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.

130140-OPC-POD-57-184

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, $b = [E(R_m) - r_f - \tau(d_m - r_f)]$ 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.

130140-OPC-POD-57-185

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 $\theta^k \equiv -E(u')/E(u'')$ 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:

$$\begin{aligned} \mathcal{L}^k \equiv{} & f^k(\mu_k, \sigma_k^2) + \lambda_1^k\left[1 - \sum_i X_i^k - X_f^k\right] \\ & + \lambda_2^k\left[\sum_i X_i^k d_i + X_f^k r_f - S_2^k\right] + \lambda_3^k\left[(1-\alpha)\sum_i X_i^k + X_f^k - S_3^k\right], \end{aligned} \tag{10}$$

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:

$$\begin{aligned} \frac{\partial \mathcal{L}^k}{\partial X_i^k} ={} & f_1^k\{E(\tilde{R}_i) - [t^k + Y_1^k g'(Y_1^k)]d_i\} - \lambda_1^k + \lambda_2^k d_i \\ & + \lambda_3^k(1-\alpha) + 2f_2^k \sum_j X_j^k \operatorname{cov}(\tilde{R}_i, \tilde{R}_j) = 0, \qquad i = 1, 2, \ldots, N, \end{aligned} \tag{11}$$

$$\frac{\partial \mathcal{L}^k}{\partial X_f^k} = f_1^k\{r_f - [t^k + Y_1^k g'(Y_1^k)]r_f\} - \lambda_1^k + \lambda_2^k r_f + \lambda_3^k = 0, \tag{12}$$

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

$$\beta_i \equiv \operatorname{cov}(\tilde{R}_i, \tilde{R}_m)/\operatorname{var}(\tilde{R}_m),$$

$$a \equiv \alpha \sum_k (\theta^k/\theta^m)(\lambda_3^k/f_1^k),$$

$$b \equiv \operatorname{var}(\tilde{R}_m)(W^m/\theta^m),$$

$$c \equiv \sum_k (\theta^k/\theta^m)[T^k - (\lambda_2^k/f_1^k)],$$

$$\theta^m \equiv \sum_k \theta^k.$$

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, $\lambda_3^k > 0$ 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, ($\lambda_3^k = 0$ 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 $b = [E(R_m) - r_f - c(d_m - r_f) - a]$. 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 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 $k \in N$, $\lambda_2^k = 0$, $Y_1^k \geq 0$ and $T^k \geq t^k \geq 0$. 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

$$\tilde{R}_{it}-r_{ft}=\gamma_0+\gamma_1\beta_{it}+\gamma_2(d_{it}-r_{ft})+\tilde{\varepsilon}_{it}, \qquad i=1,2,\dots,N_t, \quad t=1,2,\dots,T, \tag{23}$$

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, $\tilde{\varepsilon}\equiv\{\tilde{\varepsilon}_{it}: i=1,2,\dots,N_t,\ t=1,\dots,T\}$, 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 '{ · }' 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 i, $i=1,2,\dots,N_t$, securities in each period t, $t=1,\dots,T$. 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

$$\bar{R}_t \equiv \{(\bar{R}_{1t} - r_{ft})(\bar{R}_{2t} - r_{ft})(\bar{R}_{it} - r_{ft}), \ldots, (\bar{R}_{N_t t} - r_{ft})\},$$

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 [illegible] 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:

$$\tilde{R}_{it} = \alpha_{it} + \beta_{it}\tilde{R}_{mt} + \tilde{\varepsilon}_{it}, \qquad i = 1, 2, \ldots, N_t, \tag{34}$$

$$\begin{aligned} \operatorname{cov}(\tilde{\varepsilon}_{it}, \tilde{\varepsilon}_{jt}) &= 0, \qquad i \neq j, \\ &= s_{ii}, \qquad i = j, \end{aligned} \tag{35}$$

$$\alpha_{it} = E(\tilde{R}_{it} \mid R_{mt} = 0).$$

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

$$\begin{aligned} V_t(i,j) &= \beta_{it}\beta_{jt}\sigma_{mm}, \qquad i \neq j, \\ &= \beta_{it}^2\sigma_{mm} + s_{ii}, \qquad i = j, \end{aligned} \qquad i, j = 1, 2, \ldots, N_t, \tag{36}$$

[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

$$\begin{aligned} \Omega_t(i,j) &= 0, \qquad i \neq j, \\ &= s_{ii}, \qquad i = j, \end{aligned} \qquad i, j = 1, 2, \ldots, N_t. \tag{38}$$

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

$$\begin{aligned} P_t(i,j) &= \phi/s_i \equiv \phi/\sqrt{s_{ii}}, \qquad i = j \\ &= 0, \qquad i \neq j, \end{aligned} \tag{39}$$

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 $E(\tilde{e}_{i\tau}) = 0$ 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 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, $s_i = (\mathrm{var}(\tilde{e}_{it}))^{\frac{1}{2}}$. 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

---

[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 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, $\tilde{R}_{it} - r_{ft}$ was regressed on $\tilde{R}_{mt} - r_{ft}$ to estimate β_{it} as explained in section 4 below. Thus in the computation in (43), $(\tilde{R}_{mt} - r_{ft} - \overline{R_m - r_f})^2$ is used in place of $(\tilde{R}_{mt} - \bar{R}_m)^2$.

$$\hat{\Gamma} = \sum_{t=1}^{T} (\hat{\Gamma}_t/T), \tag{46}$$

where

$$\hat{\Gamma}_t = (\hat{X}_t^{*\prime} \hat{X}_t^{*})^{-1} \hat{X}_t^{*\prime} \tilde{R}_t^{*} \tag{47}$$

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:

$$\hat{\Gamma}_t = \left(\frac{\hat{X}_t^{*\prime} \hat{X}_t^{*}}{N_t} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right)^{-1} \frac{\hat{X}_t^{*\prime} R_t^{*}}{N_t}. \tag{49}$$

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

$$\operatorname*{plim}_{N_t} \left(\frac{\hat{X}_t^{*\prime} \hat{X}_t^{*}}{N_t} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right)^{-1} \frac{\hat{X}_t^{*\prime} X_t^{*}}{N_t} = I, \tag{52}$$

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,

$$\tilde{R}_{i\tau} - r_{f\tau} = \alpha_{it} + \beta_{it}(\tilde{R}_{m\tau} - r_{f\tau}) + \tilde{e}_{i\tau}, \quad \tau = t-60, t-59, \ldots, t-1. \tag{55}$$

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),

$$\tilde{R}_{it} - r_{ft} = \gamma'_0 + \gamma'_1 \beta_{it} + \tilde{u}_{it}, \qquad i=1,2,\ldots,N_t, \quad t=1,2,\ldots,504, \tag{56}$$

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,

$$\hat{\gamma}_k = \sum_{t=1}^{504} \hat{\gamma}_{kt} \Big/ 504, \qquad k=0,1,2, \tag{57}$$

$$\operatorname{var}(\hat{\gamma}_k) = \sum_{t=1}^{504} (\hat{\gamma}_{kt} - \hat{\gamma}_k)^2/(504 \cdot 503). \tag{58}$$

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 $j=1,2,3$.

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, 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.

[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.

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.

130140-OPC-POD-57-195

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

$$\tilde R_{it} - r_{ft} = \gamma_0 + \gamma_1\beta_{it} + \gamma_2(d_{it} - r_{ft}) + \tilde\epsilon_{it}, \quad i = 1, 2, \ldots, N_t, \; t = 1, 2, \ldots, T.$$

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:

$$\tilde R_{it} - r_{ft} = \gamma_0 + \gamma_1\beta_{it} + \gamma_2\{\delta_{it} d^0_{it} - r_{ft}\} + \gamma_3\{(1-\delta_{it})d^0_{it}\} + \tilde\epsilon_{it}, \quad i = 1, 2, \ldots, N_t \qquad (59)$$

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

$$\tilde R_{it} - r_{ft} = \gamma_0 \; \gamma_1\beta_{it} + \gamma_2\{\delta_{it} d^0_{it} - r_{ft}\} + \gamma_3(1-\delta_{it})d^0_{it} + \tilde\epsilon_{it}, \quad i = 1, 2, \ldots, N_t, \; t = 1, 2, \ldots, T.$$

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

$$c_i=k-hd_i, \qquad (60)$$

where $k, h>0$, and the hypothesized relationship is

$$E(\tilde{R}_i)-r_f=a+b\beta_i+(k-hd_i)(d_i-r_f). \qquad (61)$$

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 $(d^e_{it}/3)$ 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 γ_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.

$$\tilde{R}_{it}-r_{ft}=\gamma_0+\gamma_1\beta_{it}+\gamma_2(d_{it}-r_{ft})+\gamma_4 d_{it}(d_{it}-r_{ft})+\tilde{\varepsilon}_{it}, \quad i=1,2,\ldots,N_t, \qquad (62)$$

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:

$$\tilde{R}_{it}-r_{ft}=\gamma_0+\gamma_1\beta_{it}+\gamma_2(d_{it}-r_{ft})+\gamma_4 d_{it}(d_{it}-r_{ft})+\tilde{\varepsilon}_{it} \quad i=1,2,\ldots,N_t, \quad t=1,2,\ldots,T.$$

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 $\{h_{it}; i=1,2,\dots,N_t\}$ are chosen in each period, for investment in assets $i=1,2,\dots,N_t$. Using eq. (23) from the text the excess return on such a portfolio is given by

$$\sum_i h_{it}(\tilde{R}_{it} - r_{ft}) = \gamma_0\Big(\sum_i h_{it}\Big) + \gamma_1\Big(\sum_i h_{it}\hat{\beta}_{it}\Big) + \gamma_2\Big[\sum_i h_{it}(d_{it} - r_{ft})\Big] + \sum_i h_{it}\tilde{\varepsilon}_{it}.$$

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,

$$\operatorname{var}\Big(\sum_i h_{it}(\tilde{R}_{it} - r_{ft})\Big) = \Big(\sum_i h_{it}\beta_{it}\Big)^2 \sigma_{mm} + \sum_i h_{it}^2 s_{it}.$$

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

$$\Big(\sum_i h_{it}\beta_{it}\Big)^2 \sigma_{mm} + \sum_i h_{it}^2 s_{it},$$

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

$$\tilde{R}^*_{it} = \gamma_0 p^*_{it} + \gamma_1 \beta^*_{it} + \gamma_2 d^*_{it} + \tilde{\varepsilon}^*_{it}, \qquad \text{(B.1)}$$

and the observed beta as

$$\tilde{\beta}^*_{it} = \beta^*_{it} + \tilde{v}^*_{it}, \qquad \text{(B.2)}$$

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

$$m_{xy} = \sum_{i=1}^{N_t} x_{it} y_{it} \Big/ N_t, \qquad \text{(B.3)}$$

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),

$$m_{R^* p^*} = \gamma_0 m_{p^* p^*} + \gamma_1 m_{\beta^* p^*} + \gamma_2 m_{d^* p^*} + m_{\varepsilon^* p^*}, \qquad \text{(B.4)}$$

$$m_{R^* \tilde{\beta}^*} = \gamma_0[m_{p^* \beta^*} + m_{p^* v^*}] + \gamma_1[m_{\beta^* \beta^*} + m_{v^* \beta^*}] + \gamma_2[m_{d^* \beta^*} + m_{d^* v^*}] + m_{\varepsilon^* \beta^*} + m_{\varepsilon^* v^*}, \qquad \text{(B.5)}$$

$$m_{R^* d^*} = \gamma_0 m_{p^* d^*} + \gamma_1 m_{\beta^* d^*} + \gamma_2 m_{d^* d^*} + m_{\varepsilon^* d^*}, \qquad \text{(B.6)}$$

$$m_{\tilde{\beta}^* p^*} = m_{\beta^* p^*} + m_{v^* p^*}, \qquad \text{(B.7)}$$

$$m_{\tilde{\beta}^* \tilde{\beta}^*} = m_{\beta^* \beta^*} + 2m_{\beta^* v^*} + m_{v^* v^*}, \qquad \text{(B.8)}$$

$$m_{d^* \tilde{\beta}^*} = m_{d^* \beta^*} + m_{d^* v^*}. \qquad \text{(B.9)}$$

In these six equations, take expectations and use the fact that

$$E(\tilde{v}^*_{it}) = E(\tilde{\varepsilon}^*_{it}) = 0,$$

$$E(\tilde{v}^*_{it}\tilde{\varepsilon}^*_{it}) = 0, \qquad \text{(B.10)}$$

$$E(\tilde{v}^*_{it}\tilde{v}^*_{it}) = E[\tilde{v}^2_{it}/s^2_i] = 1.$$

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 v_{it} and ε_{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.

$$m_{R^* p^*} = \gamma_0 m_{p^* p^*} + \gamma_1 m_{\tilde{\beta}^* p^*} + \gamma_2 m_{d^* p^*}, \qquad \text{(B.11)}$$

$$m_{R^* \tilde{\beta}^*} = \gamma_0 m_{p^* \tilde{\beta}^*} + \gamma_1 (m_{\tilde{\beta}^* \tilde{\beta}^*} - 1) + \gamma_2 m_{d^* \tilde{\beta}^*}, \qquad \text{(B.12)}$$

$$m_{R^* d^*} = \gamma_0 m_{p^* d^*} + \gamma_1 m_{\tilde{\beta}^* d^*} + \gamma_2 m_{d^* d^*}. \qquad \text{(B.13)}$$

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 $m_{R^* R^*}$.

$$m_{\tilde{R}^* \tilde{R}^*} = \gamma_0^2 m_{p^* p^*} + \gamma_1^2 m_{\beta^* \beta^*} + \gamma_2^2 m_{d^* d^*} + 2\gamma_0\gamma_1 m_{p^* \beta^*} + 2\gamma_0\gamma_2 m_{p^* d^*} + 2\gamma_1\gamma_2 m_{\beta^* d^*} + 2\gamma_0 m_{p^* \varepsilon^*} + 2\gamma_1 m_{\beta^* \varepsilon^*} + 2\gamma_2 m_{d^* \varepsilon^*} + m_{\varepsilon^* \varepsilon^*}. \qquad \text{(B.14)}$$

If we take expectations, using (B.10) and the fact that

$$E(m_{g^*g^*}) = E\left(\sum_{i=1}^{N_t} \frac{\tilde{\varepsilon}_{it}^2}{\sigma_i^2 N_t}\right)$$

$$= \frac{1}{N_t}\sum_{i=1}^{N_t} \frac{E(\tilde{\varepsilon}_{it}^2)}{\sigma_i^2} = \frac{1}{N_t}\cdot N_t\phi^2 = \phi^2,$$

we have

$$E(m_{g^*g^*}) = \gamma_0^2 m_{\beta^*\beta^*} + \gamma_1^2 m_{\beta^*\beta^*} + \gamma_2^2 m_{d^*d^*} + 2\gamma_0\gamma_1 m_{\beta^*\beta^*} + 2\gamma_0\gamma_2 m_{\beta^*d^*} + 2\gamma_1\gamma_2 m_{\beta^*d^*} + \phi^2, \quad \text{(B.15)}$$

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).