Extra-Market Components of Covariance Among Security Prices

Barr Rosenberg

University of California, Berkeley

Michel Houglet

University of California, Berkeley

Vinay Marathe

University of California, Berkeley

1973

Research Program in Finance
Working Paper RPF-013

EXTRA-MARKET COMPONENTS OF COVARIANCE AMONG SECURITY PRICES

Barr Rosenberg, Michel Houglet, and Vinay Marathe*

I. Introduction

In a pioneering study [4], King concluded that there are significant covariances among securities beyond those attributable to an overall market factor. In the small sample that King studied, the pattern of these covariances closely corresponded to industry groupings. Despite this initial effort, the magnitude of extra-market covariance remains uncertain. Nevertheless, portfolio managers continue to diversify across industry groups and across economic and financial characteristics in the hope of reducing portfolio variance.

In Section II, below, the relationship between a multiple-factor model of security returns and the "market" model is explored. Then, in Section III, a new method of estimating extra-market covariance is proposed that avoids some of the limitations of earlier methods by relying on the assumption that the factor loadings of individual security returns are linear functions of measurable characteristics of the firm.

---

*The authors are Assistant Professor and graduate students, respectively, at the School of Business Administration, University of California, Berkeley. We gratefully acknowledge helpful discussions with William F. Sharpe, Walt McKibben, William Fouse, William Jahnke, and Robert Jackson. The development of Section II parallels Sharpe's treatment of a related problem in [14], and has particularly benefited from his contributions. All errors, of course, are our own responsibility. This research was supported by NSF Grant GS 3306.

The method is applied to an extensive data base, explained in Section IV, and the more interesting results are reported in Section V. A concluding discussion completes the paper.

II. A Multiple Factor Model of Security Returns

Let r_{nt} denote the return on security n in period t, $n=1,, N_t$, $t=1,, T$. Let r_{Ft} denote the risk-free return in period t. Let $z_{nt} = r_{nt} - r_{Ft}$ denote a return in excess of the risk-free rate. Then consider the model

$$(1) \qquad z_{nt} = \mu + \sum_{j=1}^{J} \lambda_{jnt} f_{jt} + u_{nt} = \underset{\sim}{\lambda}'_{nt} \underset{\sim}{f}_t + u_{nt} \qquad \begin{array}{l} n=1,,N_t \\ t=1,,T \end{array}$$

where $\underset{\sim}{f}_t = (f_{1t} \vdots f_{2t} \vdots \ldots \vdots f_{Jt})'$ is a column vector of the values of J factors in period t, and where $\underset{\sim}{\lambda}_{nt} = (\lambda_{1nt} \vdots \lambda_{2nt} \vdots \ldots \vdots \lambda_{Jnt})'$ is the column vector of "loadings" of the return r_{nt} onto these factors. The random terms u_{nt} introduce risk that is specific to the return on security n in period t, and are therefore assumed to be uncorrelated across securities and over time, with mean equal to zero and variance equal to σ^2_{nt}. Notice that if $\underset{\sim}{z}_t$ is the column vector of returns on all securities in period t, $\underset{\sim}{z}_t = (z_{1t} \vdots z_{2t} \vdots \ldots \vdots z_{N_t t})'$, if $\underset{\sim}{u}_t = (u_{1t} \vdots u_{2t} \vdots \ldots \vdots u_{N_t t})'$, and if $\underset{\sim}{\Lambda}_t = (\underset{\sim}{\lambda}_{1t} \vdots \underset{\sim}{\lambda}_{2t} \vdots \ldots \vdots \lambda_{N_t t})'$, then the model for all returns in period t may be written in matrix form as

$$(1a) \qquad \underset{\sim}{z}_t = \mu \underset{\sim}{\iota} + \underset{\sim}{\Lambda}_t \underset{\sim}{f}_t + \underset{\sim}{u}_t \;,$$

where $E(\underset{\sim}{u}_t) = \underset{\sim}{0}$, $E(\underset{\sim}{u}_t\underset{\sim}{u}_t') = \text{Diag}(\sigma^2_{nt}) \equiv \underset{\sim}{\Sigma}_t$, and where $\underset{\sim}{\iota} = (1{:}1{:} \,..\, {:}\,1)'$ is a column vector of units. This model was introduced by King [4: Secs. I and II), in slightly different notation.

The J factors represent random events that may influence the returns on several securities, such as a change in inflationary expectations, a change in risk aversion on the part of investors, or a change in the prospects of some industry. The factors are assumed to be serially uncorrelated and to be uncorrelated with the specific risk components. For simplicity, the distribution of the factors will be assumed to be stationary, although evidence reported below strongly suggests that this is not the case. Let $\underset{\sim}{\overline{f}} \equiv E(\underset{\sim}{f}_t)$ and $\underset{\sim}{F} \equiv VAR(\underset{\sim}{f}_t)$ be the stationary mean vector and variance matrix of the factors. Then the mean vector and variance matrix of securities returns are

$$(2) \qquad E(\underset{\sim}{z}_t) = \mu\underset{\sim}{\iota} + \underset{\sim}{\Lambda}_t\underset{\sim}{\overline{f}} \quad , \quad VAR(\underset{\sim}{z}_t) = \underset{\sim}{\Lambda}_t\underset{\sim}{F}\underset{\sim}{\Lambda}_t' + \underset{\sim}{\Sigma}_t \; .$$

Thus the variance matrix of returns in period t is the sum of the contribution of the factors, a matrix of rank J, plus the contribution of specific risk, a diagonal matrix.

Let the "market" return be defined by $r_{Mt} \equiv \sum_n c_{nt}r_{nt} = \underset{\sim}{c}_t'\underset{\sim}{r}_t$, where $\underset{\sim}{c}_t = (c_{1t}{:}\; c_{2t}{:} \,..\, {:}\; c_{N_t t})'$ is a vector of weights satisfying $\sum_n c_{nt} = 1$. Let $z_{Mt} = r_{Mt} - r_{Ft}$ be the excess return on the market. Then, from (1a),

$$(3) \qquad z_{Mt} = \underset{\sim}{c}_t'\underset{\sim}{z}_t = \mu + \underset{\sim}{c}_t'\underset{\sim}{\Lambda}_t\underset{\sim}{f}_t + \underset{\sim}{c}_t'\underset{\sim}{u}_t = \mu + \underset{\sim}{a}_t'\underset{\sim}{f}_t + \underset{\sim}{c}_t'\underset{\sim}{u}_t \; ,$$

where $\underset{\sim}{a}_t = \underset{\sim}{\Lambda}_t'\underset{\sim}{c}_t$ is the vector of loadings of the market onto the factors.

On the assumption that the weights in $\underset{\sim}{c}_t$ are similar to $1/N_t$, the variance of the last term is of order $1/N_t$; if N_t is large, this term may be neglected, yielding

$$z_{Mt} \simeq \mu + \underset{\sim}{a}_t'\underset{\sim}{f}_t \ . \tag{3a}$$

As is customary in the literature on the Capital Asset Pricing Model, the approximation (3a) will be used to simplify subsequent derivations. (There would be no significant change if the exact expression (3) were used instead.) The moments of the market return are easily derived in terms of the moments of the factors:

$$E(z_{Mt}) = \mu + \underset{\sim}{a}_t'\overline{\underset{\sim}{f}} \ , \quad VAR(z_{Mt}) = \underset{\sim}{a}_t'\underset{\sim}{F}\underset{\sim}{a}_t \ , \quad COV(\underset{\sim}{f}_t, z_{Mt}) = \underset{\sim}{F}\underset{\sim}{a}_t \ . \tag{4}$$

Moreover, the covariance between an individual security return and the market is

$$COV(z_{nt}, z_{Mt}) = \underset{\sim}{\lambda}_{nt}'\underset{\sim}{F}\underset{\sim}{a}_t \quad , \quad n=1,,N_t. \tag{5}$$

Consequently, the regression coefficient of the individual security return on the market return is

$$\beta_{nt} \equiv COV(z_{nt}, z_{Mt})/VAR(z_{Mt}) = \underset{\sim}{\lambda}_{nt}'\underset{\sim}{F}\underset{\sim}{a}_t/(\underset{\sim}{a}_t'\underset{\sim}{F}\underset{\sim}{a}_t) = \underset{\sim}{\lambda}_{nt}'\underset{\sim}{\theta}_t \ , \tag{6}$$

where $\underset{\sim}{\theta}_t = \underset{\sim}{F}\underset{\sim}{a}_t/(\underset{\sim}{a}_t'\underset{\sim}{F}\underset{\sim}{a}_t)$. Thus, the security beta is a linear combination of the factor loadings for that security, with weights $\underset{\sim}{\theta}_t = (\theta_{1t} : \theta_{2t} : .. : \theta_{Jt})'$, where, from (4), $\theta_{jt} = COV(f_{jt}, z_{Mt})/VAR(z_{Mt})$ is the regression coefficient of factor j onto the market return.

Equilibrium in the basic one-period Capital Asset Pricing Model with unlimited borrowing and lending at the risk-free rate implies the widely known condition that $E(z_{nt}) = \beta_{nt}E(z_{Mt})$ for all n (see, for example, Lintner [5] and Sharpe [13]). For the multiple-factor model, this condition becomes

$$(7) \qquad E(z_{nt}) = \mu + \lambda_{nt}'\bar{f} = \left(\lambda_{nt}'Fa_t/(a_t'Fa_t)\right)(\mu + a_t'\bar{f}) = \beta_{nt}E(z_{Mt}), \quad \text{for all } n.$$

After this requirement has been restated as

$$\mu + \lambda_{nt}'\bar{f} = \lambda_{nt}'\left(\frac{Fa_t\mu + Fa_ta_t'\bar{f}}{a_t'Fa_t}\right),$$

it is easily shown that if this is to be satisfied for all possible loading vectors λ_{nt}, then it is necessary and sufficient that

$$(8) \qquad \mu = 0 \quad \text{and} \quad \bar{f} = kFa_t$$

for some constant k. This constant may be interpreted as the price of risk, since when $\mu = 0$,

$$E(z_{Mt}) = a_t'\bar{f} = a_t'(kFa_t) = k\mathrm{VAR}(z_{Mt}) \rightarrow k = E(z_{Mt})/\mathrm{VAR}(z_{Mt}).$$

Thus the Capital Asset Pricing Model imposes a very sharp restriction on the mean values of the factors, namely that these are proportional to the covariances of the factors with the market return. Substitution of condition (8) on $\bar{f}$ into equation (6) for beta yields

$$\beta_{nt} = \underset{\sim}{\lambda}'_{nt}\underset{\sim}{\theta}_t = \underset{\sim}{\lambda}'_{nt}\left(\bar{\underset{\sim}{f}}/E(z_{Mt})\right). \tag{9}$$

Thus the weights of the factor loadings in the equation for beta are equal to the mean values for the factors divided by the excess return for the market. When the condition that $\mu = 0$ is resubstituted into the original multiple-factor model (1), the model may be rewritten in the equivalent form

$$z_{nt} = \underset{\sim}{\lambda}'_{nt}\underset{\sim}{f}_t + u_{nt} = \underset{\sim}{\lambda}'_{nt}\left(\underset{\sim}{\theta}_t\underset{\sim}{a}'_t + (\underset{\sim}{I}-\underset{\sim}{\theta}_t\underset{\sim}{a}'_t)\right)\underset{\sim}{f}_t + u_{nt}$$

$$= \underset{\sim}{\lambda}'_{nt}\underset{\sim}{\theta}_t z_{Mt} + \underset{\sim}{\lambda}'_{nt}\underset{\sim}{g}_t + u_{nt}, \text{ where } \underset{\sim}{g}_t = (\underset{\sim}{I}-\underset{\sim}{\theta}_t\underset{\sim}{a}'_t)\underset{\sim}{f}_t = \underset{\sim}{f}_t-\underset{\sim}{\theta}_t z_{Mt}$$

$$= \beta_{nt}z_{Mt} + \underset{\sim}{\lambda}'_{nt}\underset{\sim}{g}_t + u_{nt} \quad , \quad n = 1,,N_t; \quad t=1,,T. \tag{10}$$

Thus the model has been rewritten to include the familiar contribution of the market, $\beta_{nt}z_{Mt}$, the familiar contribution of specific risk, u_{nt}, and an additional term $\underset{\sim}{\lambda}'_{nt}\underset{\sim}{g}_t$, where $\underset{\sim}{g}_t = (g_{1t} : g_{2t} : .. : g_{Jt})'$ is a vector of transformed factors. Each of these transformed factors is equal to the original factor minus the prediction for that factor based on market return (or, in other words, is equal to the component of that factor that is uncorrelated with or "orthogonal to" the market), for

$$g_{jt} = f_{jt} - \theta_{jt}z_{Mt} = f_{jt} - \left(\frac{COV(f_{jt},z_{Mt})}{VAR(z_{Mt})}\right)z_{Mt}. \tag{11}$$

Thus the transformed factors g_j may be thought of as "residual" factors. From condition (8) on the mean values of the factors,

$$E(\underset{\sim}{g}_t) = E\left((\underset{\sim}{I}-\underset{\sim}{\theta}_t\underset{\sim}{a}'_t)\underset{\sim}{f}_t\right) = \left(\underset{\sim}{I} - (\underset{\sim}{F}\underset{\sim}{a}_t\underset{\sim}{a}'_t)/(\underset{\sim}{a}'_t\underset{\sim}{F}\underset{\sim}{a}_t)\right)k\underset{\sim}{F}\underset{\sim}{a}_t = \underset{\sim}{0} \,. \tag{12}$$

By construction,

$$\text{(13)} \qquad \mathrm{COV}(\mathbf{g}_t, z_{Mt}) = E\Big(\big(\mathbf{I} - (\mathbf{F}\mathbf{a}_t\mathbf{a}_t')/(\mathbf{a}_t'\mathbf{F}\mathbf{a}_t)\big)(\mathbf{f}_t-\bar{\mathbf{f}})(\mathbf{f}_t-\bar{\mathbf{f}})'\mathbf{a}_t\Big)$$

$$= \big(\mathbf{I} - (\mathbf{F}\mathbf{a}_t\mathbf{a}_t')/(\mathbf{a}_t'\mathbf{F}\mathbf{a}_t)\big)\mathbf{F}\mathbf{a}_t = \mathbf{F}\mathbf{a}_t - \mathbf{F}\mathbf{a}_t = \mathbf{0} \; .$$

Thus the residual factors have mean equal to zero and are uncorrelated with the market. Finally,

$$\mathbf{G} \equiv \mathrm{VAR}(\mathbf{g}_t) = E\Big(\big(\mathbf{I}-(\mathbf{F}\mathbf{a}_t\mathbf{a}_t')/(\mathbf{a}_t'\mathbf{F}\mathbf{a}_t)\big)(\mathbf{f}_t-\bar{\mathbf{f}})(\mathbf{f}_t-\bar{\mathbf{f}})'\big(\mathbf{I}-(\mathbf{a}_t\mathbf{a}_t'\mathbf{F})/(\mathbf{a}_t'\mathbf{F}\mathbf{a}_t)\big)\Big) \; .$$

$$\text{(14)} \qquad = \mathbf{F} - \mathbf{F}\mathbf{a}_t(\mathbf{a}_t'\mathbf{F}\mathbf{a}_t)^{-1}\mathbf{a}_t'\mathbf{F} \; .$$

Thus the variance matrix of the residual factors is equal to the original variance matrix $\mathbf{F}$ minus that part of the original variances and covariances that may be explained by correlation with the market return.

In summary, the multiple-factor model (1) implies the market model (10) with residual factors $\mathbf{g}_t$ having zero mean, zero covariance with the market return, and reduced variance matrix $\mathbf{G}$. The security beta in the implied market model is determined by the contributions of the factors to market variance, so that beta is a linear combination (given in (6)) of the loadings on the residual factors.

It remains to explore the implications of the model for the returns on a portfolio. Let z_{Pt} denote the excess return on a portfolio, with $z_{Pt} = \sum_{n=1}^{N} x_n z_{nt} = \mathbf{x}'\mathbf{z}_t$, where $\mathbf{x} = (x_1 : x_2 : .. : x_n)'$ is a vector of weights satisfying $\Sigma x_n = 1$. Then

(15) $$z_{Pt} = \underset{\sim}{x}'\underset{\sim}{\Lambda}_t\underset{\sim}{f}_t + \underset{\sim}{x}'\underset{\sim}{u}_t = \underset{\sim}{p}'\underset{\sim}{f}_t + \underset{\sim}{x}'\underset{\sim}{u}_t ,$$

where $\underset{\sim}{p} = \underset{\sim}{\Lambda}_t'\underset{\sim}{x} = \sum_n x_n\underset{\sim}{\lambda}_{nt}$ is the vector of portfolio loadings on the factors, equal to the weighted average of the loadings of the individual securities. Therefore, the moments of the portfolio return are

(16) $$E(z_{Pt}) = \underset{\sim}{p}'\overline{\underset{\sim}{f}} \quad , \quad VAR(z_{Pt}) = \underset{\sim}{p}'\underset{\sim}{F}\underset{\sim}{p} + \sum_n x_n^2\sigma_{nt}^2 .$$

The variance of portfolio return can be decomposed into that part due to covariance with the market, that part due to the residual factors, and that part due to specific risk:

$$VAR(z_{Pt}) = \underset{\sim}{p}'\underset{\sim}{F}\underset{\sim}{a}_t(\underset{\sim}{a}_t'\underset{\sim}{F}\underset{\sim}{a}_t)^{-1}\underset{\sim}{a}_t'\underset{\sim}{F}\underset{\sim}{p} + \underset{\sim}{p}'\left(\underset{\sim}{F}-\underset{\sim}{F}\underset{\sim}{a}_t(\underset{\sim}{a}_t'\underset{\sim}{F}\underset{\sim}{a}_t)^{-1}\underset{\sim}{a}_t'\underset{\sim}{F}\right)\underset{\sim}{p} + \sum_n x_n\sigma_{nt}^2$$

(17) $$= \beta_{Pt}^2\ VAR(z_{Mt}) + \underset{\sim}{p}'\underset{\sim}{G}\underset{\sim}{p} + \sum_n x_n\sigma_{nt}^2 \quad , \text{ where}$$

(18) $$\beta_{Pt} \equiv COV(r_{Pt},r_{Mt})/VAR(r_{Mt}) = \underset{\sim}{p}'\underset{\sim}{F}\underset{\sim}{a}_t/(\underset{\sim}{a}_t'\underset{\sim}{F}\underset{\sim}{a}_t).$$

A matched-beta market portfolio may be defined as the portfolio with factor loadings equal to $\beta_{Pt}\underset{\sim}{a}_t$: this portfolio has beta equal to β_{Pt} and has factor loadings proportional to those of the market. It is easily shown that the residual factors make no contribution to the variance of the matched-beta market portfolio, so that it is the minimum-variance portfolio with beta equal to β_{Pt}. Moreover, if the loadings of the original portfolio are written as $\underset{\sim}{p} = \beta_{Pt}\underset{\sim}{a}_t + \underset{\sim}{\delta}$, where $\underset{\sim}{\delta}$ is the discrepancy relative to the loadings of the matched-beta market portfolio, then it is easily shown that

$$\underset{\sim}{p}'\underset{\sim}{G}\underset{\sim}{p} = \underset{\sim}{\delta}'\underset{\sim}{G}\underset{\sim}{\delta} = \underset{\sim}{\delta}'\underset{\sim}{F}\underset{\sim}{\delta} \ . \tag{19}$$

Therefore,

$$\mathrm{VAR}(z_{Pt}) = \beta_{Pt}^2 \ \mathrm{VAR}(z_{Mt}) + (\underset{\sim}{p}-\underset{\sim}{a}_t)'\underset{\sim}{F}(\underset{\sim}{p}-\underset{\sim}{a}_t) + \Sigma x_n^2 \sigma_{nt}^2 \ . \tag{20}$$

The second term may be diversified away only by making the portfolio loadings similar to those in the matched-beta market portfolio, for with $\underset{\sim}{F}$ being a positive-definite variance matrix, the second term will go to zero only if $\underset{\sim}{\delta}$ goes to zero.

III. Empirical Estimation of a Multiple-Factor Model

A multiple-factor model is specified by the factor loading vectors $\underset{\sim}{\lambda}_{nt}$, $n=1,,N_t$, $t=1,,T$, and by the values of the factors f_{jt}, $j=1,,J$, $t=1,,T$. There are three distinct approaches that may be followed in fitting such a model.

The first approach consists of the factor analytic methods that are designed to operate when no a priori knowledge whatsoever is available about the factors and the loadings (see [6]). In these methods, criteria of maximal explained variance can be used to extract factors and loadings from the data, provided that the loadings can be assumed to be fixed over time, i.e., that for each n, $\underset{\sim}{\lambda}_{nt} \equiv \underset{\sim}{\lambda}_n$ for all t. King's study [4, Secs. III-VI, VIII-IX] and Farrar's earlier work [3] are excellent illustrations of these methods. Unfortunately, the methods have several telling weaknesses, in addition to the requirement that loadings be constant over time. The methods become computationally infeasible as N increases beyond 200, so that analysis of the entire universe of securities is impossible. The methods are inefficient in that they do not take advantage of knowledge, based upon economic theory, that factors should be associated with meaningful economic events, and that loadings should be associated with characteristics of firms such as their markets and their financial positions. Finally, there are difficulties of interpretation associated with these methods: statistical tests of the significance of the estimated factors are lacking, and the interpretation of the results is a matter of judgment.

The second approach assumes that the values of the factors are known. This approach at its simplest is taken when the "market return" is assumed to be the common factor underlying returns in the one-factor "market model." The same approach to a multiple-factor model is taken in King's study [4, Sec. VII] and in Cohen and Pogue's analysis [2]; in both cases, the factors are identified with average returns on various industry groups, so that $f_{jt} = \frac{1}{m_j} \sum_{\mathcal{I}_j} z_{nt}$, where the sum is taken over the m_j securities in industry $\mathcal{I}_j$. When the factors are known, the estimation problem is simple: the loadings for each security are estimated by regressing the time series of returns for that security on the time series of factor values. The loadings may be permitted to vary over time by the methods in [9] and [12]. The limitation of this approach is that it may be applied only where one has complete a priori knowledge of a factor. It may not be used to study factors for which the time series of values is unknown.

The third approach, which will be taken in this article, is to assume something about the loadings. Specifically, it is assumed that the factor loading for security n in period t is a fixed linear function of some "descriptors" of that security at that date:

$$\lambda_{nt} = D w_{nt} \, , \tag{21}$$

where $w_{nt} = (w_{1nt} : w_{2nt} : \cdots : w_{Jnt})'$ is a vector of descriptors that apply to security n at the beginning of period t, and where D is a $J \times J$ matrix of regression coefficients. The descriptors are assumed to be known predetermined variables that are independent of f_t and u_t.

It seems reasonable to assume that the loadings are functions of measurable characteristics of the securities. For example, the loading on a factor capturing the effects of "changes in anticipated inflation" should be a function of, among other things, a descriptor that captures "leverage"; the loading on a factor capturing "changes in anticipated monetary stringency" should be a function of a descriptor that measures "liquidity"; the loading on a factor that captures "changed expectations as to economy-wide corporate earnings" should be a function of the responsiveness of that firm's earnings to economy-wide earnings, or an "accounting beta"; the loading on a factor capturing the "change in investor risk-aversion" should be a function of all those descriptors that are relevant to the perceived risk of a security, such as "historical specific risk" and "historical market beta"; the loading on a factor that captures "changes in the prospects of an industry" should be a function of the "percentage of earnings arising from that industry"; etc. Thus, it makes sense to assume that the loadings are, at least in part, functions of measurable characteristics. It is less justifiable to assert that the function is linear with fixed coefficients, or to assert that the relationship is exact, but these simplifications will facilitate subsequent analysis.

When relation (21) holds, the various forms of the multiple-factor model derived in Section II may be re-expressed in a manner that is ideally suited to empirical estimation. The contributions of the factors to return are

$$\text{(22)} \quad \text{and} \quad \begin{aligned} \tilde{\lambda}'_{nt}\tilde{f}_t &= \tilde{w}'_{nt}\tilde{D}'\tilde{f}_t = \tilde{w}'_{nt}\tilde{f}^*_t \quad , \quad \text{where} \quad \tilde{f}^*_t = \tilde{D}'\tilde{f}_t \ , \\ \tilde{\lambda}'_{nt}\tilde{g}_t &= \tilde{w}'_{nt}\tilde{D}'\tilde{g}_t = \tilde{w}'_{nt}\tilde{g}^*_t \quad , \quad \text{where} \quad \tilde{g}^*_t = \tilde{D}'\tilde{g}_t \ . \end{aligned}$$

The starred factor vectors are linear transforms of the original factor vectors. The starred factors are of interest because the loadings on them are identically equal to the values of the descriptors; that is, the starred factors are associated with microeconomic characteristics of the firms. By contrast, the original factors were presumably associated with variables having broad macroeconomic meaning, such as inflation, interest rates, gross product, risk aversion. A multiple-factor model may be constructed in terms of the starred factors: the implications for security returns and market return are unchanged, with only the interpretation of the factors being different.

The basic multiple factor model (1), for any period t, with $\mu = 0$ as required by the CAPM, may be rewritten by using (22) as

$$\text{(Type I)} \qquad z_{nt} = \sum_{j=1}^{J} f^*_{jt}\{w_{jnt}\} + u_{nt} \qquad n=1,,N_t;\ t \text{ fixed.}$$

In any one period t, this is nothing more than a regression relation in which the z_{nt} are the dependent variables, the w_{jnt} are the predetermined regressors, and the f^*_{jt} are the unknown coefficients to be estimated. Thus, the realized values of the factors are estimated by regressing the returns on the descriptors. There are N_t observations, one for each security in the sample, and J parameters to be estimated. Assume for convenience that the variance of specific risk is the same for all securities, i.e., $\sigma^2_{nt} = \sigma^2_t$ for all n. Then the minimum variance linear unbiased estimator of the factor vector is

$$\hat{\underset{\sim}{f}}{}^*_t = \left(\sum_n \underset{\sim}{w}_{nt}\underset{\sim}{w}'_{nt}\right)^{-1} \sum_n \underset{\sim}{w}_{nt} z_{nt} \tag{23}$$

its variance-covariance matrix of estimation error is

$$\sigma^2_t \underset{\sim}{\Phi}_t = \sigma^2_t \left(\sum_n \underset{\sim}{w}_{nt}\underset{\sim}{w}'_{nt}\right)^{-1} \tag{24}$$

and an unbiased estimate of specific risk is provided by

$$\hat{\sigma}^2_t = \frac{\sum_n (z_{nt}-\hat{z}_{nt})^2}{(N_t-J)} \quad , \quad \text{where} \quad \hat{z}_{nt} = \underset{\sim}{w}'_{nt}\hat{\underset{\sim}{f}}{}^*_t. \tag{25}$$

A regression of Type I may be repeated for each of the T periods in the sample. The result is a collection of T estimated factor vectors, one for each period.

Thus far, no assumption has been made about the distribution of the factors. Under the assumption that this distribution is stationary, an unbiased estimator of the mean is provided by

$$\hat{\bar{\underset{\sim}{f}}}{}^{*} = \sum_{t=1}^{T} \hat{\underset{\sim}{f}}{}^{*}_{t}/T \ , \tag{26}$$

and an unbiased estimator of the variance matrix is provided by[1]

$$\hat{\underset{\sim}{F}}{}^{*} = \sum_{t=1}^{T} (\hat{\underset{\sim}{f}}{}^{*}_{t}-\hat{\bar{\underset{\sim}{f}}}{}^{*})(\hat{\underset{\sim}{f}}{}^{*}_{t}-\hat{\bar{\underset{\sim}{f}}}{}^{*})'/(T-1) - \sum_{t=1}^{T} \hat{\sigma}^{2}_{t}\underset{\sim}{\Phi}_{t}/T \ . \tag{27}$$

The methods discussed thus far suffice to analyze the multiple-factor model. However, it is interesting to separate the contributions of the factors into the component explained by market return and the contributions of the residual factors. The jumping-off point for this problem is the rewritten market model:

$$z_{nt} = \beta_{nt}z_{Mt} + \underset{\sim}{\lambda}'_{nt}\underset{\sim}{g}_{t} + u_{nt}$$

$$= \underset{\sim}{w}'_{nt}\underset{\sim}{D}'\underset{\sim}{\theta}_{t}z_{Mt} + \underset{\sim}{w}'_{nt}\underset{\sim}{D}'\underset{\sim}{g}_{t} + u_{nt}$$

$$= \underset{\sim}{w}'_{nt}\underset{\sim}{b}z_{Mt} + \underset{\sim}{w}'_{nt}\underset{\sim}{g}^{*}_{t} + u_{nt} \quad , \quad \text{where } \underset{\sim}{b} = \underset{\sim}{D}'\underset{\sim}{\theta}_{t} \ ,$$

$$\text{(Type II)} \qquad = \sum_{j=1}^{J} b_{j}\{w_{jnt}z_{Mt}\} + \sum_{j=1}^{J} g^{*}_{jt}\{w_{jnt}\} + u_{nt} \qquad \begin{matrix} n=1,,N_{t} \\ t=1,,T \ . \end{matrix}$$

---

[1]Although unbiased, these estimators of the mean and variance of the factors are not the most efficient, in the sense of having minimum mean square error, unless σ^{2}_{t} and $\underset{\sim}{\Phi}_{t}$ are constant over time. In the present case, where σ^{2}_{t} turns out to be nonconstant, and where the sample size increases from 400 to 1,300 as t increases, neither requirement is satisfied. However, the presentation of the more efficient estimators is not essential to the line of argument, and will be omitted here.

The regression of Type II includes all security returns in all time periods as dependent variables. The b_j, $j=1,,J$ are unknown parameters to be estimated. The g^*_{jt}, $j=1,,J$, $t=1,,T$, are unknown stochastic parameters, with mean vector equal to zero and unknown variance matrix $\underset{\sim}{G}^*$.

This is a stochastic parameter regression problem, which may be attacked by the methods discussed in [7] and [8]. Alternatively, a simple but less efficient approach may be taken by rewriting the regression as

$$\text{(Type III)} \qquad z_{nt} = \sum_{j=1}^{J} b_j\{w_{jnt}z_{Mt}\} + v_{nt} \qquad \begin{matrix} n=1,,N_t, \\ t=1,,T \ , \end{matrix}$$

where $v_{nt} = \sum_j g^*_{jt}w_{jnt} + u_{nt}$. From (13), the random variables v_{nt} are uncorrelated with the regressors $w_{jnt}z_{Mt}$, so that a regression of Type III will yield an unbiased estimator of $\underset{\sim}{b}$. However, the v_{nt} have nonconstant variances and nonzero covariances for $n_1 \neq n_2$, so that an ordinary least squares regression, in ignoring heteroscedasticity, will be inefficient and will yield downward biased estimators of estimation error variance. Despite this defect, regressions of Type III have the appeal of great simplicity, and several will be reported below.

Let $\hat{\underset{\sim}{b}}$ be an estimate of $\underset{\sim}{b}$, obtained by some statistical method. Then it is interesting to consider a decomposition of the variance of security returns in any period t which exhibits the usefulness of market return in explaining that variance. Let $s_t^2(x)$ denote the sample variance of variable x in period t, i.e., $s_t^2(x) = \sum_{n=1}^{N_t} \frac{(x_{nt}-x_{.t})^2}{(N_t-1)}$ where

$x_{.t}$ denotes an average over the observations $n = 1,,N_t$. Then, from the regression of Type I,

$$(28) \qquad s_t^2(z_{nt}) = s_t^2(\hat{z}_{nt}) + s_t^2(z_{nt}-\hat{z}_{nt}) .$$

Total variance of return = Variance explained by factors + Unexplained variance .

Alternatively, if $\hat{\beta}_{nt} = \underset{\sim}{w}'_{nt}\underset{\sim}{\hat{b}}$ is the fitted beta,

$$(29) \quad s_t^2(z_{nt}) = \left(s_t^2(z_{nt})-s_t^2(z_{nt}-\hat{\beta}_{nt}z_{Mt})\right) + \left(s_t^2(z_{nt}-\hat{\beta}_{nt}z_{Mt})-s_t^2(z_{nt}-\hat{z}_{nt})\right)$$

Total Variance = Variance explained by predicted beta + Variance explained by factors net of beta explanation

$$+ \; s_t^2(z_{nt}-\hat{z}_{nt}) .$$

Unexplained variance .

The first term in (29) is the difference between the variance of return and the variance of residual return after subtraction of the prediction based on the fitted beta. This difference, or explanation of variance, will be positive on average, but in any one period it may be negative; this is the case because the beta relationship is fitted over all periods, and in an atypical period the variance of returns may actually be increased by subtracting the fitted predictions. The second term is equal to the total variance explained by the factors, minus the explanation by beta. It also equals the amount of variance that would be explained in a regression on the factors in which the dependent variable was $z_{nt} - \hat{\beta}_{nt}z_{Mt}$.

IV. The Data and the Choice of Descriptors

The period of observation was taken as a month: z_{nt} was taken as the monthly return relative, $(p_{t+1} + DIV_t)/(p_t)$, where p_s is the price at the beginning of month s, and DIV_s is the dividend paid in month s, and where appropriate adjustments for stock splits are made. The risk-free rate, r_{Ft}, was taken as the ninety-day treasury bill rate. The market return, r_{Mt}, was taken as the arithmetic average of returns on all NYSE listed securities. Then excess returns were defined as $z_{nt} = \ln(1+r_{nt}-r_{Ft})$, $z_{Mt} = \ln(1+r_{Mt}-r_{Ft})$. Thus the market model (1) was redefined as linear in logarithms of return. There are various advantages and disadvantages in this redefinition, which need not be gone into here: suffice it to say that for a time period as short as one month, the transformation makes little difference. An extensive file of monthly return relatives was prepared by merging a CRSP tape, containing return relatives through June 1968 on all NYSE listed securities, with COMPUSTAT quarterly industrial and utility tapes, containing monthly price relatives and quarterly dividends for a large sample of NYSE and ASE listed securities for the decade ending in 1972. Experiences in preparing this merged file are described in [11]. In the process of correcting numerous errors, we concluded that a very high percentage of those extreme monthly price relatives satisfying $|\ln(r_{nt})| > .4 + |\ln(r_{Mt})|$ were erroneous. Hence, the few such data points that were encountered were rejected from the study.

Annual income statement and balance sheet data for corporations were taken from a merged file of annual COMPUSTAT industrial data, described in [9, Sec. IV], and a 1972 COMPUSTAT Annual Utility Tape. Descriptors based upon annual reports were revised each fiscal year (quarterly data were rejected as probably containing excessive noise), and descriptors based exclusively on stock market events were revised each calendar year. Any descriptor based on an annual report was assumed to become publicly available 120 days after the close of that fiscal year. Any descriptor based upon stock market data, such as an estimate of historical specific risk, was assumed to become available immediately. For any month t, and any firm n, the values of the descriptors taken for that observation were the latest available values: specifically, descriptors incorporating information from an annual report were based on the most recent fiscal year for firm n closing four or more months before month t, and descriptors based only on stock market occurrences were taken as of the calendar year ending in the preceding fiscal year. Many descriptors are based on five years of previous data. Accordingly, a complete set of descriptors does not become available until five years after the start of the data base.

The sample period was from October 1957 (t=1) through September 1972 (t=180). Those observations for which z_{nt} and a complete set of available descriptors w_{nt} were available were included in the regression. The number of available securities increases progressively from $N_1 = 388$ to $N_{180} = 1380$. The early increase in sample size reflects increasing comprehensiveness of the COMPUSTAT annual tape. Later, a sharp increase in sample size occurs in 1967, when five calendar years of historical return data on ASE firms becomes available from the COMPUSTAT quarterly tape. In all, there are 112,086 monthly observations.

Unquestionably, this procedure led to a nonrepresentative sample. There is a bias in favor of survivors, since nonsurvivors are usually excluded from our COMPUSTAT data base, and an early bias toward bigness, since COMPUSTAT first added the largest and most important firms to its data base. (An additional potential bias, due to the retrospective addition of data to the COMPUSTAT data base for a newly listed firm, is avoided by accepting data for a security only in those years where the security is listed on an exchange.) However, the extra-market covariance in this nonrepresentative sample may not differ greatly from that in a universal sample, with the unfortunate exception that the sample excludes certain industry groups that have unquestionably behaved differently from the norm: banks, insurance companies, railroads, brokerages, and small high-technology firms traded over the counter.

The choice of descriptors was largely determined by considerations discussed in [10], with a few deletions and additions suggested by experience reported in [9]. The list of descriptors, with mnemonics, is given below, and the computational formulae defining these are given in the appendix. A descriptor based on five years of previous data is indicated by the notation (5); one based on more than five years by (5+).

1. PAY: *Dividend Payout Ratio* (average dividend/average price) (5)
2. SIZE: *Logarithm of Total Assets* (5)
3. LIQ: *Liquidity* (the quick ratio)
4. CUT: *Absolute Magnitude of Per-Share Dividend Cuts* (5)
5. LEV: *Mean Leverage* (senior securities/total assets) (5)
6. NSG: *Nonsustainable Growth Estimate*
7. PLANT: *Gross Plant Per Dollar of Total Assets*
8. STO: *Share Turnover* (annual number of shares traded/shares outstanding)
9. LNP: *Logarithm of Share Price*

10. YIELD: *Cash Yield* (average dividend/average price) (5)
11. ΔE: *Latest Annual Proportional Change in Per-Share Earnings*
12. σE: *Standard Deviation of Earnings Per Share* (5)
13. $\sigma(\Delta E)$: *Standard Deviation of a Per-Share Earnings Growth Measure* (5)
14. B/P: *Book Value of Common Equity Per Share/Price*
15. $A\beta$: *Accounting Beta* (covariability of earnings growth with overall corporate earnings growth) (5+)
16. E/P: *Earnings Per Share/Price*
17. $\sigma(E/P)$: *Standard Deviation of Earnings Per Share/Price* (5)
18. PRNEG: *Estimated Probability that Operating Earnings Will Not Cover Fixed Charges*
19. ROI: *Return on Investment* (5)
20. EGROW: *Earnings Growth* (trend/average) (5)
21. SIGMA: *Historical Specific Risk* (standard error of regression (23)) (5)
22. BBETA: *Alternative Accounting Beta*, as defined by Beaver, Kettler, and Scholes [1] (5)
23. HBETA: *Historical Beta* (regression coefficient of z_{nt} on z_{Mt} over last five calendar years (60 data points)) (5)
24. QUAL: *Standard & Poor's Quality Rating* (as of previous December)
25. DQUAL: *Dummy Variable* (equal to one if the quality rating is not available)

To minimize the effects of outlying values of the descriptors, the mean and standard deviation of each descriptor in each period was computed, and every value lying more than three standard deviations from the mean was set to be exactly three standard deviations away. Thus the tails of the distributions were "Windsorized." Then, for all descriptors except HBETA, QUAL, and DQUAL, the values in each period were standardized to have mean equal to zero and standard deviation equal to one. This standardization is consistent with the hypothesis that the loadings on the factors are determined by the relative position of the firm within the population, rather than by the actual numerical value of the descriptors.

The great weakness of the standardization is that, due to the progressive appearance of less mature firms in the sample, the population is not stationary, so that to take the position relative to the population in each period is to introduce a nonstationarity (that will be removed in subsequent studies by standardizing relative to a fixed population). The standardization also facilitates the interpretation of estimated regression coefficients, which indicate the effect of a one-standard-deviation change in the descriptor upon the dependent variable.

In addition to these descriptors, descriptors assigning firms to industry groups were used. For each group, the descriptor is a dummy variable, equal to one if the firm is in an industry in the group, equal to zero otherwise. Firms were assigned to industry groups by their S&P industry code as of 1972. No effort was made to track possible changes in industry group over the sample period, nor to distribute a firm across several groups if its activities were not confined to one. Industry groups were chosen on a priori grounds, with the requirement that the number of available firms in a group must never fall below two, and must average at least six over the 180 months of data. There is no doubt that the grouping may be improved upon. The groupings, with mnemonics for them, numbers of observations in the group, and COMPUSTAT industry identifiers included in the group are given below:

1. MISCCP - miscellaneous consumer products; N = 2547
   3914, 3269, 3229, 2951, 3911, 7231, 2771, 3069, 2591, 2841, 3953, 3871, 5999, 5211, 5971, 5912, 5722, 5712, 3831, 5661, 5952, 5621, 5988, 5812
2. MINING - mining; N = 2619
   1211, 1042, 3295, 5093, 1477, 3399, 3331, 3341

3. D.OILS - domestic oils and petroleum services; N = 1439
1311, 6792, 2911, 8736

4. I.OILS - international oils; N = 1065
2913

5. FOODSA - foods and associated products; N = 7374
0135, 2041, 5040, 2099, 2052, 2051, 2030, 2046, 2020, 2010, 2000, 2042, 2044, 2063, 0119, 2062, 2093, 2070, 0989

6. APAREL - textiles and apparel; N = 5406
3131, 3199, 3111, 2300, 2200, 2291, 2262, 3141, 2340, 2330

7. PAPERA - paper and associated products; N = 2094
2600, 2661, 2649, 5096, 0800, 2499

8. MEDIAS - newspapers, publishing, communication broadcasters; N = 1605
2711, 2700, 2731

9. CHMCLS - chemicals; N = 5604
2899, 2818, 2802, 2801, 2803, 2871, 2819, 2821, 3079, 2823, 4925

10. DRUGSA - drugs and cosmetics; N = 3480
2844, 2835, 2837, 2836, 2834, 5022, 3843, 8061, 3841

11. RUBBER - rubber and plastics; N = 1500
3031, 3000, 3011

12. CMATLS - construction materials; N = 4563
3430, 2950, 3251, 3241, 3253, 3272, 1442, 3433, 3642, 5098, 2433, 3273, 3564, 3643, 2850, 3442, 2542, 3211, 3351, 3357, 3441, 3317, 3494, 3429

13. CNTANR - container; N = 1713
3221, 2650, 2653, 2651, 2654

14. ISTEEL - iron and steel; N = 3096
3321, 3323, 3310, 3311

15. OF.M/C - office machines; N = 1518
3570, 3579, 3571, 3573

16. NE.M/C - non-electrical machinery; N = 8004
3400, 3560, 3551, 3536, 3540, 3569, 5081, 3559, 3550, 3555, 3567, 3554, 3522, 3531, 4111, 3297, 3291, 3452, 3499, 3443, 3449, 3576

17. EEQUIP - electrical equipment; N = 3345
3600, 3610, 3821, 3511, 3629, 3622, 3641, 5063, 3699, 3822

18. ELECTR - electronics; N = 2967
3662, 3679, 5065, 3670

19. H.HOLD - household white and brown goods and furniture; N = 2841
3630, 3651, 2510

20. AUTOEQ - motor vehicles and equipment; N = 4332
3714, 3713, 3711, 5012

21. AEROSP - aerospace; N = 2739
3721, 3722, 3725, 3729

22. INSTRM - instruments; N = 1377
3811, 7391, 3611, 3861

23. LEISUR - leisure time industry; N = 1179
3949, 1951, 7948, 3652, 3931, 7821, 7811, 7831, 7810, 3941, 5969, 7931, 3732, 3751

24. AIRTRN - air transportation; N = 1887
4511

25. TOBACO - tobacco; N = 1374
2121, 2111

26. RETAIL - retail merchandise; N = 4230
5311, 5331, 5322, 5600, 5321

27. FSTORS - food stores; N = 2010
5411, 5042

28. MISCBS - miscellaneous business services; N = 1452
3581, 7213, 2761, 2752, 3589, 2799, 8999, 8931, 7311, 7349, 7392, 8911, 7399, 7393, 1799, 5341, 7011

29. FINANC - finance companies; N = 1047
6125, 6145, 7394, 6140

30. BEVRAG beverages; N = 1587
2082, 2085, 2086

31. NFMETL - nonferrous metals; N = 2688
3334, 1031, 3497, 1000, 3369

32. UTLITY - utilities; N = 12621
4910-4930, 4811

Miscellaneous; N = 7905 (all other COMPUSTAT industry numbers)

V. Selected Results

V.I. The results of six regressions of Type III for beta appear in Table 1. The first six columns give the estimated coefficients for the descriptor w_1 (which is identically equal to one, so that its coefficient is the constant term in the equation for beta), and for the accounting- and market-based descriptors w_2 through w_{26} that were defined in the previous section. The last three columns give the estimated coefficients for the industry dummy variables, d_{27} through d_{58}, in the three regressions where these variables appear. The effect of these dummy variables is to add to the beta of each security n the estimated coefficient for the dummy variable corresponding to the industry in which firm n appears. Thus, if $j(n)$ denotes the index of the industry in which firm n appears, the effect is to add $b_{j(n)}$ to beta. Of course, to avoid multicollinearity, the dummy variable for one industry group must be omitted: the omitted group is the miscellaneous category. The six regressions and the implied models for beta are given by

1. $$z_{nt} = \alpha + [b_1]z_{Mt} + u_{nt}$$

2. $$z_{nt} = \alpha + [b_1 + b_2 HBETA_{nt}]z_{Mt} + u_{nt}$$

3. $$z_{nt} = \alpha + \left[b_1 + b_2 HBETA_{nt} + \sum_{j=3}^{26} b_j w_{jnt}\right] z_{Mt} + u_{nt}$$

4. $$z_{nt} = \alpha + [b_1 + b_{j(n)}]z_{Mt} + u_{nt}$$

5. $$z_{nt} = \alpha + \left[b_1 + b_2 HBETA_{nt} + \sum_{j=3}^{26} b_j w_{jnt} + b_{j(n)}\right] z_{Mt} + u_{nt}$$

TABLE 1

TYPE 3 REGRESSIONS FOR BETA

	Regression Number					
	1	2	3	4	5	6
R^2	.2563	.2719	.2774	.2701	.2793	.2794
Alpha	-.00240 (.00022)	-.00236 (.00021)	-.00237 (.00021)	-.00231 (.00021)	-.00234 (.00021)	-.00235 (.00021)
Descriptors						
CONST	.8846 (.0045)	.3665 (.0115)	.6620 (.0189)	1.1015 (.0166)	.8181 (.0255)	.8189 (.0227)
HBETA		.5446 (.0111)	.2871 (.0150)		.2038 (.0161)	.2028 (.0159)
PAY			-.0104 (.0090)		.0013 (.0093)	
SIZE			.0044 (.0063)		.0095 (.0069)	.0101 (.0066)
LIQ			.0198 (.0052)		.0136 (.0058)	.0142 (.0057)
CUT			.0071 (.0078)		.0052 (.0079)	
LEV			-.0030 (.0057)		.0384 (.0077)	.0387 (.0076)
NSG			.0041 (.1268)		.0466 (.1274)	
PLANT			-.0348 (.0052)		-.0117 (.0072)	-.0150 (.0067)
STO			.0935 (.0091)		.0956 (.0093)	.0957 (.0090)
LNP			-.0234 (.0072)		-.0342 (.0074)	-.0341 (.0070)
YIELD			-.0218 (.0105)		-.0373 (.0108)	-.0347 (.0068)
ΔE			-.0148 (.0063)		-.0109 (.0064)	-.0087 (.0056)
σE			.0209 (.0093)		.0133 (.0095)	.0144 (.0080)
σ(ΔE)			.0117 (.0093)		-.0004 (.0094)	
B/P			.0432 (.0071)		.0523 (.0074)	.0536 (.0067)
AS			.0083 (.0060)		-.0070 (.0062)	-.0073 (.0060)
E/P			.0036 (.0094)		.0074 (.0095)	
σ(E/P)			.0010 (.0113)		-.0033 (.0117)	
PRNEG			.0006 (.0098)		.0110 (.0099)	.0098 (.0082)
ROI			.0174 (.0141)		.0223 (.0142)	.0225 (.0135)
EGROW			-.0091 (.1264)		-.0456 (.1269)	
SIGMA			.0516 (.0083)		.0337 (.0085)	.0346 (.0083)
RBETA			-.0036 (.0082)		-.0004 (.0085)	
QUAL			-.0068 (.0019)		-.0080 (.0022)	-.0077 (.0021)
DQUAL			-.0643 (.0154)		-.0268 (.0169)	-.0256 (.0165)

	Regression Number		
Industry Shifts	4	5	6
MISCCP	-.1691 (.0338)	-.0410 (.0350)	-.0450 (.0312)
MINING	-.2818 (.0335)	-.1268 (.0348)	-.1244 (.0309)
D.OILS	-.2964 (.0280)	-.0982 (.0307)	-.0976 (.0259)
I.OILS	-.6116 (.0493)	-.2786 (.0531)	-.2776 (.0499)
FOODSA	-.3934 (.0240)	-.1715 (.0259)	-.1725 (.0206)
APAREL	-.1545 (.0260)	-.0728 (.0273)	-.0765 (.0229)
PAPERA	-.2384 (.0370)	-.0299 (.0387)	
MEDIAS	-.0149 (.0402)	.1526 (.0415)	.1480 (.0383)
CHMCLS	-.1906 (.0262)	-.0640 (.0275)	-.0608 (.0235)
DRUGSA	-.3015 (.0296)	-.1011 (.0319)	-.1041 (.0284)
RUBBER	-.2390 (.0412)	-.1112 (.0420)	-.1145 (.0389)
CMATLS	-.1725 (.0275)	-.0170 (.0289)	
CNTANR	-.2961 (.0400)	-.1085 (.0410)	-.1076 (.0379)
ISTEEL	-.2140 (.0316)	-.0823 (.0335)	-.0805 (.0297)
OF.M/C	.0053 (.0414)	.0357 (.0432)	
NE.M/C	-.0686 (.0234)	.0427 (.0245)	.0404 (.0192)
EEQUIP	-.0854 (.0307)	.0245 (.0320)	
ELECTR	.1982 (.0316)	.0741 (.0329)	.0696 (.0294)
H.HOLD	-.0031 (.0328)	.0239 (.0337)	
AUTOEQ	-.1534 (.0280)	-.0022 (.0293)	
AEROSP	-.0874 (.0331)	-.0864 (.0343)	-.0914 (.0314)
INSTRM	.0386 (.0438)	.1265 (.0452)	.1208 (.0420)
LEISUR	.1428 (.0470)	.0715 (.0474)	.0663 (.0454)
AIRTRN	.2500 (.0386)	.1307 (.0412)	.1273 (.0382)
TOBACO	-.4137 (.0441)	-.1680 (.0468)	-.1709 (.0439)
RETAIL	-.2869 (.0283)	-.0741 (.0302)	-.0776 (.0258)
FSTORS	-.3946 (.0368)	-.1711 (.0398)	-.1706 (.0367)
MISCRS	-.2052 (.0421)	-.0628 (.0425)	-.0628 (.0401)
FINANC	-.2576 (.0493)	-.1923 (.0537)	-.1994 (.0518)
BEVRAC	-.3856 (.0413)	-.1709 (.0424)	-.1731 (.0393)
SEMETL	-.1453 (.0328)	-.0478 (.0338)	-.0474 (.0299)
UTLITY	-.6037 (.0210)	-.1544 (.0326)	-.1506 (.0285)

Standard errors appear in parentheses.

6. Same as 5, above, but with least significant descriptors, having F-statistics less than one, deleted by a stepwise regression program.

Note: In every case, the expression inside square brackets is the implied model for β_{nt}.

Regression 1 shows that the average beta is .8846, indicating that the selected market return--the average return on NYSE listed securities--includes securities with higher average beta than this sample, undoubtedly as a result of a sample bias toward mature firms. The R^2 equal to .2563 shows that 25.63 percent of the variance in monthly logarithmic excess returns can be explained by the logarithmic market return, if beta is assumed constant.

Regression 2 gives a linear prediction rule for β_{nt} as a function of the estimated historical beta over the previous five calendar years for that security. The prediction rule is

$$\beta_{nt} = .3665 + .5446\ \mathrm{HBETA}_{nt} \simeq \overline{\mathrm{HBETA}}_{.t} + .5446(\mathrm{HBETA}_{nt} - \overline{\mathrm{HBETA}}_{.t}).$$

HBETA provides immensely significant predictive power; the explained variance increases to 27.19 percent, and the F-statistic for the added explanatory power is $F(1,\ 112084) = 2{,}404$ against a 99 percent confidence point, for rejection of the null hypothesis of no explanatory power, equal to 6.65.[1]

[1] Since the monthly logarithmic return relatives have kurtosis somewhat greater than normal, and since there is some heteroscedasticity in the regression, the critical points for F-statistics and R^2 given in the text have somewhat overstated confidence levels. The statistical tests should therefore be considered as indicative rather than conclusive: fortunately, the actual values of the statistics are generally many times the critical value.

Notice that the predicted beta is adjusted by only 54.46 percent of the difference of the historical beta from its mean value. This is the optimal linear rule over the sample period, in the sense of achieving minimum mean square error in predicting security returns. An adjustment of only 54 percent indicates that the fitted historical beta is a less than perfect predictor of current beta. This imperfection may arise in part from estimation error in HBETA, which could be reduced by more efficient estimation methods than the conventional sixty-month regression, but another probable source of imperfection is nonconstancy in beta.

In Regression 3, the other accounting- and market-based descriptors are used as additional predictors of beta. These provide an important and highly significant increase in explanatory power. The F-statistic for the 24 additional variables is 35.5 in contrast to a 99 percent confidence point of 1.79.

In Regression 4, these descriptors are removed; instead, the variation of beta across industries is modeled by the use of the industry dummy variables. The estimated coefficient for each industry gives the difference between the average beta for that industry and the average beta for the group of firms in the miscellaneous category. For instance, the average beta for utilities is -.6037 below the average for the miscellaneous category, while the average for air transportation firms is .2500 greater than that of the miscellaneous category. The constant term, 1.105, gives the average beta for the miscellaneous category. The industry differences are highly significant, with $F(32, 112052) = 66.3$ against a 99 percent confidence point of 1.67.

In Regression 5, all of the descriptors are included. When this regression is compared with Regression 3, the additional explanatory power of the industry dummies, over and above that provided by the accounting- and market-based descriptors, is highly significant, with $F(32, 112036) = 9.2$ against the 99 percent confidence point of 1.67. Regression 6 differs only in that a stepwise regression program has removed minimally significant variables from Regression 5 until no variables with F-statistic less than 1.0 remain. None of the estimated coefficients for the remaining variables have changed importantly as a result of these deletions, so that Regressions 5 and 6 are effectively interchangeable.

These regressions have several implications for the nature of beta. Among the most interesting are the following:

(i) HBETA is an important predictor of beta, but by no means the only descriptor containing useful predictive information. The incremental contribution to predictive power by any one descriptor is given by its F-statistic (the square of the ratio of the estimated coefficient to its standard error). The F-statistic of HBETA is 160.3. This indicates that if HBETA were lost as a descriptor, the explained variance would be reduced by 160.3 times the residual mean square error. But the total variance explained by all accounting- and market-based descriptors and industry dummies is 3,138 times the residual mean square error. Therefore, the incremental contribution of HBETA to predicted variance, over and above that provided by the other descriptors, is only 5.1 percent of the total predictable variance.

(ii) Three other descriptors, based only on the behavior of the security in the market, provide an important contribution to prediction of beta: proportional share turnover in the previous calendar year (STO), ($F = 106.6$); residual standard error in the beta regression over the five previous calendar years (SIGMA), ($F = 15.6$); and the logarithm of share price (LNP), ($F = 21.6$). All are significant, for the 99 percent confidence point for $F(1, {}^{11}2026)$ is 6.64. The magnitudes of the estimated effects are as follows: a one-standard-deviation increase in STO causes an increase in predicted beta of .0956; a one-standard-deviation increase in SIGMA increases predicted beta by .033; a one-standard-deviation increase in LNP decreases beta by -.034.

(iii) Several accounting-based descriptors are also highly significant and show important effects. The most noteworthy are leverage (LEV), with an increase of one standard deviation implying an increase of .038 in beta ($F = 25.0$);[1] book value per share/share price (B/P), with an increase of one standard deviation implying an increase of .052 in beta ($F = 50.1$); and the dividend yield, with a one-standard-deviation increase implying a reduction of .037 in beta ($F = 12.0$). The Standard and Poor's quality rating contains significant information ($F = 13.0$). The difference between the extreme ratings of A+ and C brings an adjustment of .056 in beta.

---

[1]Notice that in Regression 3, where the industry dummies are not included, the effect of leverage on beta is negligible. It is only when the industry dummies are included, so that the descriptors introduce "within-industry" differences only, that the effect of leverage emerges.

(iv) There are numerous highly significant differences across industries that are not explained by the accounting-based and market-based descriptors, so that they must be introduced by the industry dummies.[1] International oils have an average beta shift, net of the difference accounted for by the other descriptors, of -.28 (F = 31.0); the media--publishing firms, newspapers, radio, and TV--have a positive shift of .15 above the miscellaneous category (F = 13.5); air transportation firms (F = 10.0) have a beta shift of +.13; tobacco (F = 12.9), food and associated products (F = 43.9), food retailers (F = 18.4), and beverage producers (F = 16.2) have beta shifts of -.17; and utilities (F = 118.1) have a beta shift of -.35.

V.2 In the Type I regressions to estimate the factors, in order to economize on computations, the observations were first grouped into calendar quarters, so that there were only sixty regressions to be run--one for each quarter. It is impossible to report all the details of these regressions,

---

[1] It is interesting to contrast the estimated industry coefficients in Regression 4 with those in Regression 5. Differences between coefficients for two industries in Regression 4 equal the difference in average beta, while differences in Regression 5 coefficients equal the average difference in beta net of the difference explained by the accounting- and market-based descriptors. For example, the paper-related industries (PAPERA) have an average beta that is significantly lower than the miscellaneous category (-.2384), and that falls well below the electronics industry (ELECTR) (the difference is [-.2384 - (.1982) = -.4366]). However, the accounting- and market-based descriptors included in Regression 5 almost fully explain these differences: PAPERA is now insignificantly different from the miscellaneous firms (-.0299), and the difference between average betas in PAPERA and ELECTR, that is unexplained by the accounting- and market-based descriptors, is only [-.0299 - (.0741) = -.1040].

since each one involved the estimation of all fifty-eight factors. Instead, Table 2 summarizes the variance explained by the factors quarter by quarter. The first column gives the variance of percentage quarterly return on a typical security which is attributed to the fifty-eight factors. The second column gives the variance of percentage quarterly return which is attributed to specific risk.

Notice the progressive increase in the specific risk of the sample, which is not necessarily due to increased variance in the market, but rather to the progressive broadening of the sample to include less mature firms. There can be no doubt that a large part of the total variance in the market is associated with the factors: Column 4 gives the adjusted R^2 for each regression. The 99 percent confidence point for the R^2 decreases with increasing sample size from .09 for 1957.IV to .03 for 1972.III. Of the sixty regressions, each of which is independent from the others, all are significant.

Column 3 of the table gives the net reduction in variance achieved by the fitted predictions $\hat{\beta}_{nt} z_{Mt}$. This shows an extremely interesting pattern that can be seen in Table 3, where the quarterly averages of the statistics in Table 2 appear. If the return-generating process were stationary, the potential predictive ability would be directly proportional to the square of market return. But, it is apparent from examining Table 3 that in the first calendar quarter, and to a lesser degree in the third calendar quarter, the fitted beta model is far more successful in explaining the variance of security returns than in the fourth calendar quarter. In fact, the fitted beta model reduces the variance of security returns

TABLE 2

FACTOR REGRESSIONS BY QUARTERS

Quarter and Year	Variance of Percent Return: Explained by Factors	Variance of Percent Return: Due to Specific Risk	Variance of Percent Return: Explained by Fitted β	$\bar{R}^2$ for Type I Regression	Sample Mean Return	Percent Return on HBETA Factor for $z_{nt}-\hat{\beta}_{nt}z_{Mt}$
1957 IV	10.51	8.09	3.10	0.49	-2.32	-2.48
1958 I	3.23	5.30	-0.43	0.28	3.48	-2.27
II	1.78	4.70	-0.37	0.16	2.59	-1.46
III	5.87	4.88	2.20	0.48	3.84	1.29
IV	2.61	7.68	-0.68	0.14	3.33	-1.61
1959 I	4.44	6.46	0.33	0.32	1.97	-0.89
II	6.06	9.33	0.42	0.30	1.18	1.87
III	4.42	4.62	0.74	0.41	-1.42	-1.88
IV	3.20	6.47	0.26	0.23	1.51	0.50
1960 I	5.79	6.53	1.21	0.39	-2.43	-2.67
II	5.24	10.05	-0.26	0.25	0.91	-2.65
III	5.81	6.30	0.89	0.41	-1.75	-2.67
IV	3.10	6.88	-0.85	0.21	2.07	-1.54
1961 I	3.45	9.98	1.23	0.15	4.89	-1.09
II	2.78	9.02	0.00	0.13	-0.01	-0.35
III	5.13	8.12	0.11	0.30	0.24	-3.13
IV	2.34	7.61	-0.16	0.13	1.73	-0.96
1962 I	2.92	6.21	-0.02	0.23	-0.51	0.21
II	6.53	8.20	1.57	0.37	-8.30	1.00
III	3.27	5.86	-0.04	0.27	0.30	-0.50
IV	3.02	7.10	-0.68	0.21	3.25	-1.16
1963 I	2.45	5.06	-0.26	0.24	2.15	-1.38
II	1.97	6.63	0.30	0.13	1.29	0.54
III	2.09	5.80	-0.03	0.17	0.61	-0.07
IV	2.75	6.55	0.05	0.21	0.50	1.52
1964 I	5.94	8.00	0.69	0.35	2.24	0.35
II	2.69	7.37	-0.18	0.17	0.46	-2.08
III	2.23	6.95	-0.05	0.15	1.40	-0.81
IV	2.62	6.52	-0.00	0.20	0.13	-0.95

Quarter and Year	Variance of Percent Return: Explained by Factors	Variance of Percent Return: Due to Specific Risk	Variance of Percent Return: Explained by Fitted β	$\bar{R}^2$ for Type I Regression	Sample Mean Return	Percent Return on HBETA Factor for $z_{nt}-\hat{\beta}_{nt}z_{Mt}$
1965 I	2.75	5.68	1.09	0.25	2.34	1.10
II	1.34	5.54	-0.20	0.10	-1.65	0.86
III	6.37	8.71	1.87	0.36	2.67	1.39
IV	5.43	7.65	2.32	0.35	2.38	1.65
1966 I	5.67	10.38	0.42	0.29	-0.03	1.70
II	2.46	8.33	-0.03	0.15	-2.06	0.27
III	4.33	11.08	1.74	0.21	-4.24	-0.53
IV	5.79	11.40	-0.11	0.27	1.27	0.61
1967 I	7.74	12.56	3.82	0.32	5.02	-0.19
II	5.72	12.83	0.82	0.24	1.38	0.21
III	3.66	10.67	-0.13	0.19	2.41	-1.48
IV	3.41	12.68	0.02	0.14	-0.34	-0.78
1968 I	6.09	11.51	1.05	0.29	-2.60	-2.03
II	4.16	13.66	-0.29	0.16	4.23	-1.99
III	3.58	11.61	0.05	0.17	0.86	-0.83
IV	2.53	11.01	-0.36	0.12	1.27	-1.04
1969 I	2.30	10.53	0.82	0.11	-2.63	-0.24
II	5.21	12.17	1.89	0.24	-3.94	-1.57
III	3.85	15.91	0.03	0.13	-2.10	0.71
IV	4.95	15.54	0.60	0.18	-1.36	-0.57
1970 I	4.31	14.81	0.45	0.16	-0.65	-0.94
II	10.13	19.36	6.77	0.29	-8.92	0.92
III	5.19	13.03	2.40	0.23	5.02	-0.36
IV	4.68	11.83	-1.87	0.23	1.95	-4.22
1971 I	5.34	12.00	3.02	0.26	4.37	0.35
II	4.41	15.85	0.40	0.17	-1.70	-0.73
III	3.67	16.47	0.34	0.15	-1.54	-0.25
IV	3.18	13.79	-0.31	0.15	-0.12	-1.98
1972 I	7.13	19.08	2.28	0.24	3.42	2.23
II	4.90	27.16	0.58	0.12	-2.20	-1.27
III	5.07	19.79	0.58	0.17	-2.10	-1.86

TABLE 3

QUARTERLY AVERAGES FOR FACTOR REGRESSIONS

Quarter	Variance of Percent Return			$\bar{R}^2$ for Type I Regression	Sample Mean Return, Squared
	Explained by Factors	Due to Specific Risk	Explained by Fitted β		
I	4.64	9.61	1.05	0.26	8.75
II	4.36	11.35	0.76	0.20	13.95
III	4.30	9.99	0.71	0.26	6.02
IV	4.01	9.39	0.09	0.22	3.44

in twelve out of fifteen instances in the first quarter, but in only six out of fifteen instances in the fourth quarter. The mean square market return in the fourth quarter is smaller than in the other quarters, but not enough smaller to explain this difference. Apparently, the fitted betas are much more successful predictors of events in the first quarter than in the fourth. This is not due to obsolescence in the descriptors. The values are updated four months after the close of the firm's fiscal year, so that almost all are updated in the second or third quarters and, accordingly, are more timely as predictors of events in the fourth quarter than in the first quarter of the following year. The explanation that appears to be the most probable at this point in the research is that the return-generating process itself is nonstationary. Specifically, a plausible explanation would be that the information reaching the market in the first quarter, when the previous year's annual reports are gradually revealed, is more closely related to the factors associated with the descriptors than the information reaching the market in the fourth quarter.

Column 5 of Table 2 gives the monthly mean of z_{nt} (× 100) for comparison with the explained variances. For illustrative purposes, Column 6 of Table 2 gives the estimated factor $\hat{f}_{2t}$ for each quarter, from the Type I regression

$$z_{nt} - \hat{\beta}_{nt} z_{Mt} = f_{1t} + f_{2t}\,\mathrm{HBETA}_{nt} + u_{nt} \qquad n=1,,N_t,\ t\ \text{fixed.}$$

The estimated values are multiplied by 100 to approximate monthly return. The estimated values are substantial. A nonzero value in any quarter implies that the distribution of returns along the systematic risk axis was not distributed in that quarter as predicted by the fitted betas. Such an occurrence could be the result of the selected risk-free rate differing from the true value, of the chosen market return being erroneous, or of a shift in the valuation of securities in that period that was somehow related to the systematic risk of securities. Such a shift could be the result of a change in risk aversion, a change in the assessment of overall market risk, or any other expectational change that affected anticipated earnings in a manner that was correlated with systematic risk.

Finally, Table 4 gives the significance of the descriptors, averaged over the sample history. The column labeled "Simple F-statistic" gives, for each descriptor j, the F-statistic in the Type I regression z_{nt} (or, alternatively, $z_{nt} - \hat{\beta}_{nt} z_{Mt}$) $= f_1 + f_j w_{jnt} + u_{nt}$. This F-statistic measures the significance of the factor, taken by itself, with no other extra-market components in the regression. The column labeled "Incremental F-statistic" gives the F-statistic of the factor in the Type I regression with all factors included. This F-statistic measures the importance of the additional risk attributable to this factor, over and above that which may be attributed to other factors. In each case, the figure given is the average F-statistic over the sixty quarters. The 99 percent confidence point for this average is 1.47, so that almost all descriptors are significant, both singly and incrementally.

After these quarterly regressions were completed, those of the descriptors that were the least significant in explaining beta and as components of risk were deleted, and 180 monthly regressions were used

TABLE 4

SUMMARY STATISTICS ON INDIVIDUAL FACTORS

Descriptors	Explanation of z_{nt}: Average Simple F	Average Incremental F	$\sqrt{\hat{F}^*_{jj}}$	Explanation of $(z_{nt} - \hat{\beta}_{nt} z_{Mt})$: Average Simple F	Average Incremental F
HBETA	33.06	4.74	1.68	16.21	2.89
PAY	16.25	2.33	.43	8.81	2.33
SIZE	18.98	2.41	.54	12.51	2.44
LIQ	3.87	1.69	.25	4.02	1.55
CUT	5.80	0.83		4.85	0.83
LEV	7.46	1.45	.40	6.54	1.44
NSG	3.61	1.75		3.67	1.75
PLANT	11.46	1.78	.34	6.55	1.81
STO	24.13	3.73	1.02	13.74	3.85
LNP	22.56	3.79	.69	18.34	3.57
YIELD	17.81	2.12	.62	13.04	2.13
ΔE	6.40	1.77	.30	5.77	1.81
σE	11.55	5.50	.48	10.20	5.60
σ(ΔE)	28.71	1.98		16.05	1.98
B/P	12.88	2.57	.42	12.02	2.28
AB	7.60	2.07	.29	5.60	2.05
E/P	7.37	1.41		5.60	1.41
σ(E/P)	22.31	1.83	.51	12.08	1.84
PRNEG	13.63	2.15	.46	9.60	2.15
ROI	7.41	1.45		2.33	1.48
EGROW	3.56	1.71	.70	3.61	1.71
SIGMA	32.91	2.00	.46	18.09	1.83
BBETA	6.19	1.41		6.40	1.40
QUAL			.06		

Industry	Explanation of z_{nt}: Average Simple F	Average Incremental F	$\sqrt{\hat{F}^*_{jj}}$	Explanation of $(z_{nt} - \hat{\beta}_{nt} z_{Mt})$: Average Simple F	Average Increment. F
MISCCP	1.27	1.24		1.32	1.21
MINING	3.11	2.82	2.62	3.20	2.61
D.OILS	4.88	3.92	2.59	4.88	3.61
I.OILS	2.57	1.92	2.65	2.01	1.52
FOODSA	2.44	1.99	1.10	2.09	1.74
APAREL	3.73	3.14	1.45	3.77	3.04
PAPERA	1.93	2.24	1.90	2.09	2.21
MEDIAS	2.02	2.13	2.27	1.81	2.02
CHMCLS	2.26	1.48	1.20	2.38	1.48
DRUGSA	3.66	2.83	2.02	3.77	2.88
RUBBER	2.10	2.64	2.22	2.19	2.41
CHATLS	3.20	1.68		3.48	1.66
CNTANR	1.10	1.48		1.11	1.38
ISTEEL	3.52	2.26	2.12	3.58	2.07
OF.M/C	2.45	1.97		2.13	1.97
NE.M/C	4.42	2.06	1.31	3.50	2.01
EEQUIP	3.25	2.15	1.66	3.31	2.18
ELECTR	6.76	3.78	3.32	5.54	3.74
H.HOLD.	2.60	2.47	1.83	2.38	2.46
AUTOEQ	3.20	2.28	1.52	3.23	2.28
AEROSP	5.37	3.59	3.36	5.64	3.57
INSTRM	2.08	1.61	1.77	1.97	1.57
LEISUR	2.73	1.66	2.96	2.07	1.64
AIRTRN	10.51	6.69	5.12	9.55	6.67
TOBACO	2.36	2.37	2.50	2.48	2.42
RETAIL	3.35	3.08	1.62	3.34	3.02
FSTORS	3.46	3.06	1.95	3.32	2.82
MISCBS	1.45	1.50		1.48	1.50
FINANC	2.03	2.36	2.69	2.08	2.36
BEVRAG	1.63	1.72	1.65	1.76	1.66
NFMETL	2.80	2.31	2.16	3.01	2.19
UTLITY	18.63	2.92	2.31	11.46	2.25

to generate unbiased estimates of F by the method of equation (27). The estimated standard deviation of *monthly percent return* associated with each factor is given in the column of Table 4 labeled $\sqrt{\hat{F}^*_{jj}}$. For the accounting- and market-based descriptors other than HBETA and QUAL, these figures are to be interpreted as the standard deviation of the difference in return to be observed between two securities (or two portfolios) differing by one standard deviation in their (average) values on that descriptor, but possessing identical (average) values for all other descriptors. The figure for HBETA gives the standard deviation associated with a difference of one in (average) HBETA, and the figure for QUAL gives the standard deviation for a difference of one in (average) QUAL (e.g., equal to the difference between an A+ and an A rating). The figures for industries give the standard deviation of monthly percentage return associated with a security (or portfolio) in that industry, relative to the return of a market portfolio made up of miscellaneous firms but identical in other characteristics.

For example, a portfolio with average share turnover (STO) differing from the market portfolio by one standard deviation would have, as a result, a component of variance in monthly percentage portfolio return of $(1.02)^2$, leading to a standard deviation in monthly percentage return of 1.02 if the portfolio were identical to the market portfolio in all other characteristics. These figures give the variance of each separate factor. In order to define the variance for any portfolio, which will differ from the market on numerous descriptors, it is necessary to know not only the variances of the various factors, but also the covariances among them, so

Tables 2, 3, and 4 survey the magnitude of extra-market covariance in security returns in the past fifteen years insofar as extra-market covariance was transmitted through factors having associated loadings that were linear in the descriptors. Our general impression, after having examined the results summarized in these tables, would not surprise a practicing portfolio manager in the least. It appears that, regardless of the number of securities included in a portfolio, the elimination of portfolio risk relative to the market requires that the portfolio look like the market--that is, that the average values for the portfolio on the accounting- and market-based descriptors and the percentages of portfolio value invested in the various industry groups be similar to the market portfolio.

APPENDIX

This appendix provides details not given in the text about the descriptors. The descriptors are defined either directly from the data or through the operators defined below. {m} denotes annual data item number m, as defined in the *Compustat Manual*, Section 9, for the n^{th} firm in the t^{th} year unless otherwise specified, the firm and year subscripts being dropped when not confusing. A data item that was not available in the source or that was detected as obviously spurious by a screen was flagged as not available. The response to a nonavailable item is specified below through the operators. For descriptors based on a number of years of data, if more than one observation was missing, the descriptor was flagged as not available. The descriptors were assumed to be constant over each calendar quarter and were computed from the latest annual data available at the beginning of the quarter as detailed in the text. In a particular quarter, if the data for any descriptor (except quality rating) or any monthly return on a security are not available, the security was dropped from the sample in that quarter.

Operators Used in Constructing the Descriptors

X_s exists unless flagged "not available"

$$\Delta X \equiv \Delta_t(X) = 2.0(X_t - X_{t-1})/(|X_t|+|X_{t-1}|): \quad \Delta_t X \text{ exists if}$$

X_t and X_{t-1} exist and are not both = 0.

$$M_t(X) = \sum_{\substack{s=t-4 \\ X_s \text{ exists}}}^{t} X_s \Big/ L_t(X), \text{ where } L_t(X) = \sum_{\substack{s=t-4 \\ X_s \text{ exists}}}^{t} 1.0 \,.$$

$$\sigma_t(X) = U_t[X - M_t(X)], \text{ where } U_t(Y) = \left[M_t(Y^2)/[L_t(Y) - 1]\right]^{1/2}.$$

A smoothed variable, H, is defined by the algorithm

$$W_0 = 0 \quad , \quad W_t = .8W_{t-1} + 1 \qquad t = 1,2,\ldots$$

$$H_0(X) = 0 \quad , \quad H_t(X) = [.8H_{t-1}(X) + X_t]/W_t \qquad t = 1,2,\ldots$$

When X_t is missing in a period, the missing item is deleted from the sum, and the divisor is adjusted accordingly.

Basic Variables Used in Defining the Descriptors

J = adjusted number of common shares outstanding = {25}{27}

A = total assets = {6}

B = adjusted book value of common equity per common share = {11}/J

D = adjusted dividends per common share = {26}/{27}

E = adjusted earnings per common share = {20}/J

O = adjusted operating income per common share = {13}/J

F = adjusted fixed charges per common share = {15}/J

P = adjusted closing price per common share = {24}/{27}

Descriptors

(1) $$PAY_{nt} = M_{nt}(D)/M_{nt}(E)$$

(2) $$SIZE_{nt} = \log[M_{nt}(A)]$$

(3) $$LIQ_{nt} = \begin{Bmatrix} \text{quick} \\ 1 \end{Bmatrix} \text{ if } \begin{cases} \text{quick} < 1 \\ \text{quick} \geq 1 \end{cases}, \text{ where quick} = (\{1\} + \{2\})/\{5\}$$

(4) $$CUT_{nt} = M_{nt}(\Delta D^-), \text{ where } \Delta D^- = \begin{cases} |D_t - D_{t-1}| & \text{if } D_t - D_{t-1} < 0 \\ 0, \text{ otherwise} \end{cases}$$

(5) $$LEV_{nt} = M_{nt}(v), \text{ where } v = (\{5\} + \{9\} + \{10\})/\{6\}$$

(6) $$NSG_{nt} = EGROW_{nt} - (1 - PAY_{nt})H_{nt}(E/P)$$

(7) $$PLANT_{nt} = \{7\}/\{6\}$$

(8) $$STO_{nt} = \{28\}/\{25\}$$

(9) $$LNP_{nt} = \ln\{24\}$$

(10) $$YIELD_{nt} = M_{nt}(D)/M_{nt}(P)$$

(15) $$A\beta_{nt} = \left(\sum_{\substack{s=1 \\ E_{ns} \text{ exists}}}^{t} \Delta E_{ns} \Delta E^n_{Ms} \right) \Bigg/ \left(\sum_{\substack{s=1 \\ E_{ns} \text{ exists}}}^{t} \Delta E^{n\,2}_{Ms} \right)$$

where E^n_{Ms} = earnings for U.S. corporations as reported by the Bureau of Labor Statistics, for the fiscal year of firm n.

(18) $$PRNEG_{nt} = Z[(F_{nt}-O_{nt})/\sigma O_{nt}], \text{ where } Z(X) = \text{Prob}[\eta(0,1) < X]$$

(19) $$ROI_{nt} = \frac{E_{nt} - E_{n,t-5}}{M_{n,t-1}(E-D)}$$

(20) $$EGROW_{nt} = \frac{M_t(E_s - E_{s-1})}{M_t(E_t)}$$

(21) $$SIGMA_{nt} = \left[\sum_{\substack{s=t-60 \\ z_{ns} \text{ exists}}}^{t} (\dot{z}_{ns} - H\beta_{nt}\dot{z}_{Ms})^2/58 \right]^{1/2}$$

where the "." indicates a deviation from the mean of this sample.

(22) $$BBETA_{nt} = \left(\sum_{s \geq t-10}^{t} E^*_{ns} E^*_{Ms} \middle/ \sum_{s \geq t-10}^{t} {E^*_{Ms}}^2 \right) \quad \text{where}$$

$$E^*_{nt} = \frac{M_{nt}(E)(1+2EGROW_{nt})}{P_{nt}} \quad , \quad E^*_{Mt} = \frac{M_t(E^n_M) + 2M_t(E^n_{Ms} - E^n_{M,s-1})}{P_{Mt}}$$

(23) $$HBETA_{nt} = \left(\sum_{s=t-60}^{t} \dot{z}_{ns}\dot{z}_{Ms} \right) \Big/ \left(\sum_{s=t-60}^{t} \dot{z}_{Ms}^{\,2} \right)$$

(24) $$QUAL_{nt} = \begin{Bmatrix} 9 \\ 8 \\ 7 \\ 6 \\ 5 \\ 4 \\ 3 \\ 2 \\ 0 \end{Bmatrix} \text{ if S\&P quality rating} = \begin{Bmatrix} A+ \\ A \\ A- \\ B+ \\ B \\ B- \\ C+ \\ C \\ \text{Other} \end{Bmatrix} \text{ in December of year } t-1$$

(25) $$DQUAL_{nt} = \begin{cases} 1 & \text{if } QUAL_{nt} = 0 \\ 0 & \text{otherwise} \end{cases}$$

REFERENCES

[1] Beaver, William; Kettler, Paul; and Scholes, Myron. "The Association Between Market Determined and Accounting Determined Risk Measures." *The Accounting Review* 45 (October 1970): 654-682.

[2] Cohen, Kalman J., and Pogue, Jerry A. "An Empirical Evaluation of Alternative Portfolio Selection Models." *Journal of Business* 40 (April 1967): 166-193.

[3] Farrar, D. E. *The Investment Decision Under Uncertainty*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1962.

[4] King, Benjamin F. "Market and Industry Factors in Stock Price Behavior." *Journal of Business* 39 (January 1966): 139-190.

[5] Lintner, John. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics* 47 (February 1965): 13-37.

[6] Morrison, Donald F. *Multivariate Statistical Methods*. New York: McGraw-Hill, Inc., 1967.

[7] Rosenberg, Barr. "Linear Regression with Randomly Dispersed Parameters." *Biometrika* 60 (April 1973): 65-72.

[8] ________. "A Survey of Stochastic Parameter Regression." *Annals of Economic and Social Measurement* 2 (October 1973):

[9] Rosenberg, Barr, and McKibben, Walt. "The Prediction of Systematic and Specific Risk in Common Stocks." *Journal of Finance and Quantitative Analysis* 8 (March 1973): 317-333.

[10] ________. "The Prediction of Systematic and Specific Risk in Common Stocks." Working Paper No. 12, Research Program in Finance. Berkeley, Calif.: Institute of Business and Economic Research, University of California, 1972.

[11] Rosenberg, Barr, and Houglet, Michel. "CRSP and Compustat Stock Price Data Bases: Error Rates and Suggested Filters." Working Paper No. 13, Research Program in Finance. Berkeley, Calif.: Institute of Business and Economic Research, University of California, 1973.

[12] Rosenberg, Barr, and Ohlson, James. "Stochastic Variation in Security Risk." (Manuscript in preparation.)

[13] Sharpe, William F. *Portfolio Theory and Capital Markets*. New York: McGraw-Hill, Inc., 1970.

[14] ________. "The Capital Asset Pricing Model: A 'Multi-Beta' Interpretation." Paper presented at the Sixth Annual Wells Fargo Conference on Capital Market Theory and Evidence, San Francisco, July 1973.