free rate of interest,

[2]Even if the compensation per unit of variance is identical for large and small firms, the fact that small companies tend to have higher variances could also explain part of the return differential.

[3]The possibility that beta is not the sole measure of risk is suggested in a number of both theoretical and empirical papers including Levy (1978), Mayshar (1979, 1981, 1983), Klein and Bawa (1977), Lintner (1969), Friend and Westerfield (1981), and Basu and Cheung (1982).

γ_1 = premium for bearing market risk,
s_i = standard deviation of R_i,
γ_2 = premium for bearing total risk,
ϕ_i = market value of security i,
ϕ_m = average market value of all securities, and
γ_3 = constant measuring the effect of size on security i's return.

The inclusion of total risk (s_i) is justified by Levy (1978) and Mayshar (1979, 1981, 1983) who maintain that where holdings of an asset are concentrated in a relatively few, undiversified portfolios, the asset's own variance (standard deviation) will significantly affect its equilibrium required returns.

According to the CAPM, γ_2 and γ_3 should be zero. However, the arguments of Levy and Mayshar may have special relevance with regard to explaining the small firm effect. As shown in table 1, institutional investors, who are generally conceded to be the most diversified of investors, tend to underinvest in small firms, providing tentative support for the argument that the average investor in small firms is relatively less diversified than the average investor in larger firms.[4] Furthermore, in Vermaelen's (1981) study of

Table 1

Information collection and holdings of institutional investors by size.

Size	Number of firms followed by analysts	Percentage of all firms followed	Average number of price forecasts/ firm[a]	Average number of earnings estimates/ firm[b]	Percentage held by investment companies[c]	Percentage held by banks[d]
Below 100M	73	7.866	4.137	6.671	6.256	20.149
100–200	111	11.961	4.847	8.360	6.512	29.015
200–400	221	23.815	5.276	9.833	5.937	32.273
400–1000	264	28.448	7.572	13.098	4.863	36.189
1000–5000	226	24.353	10.566	16.903	4.085	43.331
Over 5000M	33	3.556	14.909	21.787	2.861	41.261
	928	100				

[a]From *Institutional Brokers Opinion Survey*, published by Lynch, Jones and Ryan. Provides average of stock performance forecasts by securities analysts and other statistics.
[b]From *Institutional Brokers Estimate System*, published by Lynch, Jones and Ryan. Contains data on expected Earnings/Share by securities analysts.
[c]From *Spectrum 1.*
[d]From *Spectrum 3.*

[4]Support is only tentative because these institutional investors serve as intermediaries for the owners of these funds who may themselves own shares in small firms. However, to the extent that shares of small firms comprise a disproportionate percentage of their holdings, these owners will still remain relatively undiversified.

common stock repurchases, the fraction of insider holdings in his sample, which was comprised primarily of small firms, was 17.5 percent. The implication is that the concentration of ownership in small firms is quite high.

There is a fundamental problem with the Levy–Mayshar line of reasoning, however; its conclusions are derived within a partial equilibrium framework. Specifically, their models and their results are driven by the behavior of the average, relatively undiversified investor in stocks with a low degree of ownership dispersion; they ignore the possibility of fully diversified investors arbitraging between securities whose returns are a function of systematic risk only and securities whose returns provide compensation for bearing total risk. Since it is the marginal investor, and not the average investor, who determines required returns in equilibrium, their theoretical models and the generally supportive empirical findings of Levy (1978), and Basu and Cheung (1982) deserve to be treated with some skepticism. We view our paper as providing a further and more comprehensive test of the basic hypothesis implied by the work of Levy and Mayshar. In the course of our investigation, we also test indirectly the hypothesized important role played by the marginal investor in theories of equilibrium pricing.

2.2. A stochastic return generating model

In order to test the hypothesis that total risk should be an important determinant of required return in addition to systematic risk (β), as well as to examine the importance of size, we run cross-sectional regressions for each month t based on the following stochastic version of (1):

$$R_{it}-R_{ft}=\gamma_{1t}\beta_{it}+\gamma_{2t}\sigma_{it}+\gamma_{3t}\ln\phi_{it}/\phi_{mt}+\varepsilon_{it}, \tag{2}$$

where σ_i is a measure of security i's unsystematic risk (estimated as the standard deviation of the least-squares residuals from the market model for security i) and ε_{it} is a random error term with mean zero. We use the residual risk (σ_i) instead of the total risk in (2) in order to reduce the problem of multicollinearity caused by the high correlation between the return variance and beta.[5]

The actual cross-sectional regressions use estimated betas $\hat{\beta}_{it}$ and residual standard deviations $\hat{\sigma}_{it}$, resulting in an errors-in-the-variables problem. We deal with this problem by following the standard technique of using portfolios instead of individual securities to estimate the gamma coefficients in (2). Securities are grouped into portfolios on the basis of size, beta and residual standard deviation (sigma) in order to increase the between-group

[5]This procedure may bias our results against finding the residual standard deviation significant.

variation in the independent variables. To cope with the 'regression-towards-the-mean' phenomenon (high observed $\hat{\beta}_i$ and $\hat{\sigma}_i$ tend to be above the true β_i and σ_i and vice versa for low observed $\hat{\beta}_i$ and $\hat{\sigma}_i$), we use a method developed by Fama and MacBeth (1973). Data from one period are used to estimate the betas and sigmas employed in forming the portfolios while data from a subsequent period are used to obtain the portfolio betas ($\hat{\beta}_p$) and sigmas ($\hat{\sigma}_p$). The $\hat{\beta}_p$ and $\hat{\sigma}_p$ are then used in the cross-sectional regression for the following period.

We estimate (2) using an ordinary least squares (OLS) regression, which assumes homoscedastic errors, as well as a generalized least squares (GLS) regression, which allows for heteroscedastic errors. In our particular case, the difference between OLS and GLS results is very small, the same result reported by Litzenberger and Ramaswamy (1979). We also estimate (2) using individual securities because of the potential loss of efficiency due to grouping as discussed by Litzenberger and Ramaswamy (1979).

It is important to note that our procedure examines the relationship between risk and return by using a predictive test; the explanatory variables in our analysis are estimated using data from prior years. A contemporaneous test, which uses the same period to estimate both the explanatory variables and the relationship between the variables, tends to overestimate the importance of the explanatory variables. It also biases upwards the explained variability of the dependent variable (the adjusted R^2). In addition, from an investor's standpoint, a contemporaneous test is not particularly relevant – an investor needs data from prior years to estimate beta and total risk. Unfortunately, both Levy (1978), and Basu and Cheung (1982), perform contemporaneous tests; their cross-sectional regressions use betas and variances estimated from the same period, thereby suffering from the problems of bias just described. Their analyses also ignore the potential bias associated with the 'January effect' and fail to account for the possible heteroscedasticity of the error terms.

3. Methodology and data

The sample for this study includes all stocks traded on the New York Stock Exchange for at least eight years between January 1954 and December 1981 and for which we had adequate return and market capitalization data. Monthly return data (dividends and capital gains, with appropriate adjustments) came from the monthly returns file of the Center for Research in Security Prices (CRSP) of the University of Chicago while the Annual Industrial Compustat and Research Annual Industrial Compustat tapes were used to provide market capitalization values.[6]

[6]Survivorship bias is not a serious problem because the Research Tape from 1971 on contains all firms, including those later eliminated. An addition, we performed tests on the two subperiods, 1962–1971 and 1972–1981, without detecting any difference in results.

Except for ranking on the additional explanatory variables as well as on beta, our methodology is similar to that described at length in Fama and MacBeth (1973). The tradeoff we faced in deciding on the number of portfolios to create was between forming more portfolios (which would increase the number of observations available in the monthly cross-sectional regressions) and having more companies in each group (with a corresponding increase in the precision of the estimated portfolio betas and sigmas). We decided to create five portfolios based on size and then to form four portfolios based on ranked $\hat{\beta}$ within each size group. We then divided each $\hat{\beta}$ portfolio into four additional portfolios ranked on $\hat{\sigma}$, yielding a total of 80 ($5 \times 4 \times 4$) portfolios for each test period. Over the period of our test, 1954–1981, the number of firms within each cell ranged from a minimum of six to a maximum of 14. To see how robust the results were to our grouping procedure, we re-ran all our tests using portfolios formed by ranking first on $\hat{\sigma}$ and then on $\hat{\beta}$. We also substituted total risk for residual risk and residual variance for residual standard deviation.

We began by using monthly return data from 1954–1957 to compute security betas and sigmas. The market index employed in these computations was the CRSP equally-weighted index.[7] Within each of the five size groups (based on size as of December 1961) we then formed 16 portfolios based on ranked $\hat{\beta}_i$ and $\hat{\sigma}_i$. Data for the next four years (1958–1961) were used to re-estimate the $\hat{\beta}_i$ and $\hat{\sigma}_i$ and these were averaged across securities within each portfolio to obtain 80 initial portfolio betas ($\hat{\beta}_p$) and sigmas ($\hat{\sigma}_p$). In estimating the $\hat{\beta}_i$ and $\hat{\sigma}_i$, we used monthly instead of daily data to reduce the problems caused by thin trading.

For each month of the next year, 1962, we ran the cross-sectional regression described in (2). This yields a time series of 12 estimated gamma vectors ($\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$). The portfolio betas and sigmas are recomputed each month to reflect the actual composition of the portfolios (any security with a missing return for the month is deleted from the portfolio for that month). The individual securities $\hat{\beta}_i$ and $\hat{\sigma}_i$ used in these calculations, however, are not updated.

This procedure was repeated for each year, e.g., data from 1955–1958 was used to estimate security betas and sigmas, data from 1959–1962 was used to re-estimate portfolio betas and sigmas, and monthly cross-sectional regressions were performed for 1963. This gave us 20 estimation periods (each comprising one year), yielding a total of 240 estimates of the various gammas.[8] We checked the robustness of our results by estimating the $\hat{\beta}$s and

[7]Our initial results when using a value-weighted index were similar to those reported here. Others, such as Banz (1981) and Stambough (1982), also report that cross-sectional tests of the CAPM are not sensitive to the choice of an equally-weighted or value-weighted index.

[8]A study by Gonedes (1973) provides evidence that the mean square error of predicted returns using estimated betas is not particularly sensitive to the estimation interval used, at least for intervals varying from three to 20 years. An advantage to using a shorter period is that it reduces any bias caused by the non-constancy of the β_i and σ_i.

$\hat{\sigma}$s using two-year estimation periods and obtained essentially the same results. This reduces the possibility that our results are affected by the error in estimating the risk parameters.

The proper procedure to use in examining the effects of size, beta and residual risk on actual returns is to aggregate the estimates of each gamma derived from the 240 monthly cross-sectional regressions and then test these aggregated gammas for their statistical significance. We use two different methods to perform these aggregations, the same two described by Litzenberger and Ramaswamy (1979). The first method involves taking the simple average of estimated gammas while the second method uses a weighted average of the estimated gammas, with the weights being inversely proportional to the standard deviation of each estimate. We perform these aggregations for the entire 240-month period January 1962 through December 1981, two subperiods, 1962–1971 and 1972–1981, and the entire period but excluding the month of January (to test for the influence of the 'January effect').

Table 2 summarizes the averages of the dependent and independent variables. The difference in average estimated betas over time between the smallest and largest firms in our sample (comprised of NYSE stocks only) was 0.388 (1.116–0.728), somewhat lower than the difference of 0.576 (1.230–0.654) reported by Stoll and Whaley (1983) between the betas of the smallest and the largest firms in their sample.[9]

According to table 2, the difference in average mean excess returns $(R_p - R_f)$ between the largest and smallest firms in our sample was 0.93 percent per month. On a risk-adjusted basis (using the average betas reported for the smallest and largest size groups and taking the average excess monthly

Table 2

Summary statistics for dependent and independent variables – Averaged across time.

Group	Mean market values (in millions $)	Beta	Total standard deviation	Residual standard deviation	$R_p - R_f$ (monthly)[a]
(1)	32	1.116	0.106	0.087	1.04%
(2)	96	0.996	0.091	0.073	0.71%
(3)	220	0.925	0.086	0.069	0.46%
(4)	509	0.837	0.078	0.064	0.36%
(5)	2766	0.728	0.070	0.058	0.11%

[a]Mean excess return by size group, unadjusted for beta.

[9]This difference is probably due to the fact that the spread in mean market values in their ten size portfolios is $15–3,348 million as compared to our five portfolio spread of $32–2,766 million.

return on the equally-weighted market index of 0.56 percent), the monthly difference should have been 0.22 percent [0.56% × (1.116–0.728)]. Thus the smallest firms outperformed the largest firms by 0.71 percent (0.93%–0.22%) on a monthly basis or around 9 percent on an annual basis. This was somewhat lower than the 12 percent risk-adjusted differential found by Stoll and Whaley (1983). The discrepancy is probably due to their use of ten instead of five size portfolios, with a consequent wider spread in the size differential. The risk-adjusted differential in Reinganum's (1981) sample was about 20 percent.

Table 3 contains information on the variability of the portfolio betas and sigmas used in our analysis. For a given size category, group (i, j) refers to the portfolio containing securities with sigmas in the jth quartile among the ith quartile of betas. The group (i, j) betas and sigmas referred to in columns (1), (2) and (3) are averages, taken across all five size categories and all twenty years. Thus, for example, the group (2, 3) beta, whose value is 0.88, is the average beta found in the third sigma portfolio within the second beta portfolio. The portfolios themselves are formed by ranking first on size, then beta and finally sigma. The correlation between beta and sigma is 0.89. Columns (4), (5) and (6) are based on portfolios formed by ranking first on size, then sigma and finally beta. The portfolio betas and sigmas reported are the actual independent variables used in our January 1962 cross-sectional

Table 3

Variability of portfolio betas and sigmas.

(1) Group[a]	(2) Beta[a]	(3) Sigma[a]	(4) Group[b]	(5) Beta[b]	(6) Sigma[b]
(1,1)	0.55	0.046	(1,1)	0.18	0.032
(1,2)	0.60	0.053	(2,1)	0.44	0.037
(1,3)	0.71	0.061	(3,1)	0.75	0.044
(1,4)	0.76	0.077	(4,1)	1.12	0.041
(2,1)	0.72	0.056	(1,2)	0.30	0.054
(2,2)	0.82	0.064	(2,2)	0.73	0.050
(2,3)	0.88	0.069	(3,2)	1.13	0.052
(2,4)	1.00	0.080	(4,2)	1.61	0.056
(3,1)	0.91	0.063	(1,3)	0.68	0.065
(3,2)	0.97	0.068	(2,3)	1.09	0.064
(3,3)	1.01	0.074	(3,3)	1.23	0.065
(3,4)	1.10	0.086	(4,3)	1.63	0.065
(4,1)	1.06	0.069	(1,4)	0.88	0.105
(4,2)	1.13	0.076	(2,4)	1.32	0.089
(4,3)	1.23	0.085	(3,4)	1.71	0.101
(4,4)	1.30	0.096	(4,4)	2.15	0.116

[a]Averages across all five size categories using all 240 observations from 1962–1981.

[b]Actual independent variables used in January 1962 cross-sectional regression.

regression. The correlation between beta and sigma here is 0.67. The results reported here are representative of other months and years and indicate that there is enough variability in the portfolio betas and sigmas to enable us to distinguish between the effects of these two variables. Multicollinearity is even less of a problem in the regressions using individual securities instead of portfolios.

4. Empirical results

In table 4, we present the mean values of various estimates of the gammas, represented by

$$\bar{\hat{\gamma}}_0,\ \bar{\hat{\gamma}}_1,\ \bar{\hat{\gamma}}_2 \text{ and } \bar{\hat{\gamma}}_3 \quad \text{where} \quad \bar{\hat{\gamma}}_i = \sum_{t=1}^{n} (\hat{\gamma}_{it}/n) \tag{3}$$

and n is the number of observations, along with their t-statistics. As shown in Fama and MacBeth (1973), these t-statistics are

$$t(\bar{\hat{\gamma}}_i) = \frac{\bar{\hat{\gamma}}_i}{s(\gamma_i)/\sqrt{n}}, \tag{4}$$

where $s(\gamma_i)$ is the standard deviation of the monthly estimates of γ_i.[10] Their interpretation is subject to the caveats discussed in Fama and MacBeth (1973), which are related to the evidence that distributions of common stock returns are more closely approximated by non-normal symmetric stable distributions than by normal distributions.

Panel (a) of table 4 contains the mean values of the estimated gammas over the entire period 1962–1981 using ordinary least squares. Panel (b) is similar to panel (a) except that now the gammas are estimated using generalized least squares to take account of possible heteroscedasticity in the error terms. In panel (c), the estimated gammas are the same as those used in panel (a) but with all January observations deleted, resulting in 220 observations instead of the 240 observations used in panels (a) and (b). Panels (d) and (e) contain the results derived by dividing the 240 observations used in panel (a) into those obtained during up markets (defined as the 128 months in which $R_{mt} - R_{ft} \geqq 0$) and those obtained during down markets (the 112 months in which $R_{mt} - R_{ft} < 0$). Each panel contains the results of four separate regressions. The first three regressions use portfolio betas, sigmas and size measures, where the portfolios are based on

[10]We also performed statistical tests based on normalized coefficients, the t-values. Such tests address the problem of non-stationarity of the variance by making the weights on the observations inversely proportional to their estimation errors. The results were similar.

Table 4

Summary statistics for the regression $R_{it}=\gamma_{0t}+\gamma_{1t}\hat{\beta}_{it}+\gamma_{2t}\hat{\sigma}_{it}+\gamma_{3t}\ln(\text{size}_i/\text{size}_m)+\varepsilon_{it}$.

Regression	Order of ranking	$\bar{\hat{\gamma}}_0$	$\bar{\hat{\gamma}}_1$	$\bar{\hat{\gamma}}_2$	$\bar{\hat{\gamma}}_3$	R^2
(*a*) *Full period using OLS (240 observations)*	Size, β, σ	−0.0017 (−0.63)[a]	0.0025 (0.73)	0.0404 (0.89)	−0.0016 (−3.05)	0.27
	β, σ, size	−0.0031 (−1.20)	0.0030 (0.91)	0.0594 (1.36)	−0.0014 (−2.78)	0.27
	σ, size, β	−0.0023 (−0.88)	0.0019 (0.58)	0.0652 (1.38)	−0.0014 (−2.66)	0.27
	Individual securities	0.0017 (0.23)	0.0004 (0.18)	0.0057 (0.19)	−0.0020 (−3.75)	0.07
(*b*) *Full period using GLS to correct for heteroscedasticity (240 observations)*	Size, β, σ	−0.0026 (−0.92)	0.0034 (1.03)	0.0484 (0.98)	−0.0012 (−2.74)	0.23
	β, σ, size	−0.0033 (−1.16)	0.0045 (1.34)	0.0467 (0.99)	−0.0012 (−2.56)	0.23
	σ, size, β	−0.0029 (−1.01)	0.0028 (0.90)	0.0645 (1.33)	−0.0011 (−2.39)	0.23
	Individual securities	0.0015 (0.58)	0.0017 (0.74)	0.0400 (0.15)	−0.018 (3.13)	0.07
(*c*) *Full period using OLS but excluding January data (220 observations)*	Size, β, σ	−0.0014 (−0.53)	0.0017 (0.50)	0.0180 (0.37)	−0.0004 (−0.98)	0.26
	β, σ, size	−0.0031 (−1.18)	0.0013 (0.39)	0.0530 (1.19)	−0.0002 (−0.58)	0.26
	σ, size, β	−0.0020 (−0.73)	0.0005 (0.16)	0.0494 (1.02)	−0.0002 (−0.52)	0.25
	Individual securities	0.0025 (0.42)	−0.0012 (−0.54)	0.0008 (0.03)	−0.0007 (−1.70)	0.06
(*d*) *Up markets only (128 observations)*	Size, β, σ	−0.0002 (−0.05)	0.0278 (6.41)	0.2148 (3.25)	0.0012 (−1.68)	0.29
	β, σ, size	−0.0022 (−0.53)	0.0279 (6.96)	0.2501 (4.10)	−0.0011 (−1.47)	0.29
	σ, size, β	−0.0012 (−0.29)	0.0249 (5.86)	0.2797 (4.11)	−0.0010 (−1.34)	0.28
	Individual securities	−0.0092 (0.99)	0.0192 (7.23)	0.1717 (4.37)	−0.0024 (−3.11)	0.07
(*e*) *Down markets only (112 observations)*	Size, β, σ	−0.0033 (−1.12)	−0.0265 (−7.07)	−0.1589 (−2.83)	−0.0019 (−2.72)	0.26
	β, σ, size	−0.0041 (−1.39)	−0.0254 (−6.32)	−0.1585 (−2.82)	−0.0018 (−2.58)	0.26
	σ, size, β	−0.0036 (−1.19)	−0.0245 (−6.84)	−0.1800 (−3.18)	−0.0018 (−2.53)	0.25
	Individual securities	−0.0063 (−0.63)	−0.0211 (−8.79)	−0.1840 (−4.73)	−0.0015 (−2.12)	0.07

[a] t-statistics in parentheses.

different grouping procedures. The fourth regression uses individual securities instead of portfolios.

Several conclusions can be drawn from the data presented in table 4. Most importantly, size is the only dignificant variable. Its coefficient is also quite robust to the data or technique used. The other coefficients are not statistically significant and are quite sensitive to the estimation procedure used. In addition, the coefficient on beta is much smaller than its theoretical value, which according to the CAPM should be 0.0056 $(R_m - R_f)$.[11] These conclusions hold regardless of the grouping procedure used, whether or not one corrects for heteroscedasticity, and whether portfolios or individual securities are used in eq. (2). If January data is deleted, however, the size variable loses its statistical significance. The other coefficients also show a reduction in their significance level. Although its is not clear what this phenomenon means, it demonstrates that the return generating function in January differs from the return generating function for the other eleven months.

Turning to the results during up and down markets [panels (d) and (e) in table 4], size tends to be more important and have a larger impact in down markets. In other words, small firms seem to do relatively better when the market is going down, but whether in an up or a down market, small firms yield higher returns than do larger firms. Looking further at the data in panels (d) and (e), we see that the coefficients on beta and sigma are all statistically significant, being positive in up markets and negative in down markets. This is as expected – ex post, high-beta and high-sigma stocks do better in up markets and worse in down markets than do low-beta and low-sigma stocks. Clearly, estimated beta and estimated sigma do explain cross-sectional differences in returns among firms on a month-to-month basis. Ex ante, of course, investors do not know in which months $R_{mt} - R_{ft}$ will be positive or negative. Hence, the data in panel (a) provide the best evidence on the ex ante relative returns to bearing systematic and total risk. As stated above, judging by this data, systematic risk and total risk appear to provide little, if any, help in explaining the cross-sectional variation in average asset returns over the entire period 1962–1981. The economic significance of these results is that investors appear to have no assurance whatsoever that taking higher risks, as traditionally measured, leads to higher returns, even over relatively long periods of time.

All that being said, the coefficient on beta, regardless of the procedure employed, generally has the predicted sign. One possible interpretation of our results, therefore, is that during the 20-year period we analyzed, the

[11]We also used the procedure developed by Black and Scholes (1974), which involves regressing the time series of the gammas on the excess return of the market, to correct for the measurement error associated with using estimated instead of actual parameters in our cross-sectional regressions. The new estimate of γ_1, is 0.0045, much closer to its theoretical value of 0.0056, but still not significant at the 5 percent level. Our other conclusions hold as well.

market was unusually volatile and obscured the true effects of beta. This can only be determined by performing similar tests on other time periods.

5. Additional empirical tests

A potential problem with using (2) to test the hypothesis that residual risk is priced in addition to beta is that, according to the theoretical models developed by Levy and Mayshar, the coefficients γ_{1t} and γ_{2t} are specific to asset i, i.e., they are not constant as is assumed in (2). In particular, if the securities of small firms are more highly concentrated in relatively undiversified portfolios than are the securities of larger firms, then the Levy–Mayshar hypothesis suggests that γ_{1t} increases and γ_{2t} decreases, in both size and significance, with the market capitalization of firm i.

Both section 4's empirical results and the empirical results of Basu and Cheung (1982) however, are derived from models that constrain γ_{1t} and γ_{2t} to remain constant across firms. To see the effects of this constraint, we redid our tests, this time permitting the coefficients to vary by size. Following the approach described in section 3, we formed the same 80 portfolios used in our other tests, with the β_{pt} and σ_{pt} calculated as before. This time, however, instead of running one cross-sectional regression each month using all 80 portfolios, we ran five monthly regressions – one for each size group – on the assumption that the coefficients within each size group are roughly similar. The cross-sectional regressions are of the form

$$R_{pjt}-R_{ft}=\gamma_{0jt}+\gamma_{1jt}\hat{\beta}_{pjt}+\gamma_{2jt}\hat{\sigma}_{pjt}+\varepsilon_{pjt} \tag{5}$$

where the subscripts pjt refer to portfolio p ($p=1,\dots,16$) in size group j ($j=1,\dots,5$) in month t ($t=1,\dots,240$). This yields for each size group j a time series of 240 gamma vectors $(\bar{\gamma}_{0j}, \bar{\gamma}_{1j}, \bar{\gamma}_{2j})$.

Summary statistics for (5) are presented in table 5. These include the mean values of the gammas –

$$\bar{\hat{\gamma}}_{0j},\ \bar{\hat{\gamma}}_{1j} \text{ and } \bar{\hat{\gamma}}_{2j} \quad \text{for } j=1,\dots,5, \quad \text{where, as before}$$

$$\bar{\hat{\gamma}}_{ij}=\sum_{t=1}^{240}(\bar{\gamma}_{ijt}/240)$$

– along with their t-statistics, for each of the five size groups, with (1) being the smallest and (5) the largest size group. We also present evidence on the coefficients of two additional regressions,

$$R_{pjt}-R_{ft}=\gamma^{1}_{0jt}+\gamma^{1}_{1jt}\hat{\beta}_{pjt}+\varepsilon_{pjt}, \quad \text{and} \tag{6}$$

$$R_{pjt}-R_{ft}=\gamma^{2}_{0jt}+\gamma^{2}_{2jt}\hat{\sigma}_{pjt}+\varepsilon_{pjt}. \tag{7}$$

Table 5

Entire period (1962–1981) – 240 observations. Summary statistics for three regressions.

	$R_{pjt}-R_{ft}=\gamma_{0jt}+\gamma_{1jt}\hat{\beta}_{pjt}+\gamma_{2jt}\hat{\sigma}_{pjt}+\varepsilon_{pjt}$				$R_{pjt}-R_{ft}=\gamma^1_{0jt}+\gamma^1_{1jt}\hat{\beta}_{pjt}+\varepsilon_{pjt}$			$R_{pjt}-R_{ft}=\gamma^2_{0jt}+\gamma^2_{2jt}\hat{\sigma}_{pjt}+\varepsilon_{pjt}$		
Group	$\bar{\hat{\gamma}}_{0j}$	$\bar{\hat{\gamma}}_{1j}$	$\bar{\hat{\gamma}}_{2j}$	R^2	$\bar{\hat{\gamma}}^1_{0j}$	$\bar{\hat{\gamma}}^1_{1j}$	R^2	$\bar{\hat{\gamma}}^2_{0j}$	$\bar{\hat{\gamma}}^2_{2j}$	R^2
(1)	0.0011	0.0061	0.0286	0.23	0.0007	0.0090	0.16	0.0018	0.0978	0.14
	(0.32)[a]	(1.42)	(0.47)		(0.22)	(2.45)		(0.54)	(1.79)	
(2)	0.0017	0.0001	0.0607	0.33	0.0034	0.0038	0.23	0.0007	0.0925	0.18
	(0.54)	(0.01)	(0.92)		(1.34)	(0.97)		(0.21)	(1.51)	
(3)	0.0006	0.0089	−0.0504	0.32	−0.0019	0.0074	0.23	−0.0002	0.0772	0.24
	(−0.16)	(−1.96)	(−0.62)		(−0.64)	(1.70)		(−0.06)	(1.07)	
(4)	−0.0003	−0.0023	0.1048	0.36	−0.0003	0.0045	0.24	−0.0008	0.0851	0.25
	(−0.08)	(−0.40)	(1.00)		(−0.08)	(0.93)		(−0.20)	(1.09)	
(5)	0.0040	0.0012	−0.0829	0.29	0.0043	−0.0053	0.17	0.0036	−0.0461	0.25
	(0.90)	(0.24)	(−0.88)		(1.19)	(−1.15)		(0.81)	(−0.53)	
All firms	0.0012	0.0028	0.0122	0.31	0.0012	0.0039	0.21	0.0010	0.0613	0.20
	(1.46)	(1.37)	(0.35)		(1.04)	(1.51)		(1.28)	(2.27)	

[a] t-statistics in parentheses.

It is evident from table 5, which summarizes the results for the entire period 1962–1981, that none of the coefficients is statistically significant except for $\hat{\gamma}_{11}^1$. The importance of this finding is emphasized by the fact that over this entire period the excess return on the market, $R_{mt}-R_{ft}$, averaged 0.56 percent per month or 6.93 percent per annum. Thus, although there was a significant return to bearing market risk in the aggregate (i.e., the excess return on the market portfolio was significantly positive), the rewards on the level of the individual security do not appear to be specifically related to its degree of risk, either systematic or total. In fact, there appears to be negative compensation for bearing systematic risk for the largest firms, although γ_{15}^1 is not statistically different from 0. The only discernible pattern is that the smaller the average firm size, the larger and more statistically significant is the coefficient on sigma in eq. (7). These results indicate that multicollinearity does not explain beta's lack of statistical significance. Even when beta is the only independent variable, it is still insignificant.

The last two lines in table 5 present the mean values of the γ_j, averaged across size groups, along with their t-statistics. The only coefficient that is statistically significant at the 5 percent level is the coefficient on sigma from eq. (7), even though the coefficient is never significant for any of the individual size groups. As before, the economic significance of β is much less than predicted by the CAPM. The average coefficient on β from (6) is 0.0039 or 4.78 percent annually, approximately two-thirds its predicted value of 6.93 percent. In other words, the average excess return on a portfolio which replicated the market portfolio ($\beta=1.0$) would earn only 4.78 percent for bearing market risk.

The results for the individual size groups remain qualitatively the same when we rank first on $\hat{\sigma}_i$, when we adjust for heteroscedasticity, when $\hat{\sigma}_i^2$ is used instead of $\hat{\sigma}_i$, and when we split the sample in half taking averages of the first 120 observations and then of the second 120 observations. Substituting the total variance (standard deviation) for the residual standard deviation from the market model also leaves the qualitative results unchanged. When ranking first on $\hat{\sigma}_i$ and then on $\hat{\beta}_i$, however, the mean coefficient on σ, averaging across the entire sample, does become somewhat more statistically important, while the coefficient on β remains insignificant. This result suggests that the ranking procedure could affect the regression outcomes.

Omitting January data doesn't change our results much. The only discernible effect is the tendency for the explanatory variables to become even less important when January data is excluded. This is shown in table 6.

Despite the lack of statistical significance of the independent variables, whether run separately or jointly, including the second independent variable in the regression generally improves the amount of variation explained by about 50 percent. Specifically, the average R^2 increases from about 0.20 for

Table 6

Entire period (1962–1981) but without January data – 220 observations. Summary statistics for there regressions.

	$R_{pjt}-R_{ft}=\gamma_{0jt}+\gamma_{1jt}\hat{\beta}_{pjt}+\gamma_{2jt}\hat{\sigma}_{pjt}+\varepsilon_{pjt}$				$R_{pjt}-R_{ft}=\gamma^1_{0jt}+\gamma^1_{1jt}\hat{\beta}_{pjt}+\varepsilon_{pjt}$			$R_{pjt}-R_{ft}=\gamma^2_{0jt}+\gamma^2_{2jt}\hat{\sigma}_{pjt}+\varepsilon_{pjt}$		
Group	$\bar{\hat{\gamma}}_{0j}$	$\bar{\hat{\gamma}}_{1j}$	$\bar{\hat{\gamma}}_{2j}$	R^2	$\bar{\hat{\gamma}}^1_{0j}$	$\bar{\hat{\gamma}}^1_{1j}$	R^2	$\bar{\hat{\gamma}}^2_{0j}$	$\bar{\hat{\gamma}}^2_{2j}$	R^2
(1)	−0.0006	0.0049	−0.0133	0.22	−0.0022	0.0055	0.15	−0.0001	0.0435	0.14
	(−0.18)[a]	(1.10)	(−0.22)		(−0.77)	(1.48)		(−0.03)	(0.80)	
(2)	0.0010	0.0005	0.0096	0.32	0.0012	0.0017	0.23	0.0001	0.0406	0.18
	(0.30)	(0.10)	(0.14)		(0.46)	(0.43)		(0.04)	(0.66)	
(3)	−0.0032	0.0074	0.0434	0.31	−0.0042	0.0067	0.23	−0.0022	0.0614	0.23
	(−0.82)	(1.68)	(−0.56)		(−1.34)	(1.49)		(−0.60)	(0.83)	
(4)	−0.0028	−0.0050	−0.1507	0.35	−0.0014	0.0038	0.24	−0.0032	0.0909	0.24
	(−0.70)	(−0.83)	(−1.37)		(−0.44)	(0.78)		(−0.85)	(1.12)	
(5)	0.0022	−0.0007	−0.0382	0.29	−0.0041	−0.0057	0.17	−0.0015	−0.0313	0.19
	(0.47)	(−0.14)	(−0.40)		(1.13)	(−1.23)		(0.35)	(−0.36)	
All firms	−0.0007	0.0014	0.0131	0.30	−0.0005	0.0024	0.20	−0.0008	0.0410	0.20
	(−0.65)	(0.65)	(0.37)		(−0.35)	(1.10)		(−0.92)	(2.04)	

[a] t-statistics in parentheses.

those regressions containing only $\hat{\beta}_{pjt}$ or $\hat{\sigma}_{pjt}$ to about 0.30 for those regressions including both independent variables.

The likely explanation for the relatively high average R^2 but minimal significance of the independent variables is that the $\hat{\gamma}_{ij}$ are averages of gammas calculated during down markets ($R_{mt}-R_{ft}<0$) and up markets ($R_{mt}-R_{ft}>0$). Based on eq. (2), we would expect $\hat{\gamma}_{ijt}<0$ when $R_{mt}-R_{ft}<0$ and $\hat{\gamma}_{ijt}>0$ when $R_{mt}-R_{ft}>0$, for $i=1, 2$. Taking simple averages of the $\hat{\gamma}_{ijt}$ over the entire period yields average values of the $\hat{\gamma}_{ij}$ close to zero and with large variances. Thus, estimates of the returns to risk bearing will be very noisy. We tested this hypothesis by dividing the 240 observations used in table 5 into those obtained during up markets and those obtained during down markets. Although we do not report our results here, they support our hypothesis.

6. Summary and conclusions

Banz (1981) and Reinganum (1981) have presented strong evidence that small firms earn abnormally high risk-adjusted returns in comparison to large firms. One possibile explanation for this small firm effect, implied by the work of Levy (1978) and Mayshar (1979, 1981, 1983), is that the existence of various transaction costs results in shares of small firms being held in portfolios that on average are relatively undiversified. This lack of diversification, in turn, requires that investors in small firms be compensated for bearing total risk rather than systematic risk. This theory, however, fails to account for the possibility of arbitrage on the part of the marginal investor.

Using monthly data from the period 1962–1981 and a variety of procedures to examine the relationship between return and various measures of risk, we tested the Levy–Mayshar hypothesis, which is so at odds with the Sharpe–Lintner capital asset pricing model. The examination of the various results tells us something about the robustness of our conclusions. We used four-year and two-year periods to estimate our betas, standard deviations of the residuals and total variances. We used total risk as well as its systematic and unsystematic components in our cross-sectional regressions. We ran these regressions with three explanatory variables (size, beta and either total or unsystematic risk), with two explanatory variables within each size group (beta and either residual or total risk), and with one explanatory variable within each size group (beta, total risk or unsystematic risk). We employed all possible portfolio grouping procedures, ranking on size, beta and residual or total risk. In pooling the data over time, we assumed both homoscedastic and heteroscedastic errors. We aggregated the resulting coefficients for the entire period, for two subperiods, with and without January data, and for up and down markets. In aggregating these coefficients, we gave equal weights to each of the cross-sectional gammas as well as weights related to their estimation procedures. We also performed these analyses without grouping, at the level of the individual security.

The results can be more easily summarized than the techniques used to derive them: they reject the implication of the Levy–Mayshar hypothesis that total risk, as opposed to systematic risk, is more important for small firms. Unfortunately for modern capital market theory, these results also reject – at standard levels of statistical significance – the fundamental tenet of the CAPM, that beta matters. Our conclusion is that neither the traditional measure of risk (beta) nor the alternative risk measures (variance or residual standard deviation) can explain – again, at standard levels of statistical significance – the cross-sectional variation in returns; only size appears to matter. It may be that 20 years is too short a time period in which to perform our tests, especially if the period selected was unusually volatile. It may also be that standard levels of statistical significance are not applicable when conducting tests with such important consequences. This is a particularly relevant consideration here since the coefficient on beta generally has the correct sign. Regardless of these speculations, however, the small firm effect is still a puzzle.

References

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

Basu, Joe and Sherm Cheung, 1982, Residual risk, firm size and returns for NYSE common stocks: Some empirical evidence, Working paper, Jan. (McMaster University, Hamilton).

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

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

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

Friend, Irwin W. and Randolph Westerfield, 1981, Risk and capital asset prices, Journal of Banking and Finance, Sept., 291–315.

Gonedes, Nicholas, 1973, Evidence on the information content of accounting numbers: Accounting-based and market-based estimates of systematic risk, Journal of Financial and Quantitative Analysis, June, 407–444.

Keim, Donald B., 1983, Size-related anomalies and stock market seasonality: Further empirical evidence, Journal of Financial Economics 12, 13–32.

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.

Lakonishok, Josef and Seymour Smidt, 1984, Volume and turn-of-the-year behavior, Journal of Financial Economics 13, Sept., 435–455.

Levy, Haim, 1978, Equilibrium in an imperfect market: A constraint on the number of securities in the portfolio, American Economic Review, Sept., 643–658.

Lintner, John, 1969, The aggregation of investors' diverse judgments and preferences in purely competitive security markets, Journal of Financial and Quantitative Analysis, Dec., 347–400.

Litzenberger, Robert 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, 114–164.

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

Mayshar, Joram, 1981, Transaction costs and the pricing of assets, Journal of Finance, June, 583–597.

Mayshar, Joram, 1983, On divergence of opinion and imperfections in capital markets, American Economic Review, March, 114–128.

Reinganum, Marc R., 1981, Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values, Journal of Financial Economics 9, March, 19–46.

Reinganum, Marc R., 1983, The anomalous stock market behavior of small firms in January, Journal of Financial Economics 12, June, 89–104.

Roll, Richard, 1983, The turn-of-the-year effect and the return premium of small firms, Journal of Portfolio Management, Winter, 18–28.

Stambough, Robert F., 1982, On the exclusion of assets from tests of the two-partners model: A sensitivity analysis, Journal of Financial Economics 11, Nov., 237–268.

Stoll, Hans R. and Robert E. Whaley, 1983, Transaction costs and the small firm effect, Journal of Financial Economics 12, June, 57–80.

Vermaelen, Theo, 1981, Common stock repurchases and market signalling: An empirical study, Journal of Financial Economics 9, June, 139–183.