This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Expectations and the Structure of Share Prices

Volume Author/Editor: John G. Cragg and Burton G. Malkiel

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-11668-9

Volume URL: http://www.nber.org/books/crag82-1

Publication Date: 1982

Chapter Title: Valuation Models and Earnings Growth

Chapter Author: John G. Cragg, Burton G. Malkiel

Chapter URL: http://www.nber.org/chapters/c11290

Chapter pages in book: (p. 97 - 134)

3 Valuation Models and Earnings Growth

The growth-rate expectations we have been investigating were formed by security analysts to aid them in judging the worth of common stocks. In addition to the questions about the accuracy, agreement, and completeness of the analysts' forecasts that we investigated in chapter 2, these data prompt two further queries: How are expectations of the future growth of earnings related to the value of common stock? Do security analysts' forecasts represent the expectations that actually affect the prices of securities?

Any consideration of these questions requires the use of some theory of the formation of the prices of common stocks in which the expected growth rates of earnings play an important role. The theory should also encompass explicitly major features of the expected growth rates that we uncovered in examining these data. We especially need to recognize in the theory the diversity of expectations that was so evident in our data.

3.1 The Role of Security-Valuation Models

The obvious theory to begin with is the capital-asset pricing model (CAPM) developed initially by Sharpe (1964), Lintner (1965), and Mossin (1966). This theory is now widely used as an approach to security valuation. It provides explicit hypotheses about the relative structure of security prices based on expectations about the returns to investors in securities, and it is now the dominant valuation model.

The CAPM has a number of theoretical and empirical weaknesses. Ross (1976, 1977) has developed a different model based on arbitrage considerations. This arbitrage pricing theory (APT) uses a less general assumption about the nature of security returns than the CAPM, but this limitation is balanced by the less restrictive assumptions about investor behavior adopted by the APT. The key assumption for the APT concerns the nature of returns—an assumption we describe below—and not the arbitrage considerations. We can obtain virtually the same conclusions about valuation as can be obtained with APT by using an alternative argument concerning diversification together with the central assumption of the APT about returns. In our argument, we assume that utility-maximizing behavior leads a number of investors to choose portfolios which are extensively diversified. The occurrence of such diversification then turns out to imply a particular structure of security prices. This alternative argument has the advantage of being somewhat less sensitive than other asset-pricing theories to variations in some other key assumptions. At the same time, if the CAPM should actually be the appropriate theory, then it gives the same results as the diversification theory provided that returns do follow the assumed pattern.

None of these theories explicitly relates expectations of returns to characteristics of the issuers of securities, or to investors' perceptions of them, including the growth rates of their earnings and dividends. Instead, the models concentrate on the returns to be received directly by investors. These returns include payments made by the issuers of securities—in particular, dividends. The returns also include changes in the market values of securities, that is, capital gains. Current market prices are determined in the valuation theories by expectations about future returns. Presumably market prices in the future will also be determined by the expectations that will be held at the time of that valuation as well as by the operation of the valuation process in the future. In order to judge the value of securities today, investors should logically concern themselves with predicting future expectations.

Most models ignore this aspect of the logical structure of valuation. Indeed, they are usually expressed in a rate-of-return form which hides the implicit problems created by the model's requiring expectations of how the model itself will operate later. The major explicit exceptions occur in papers by Stapleton and Subrahmanyam (1978) and by Ohlson and Garman (1980). The valuation model developed in this chapter will wrestle explicitly with the problems of expectations about future market valuation. We relate future expected prices to expectations of the growth of earnings and find that the price-determination process is internally consistent.

3.2 A Diversification Model of Security Valuation

3.2.1 The Basic Model

Heuristically, we start from the insight that prompted Sharpe (1964) to develop the CAPM. Sharpe's insight was that variations in the returns to different securities are related to each other through common dependence on a small number of events or variables.[1] Any risks specific to the security can be removed by diversification. Portfolios typically are diversified, but even investors with very widely diversified portfolios, containing only small holdings of very many securities, do not avoid the risks coming from the common elements. Investors must therefore be receiving appropriate rewards in order for them to be willing to bear these risks. There is no need to reward investors for bearing the risks specific to individual securities since such risks can be "diversified away." Security returns (and so security prices) must exhibit certain patterns in order that the appropriate rewards should exist in the sense that bearing systematic risk does receive a premium but bearing specific risk does not. We will argue that the key element in this description is that investors' portfolios may be considered to be extensively diversified. If this is the case, security prices must exhibit a certain type of pattern.

The situation which we consider may be characterized as follows:

a) There are J types of marketable securities. Associated with each security is a one-period total return of r_j (dollars) per unit of security, which accrues to the holders of the security. The (proportionate) rate of return over the period is π_j, defined as $\pi_j = r_j/p_j - 1$, where p_j is the price of the security.

b) The returns may be treated as random variables. The critical assumption for the diversification argument, as for the Ross APT, is that these random variables have the following structure:

$$r_j = \mu_j + \sum_{k=1}^{K} \gamma_{jk} f_k + e_j . \tag{3.2-1}$$

Here μ_j is a fixed parameter giving the expected return of security j, $j = 1, \ldots, J$. The f_k are K random variables, each with mean zero.[2] It is assumed that $k < J$. In fact, we would expect there to be very few factors while there are a large number of securities. The γ_{jk} are coefficients. The e_j are random variables, independent of the f_k and of the other e_m, $m \neq j$. In statistical terms, this is the factor-analysis model, which we used for other purposes in chapter 2; but here we are not assuming that the f_k are normally distributed or necessarily that they are independent of each other.

1. It is easy to overlook the fact that Sharpe's model does not really conform to this insight since he does not assume any particular structure for the covariance matrix of returns. The insight is also suggested in Myers (1968), from which the diversification model might be derived.

2. Insofar as the factors do not have mean zero, their expectations multiplied by the coefficients γ_{jk} are incorporated in the expected returns μ_j.

Use of this structure in the present context is intended to capture the hypothesis that returns are related only through a limited number of common factors and are otherwise independent of each other. These factors may be considered to include movements in the general stock market, in economic activity, in the rate of inflation, and so on. Equation (3.2-1) also embodies the highly restrictive assumption that the dependence of returns on the common factors is linear.[3] As we shall see, this restrictive and rather arbitrary aspect of the assumption is important to the results to be obtained and can be relaxed only to a limited extent. Note that (3.2-1) is expressed for returns rather than rates of return. We consider the alternative in section 3.2.2.

c) There are a number of investors, indexed by i, who choose their holdings of securities v_{ji}, $j = 1, \ldots, J$, so as to maximize the expected utility of end-of-period returns $R_i = \sum_{j=1}^{J} r_j v_{ji}$. In doing so, they are aware of the process by which returns are generated, given by (3.2-1). Each investor takes the prices of securities as given, unaffected by his own decisions. In carrying out the maximization, the investor is constrained by his initial holdings of securities $\bar{v}_{ji}$; that is, his budget constraint is

$$(3.2\text{-}2) \qquad \sum_{j=1}^{J} p_j v_{ji} = \sum_{j=1}^{J} p_j \bar{v}_{ji}.$$

Thus, letting U_i represent the investor's utility function, the investor's portfolio choice can be considered to maximize

$$(3.2\text{-}3) \qquad E(U_i(R_i)) = E\left[U_i\left(\sum_{j=1}^{J} r_j v_{ji}\right)\right] = E\left[U_i\left(\sum_{j=1}^{J}\left(\mu_j + \sum_{k=1}^{K} \gamma_{jk} f_k + e_j\right) v_{ji}\right)\right]$$

subject to constraint (3.2-2). As usual, the investor's utility function U_i is assumed to be monotonically increasing, with continuous first derivatives, and concave (so that the investor is risk-averse).

d) As a result of maximizing their utility functions given the prices that they observe in the market, at least some of these utility-maximizing investors hold extensively diversified portfolios in the sense that the holding of each of these investors of a number of securities is sufficiently small that each one regards his exposure to the specific risks of each of these securities—that is, to the risks coming from the random variations in their e_j's—as being trivial. Explicitly, we shall consider a portfolio to be extensively diversified with respect to a security when the investor's marginal disutility of the exposure to the specific risk of the security can be considered negligible relative to the marginal disutility of the system-

3. A discussion of the nature of the factors can be found in Roll and Ross (1980). The common-factor structure (3.2-1) was earlier used in a valuation model by Fama (1971), though he assumed all random quantities had stable distributions.

atic risk exposure to the common factors. We shall correspondingly define his holding of such a security as being extensively diversified.

The investor's choice can be characterized by the first-order conditions for maximization of (3.2-3) subject to (3.2-2), namely,

$$E(r_j U_i') - \lambda_i p_j = \mu_j E(U_i') + \sum_{k=1}^{K} \gamma_{jk} E(f_k U_i') + E(e_j U_i') - \lambda_i p_j = 0, \tag{3.2-4}$$

where $U_i' = dU_i(R_i)/dR_i$; λ_i is the Lagrange multiplier of the constrained maximization problem, and it has the standard interpretation of being the marginal utility of wealth. Dividing (3.2-4) by λ_i yields

$$p_j = \mu_j a_{0i} + \sum_{k=1}^{K} \gamma_{jk} a_{ki} + E(e_j U_i')/\lambda_i. \tag{3.2-5}$$

Note that the coefficients a_{ki} are the same for all securities j.

All terms of (3.2-5) depend on the overall variability of return since they arise from taking the expectations of U_i' and its covariances with other random variables. Within this overall dependence, we may regard the first term on the right-hand side of (3.2-5) as giving the marginal effect of variation in the holding of security j through the expected return of this security. The second term reflects the marginal effect arising through variation with the various common factors of the return of the jth security. The final term reflects the marginal specific risk being introduced into the portfolio through security j. These last two terms reflect the effect of security j on the risk of the portfolio since they arise only from the concavity of the utility function and from the stochastic nature of returns.

If we denote the distribution of the random variables as $G(\cdot)$, the numerator of the last term in (3.2-5) can be written as

$$E(e_j U_i') = \int e_j \left\{ dU_i \left[\sum_{j=1}^{J} \left(\mu_j + \sum_{k=1}^{K} \gamma_{jk} f_{jk} + e_j \right) v_{ji} \right] / dR_i \right\} \times dG(f_1, \ldots, f_k, e_1, \ldots, e_j). \tag{3.2-6}$$

Since U_i is concave, for any given values of the other random variables U_i' is monotonically decreasing (increasing) in e_j as v_{ji} is positive (negative). Since we assume that e_j is independent of other random variables and has mean zero, the expectation in (3.2-6) therefore has a sign opposite to that of v_{ji}. The continuity of U_i' in R_i ensures that (3.2-6) is continuous in v_{ji}. The last term of (3.2-6) can thus be made arbitrarily small by choosing v_{ji} small enough.

One might also presume that the concavity of the utility function would mean that the absolute size of (3.2-6) increases with the absolute value of v_{ji}. This ignores the effect of changes in v_{ji} on (3.2-6) through changes in exposure to other factors. In other words,

$$\partial(e_j U_i')/\partial v_{ji} = U_i'' e_j^2 + U_i'' e_j \mu_j + U_i'' e_j \sum_{k=1}^{K} \gamma_{jk} f_k, \tag{3.2-7}$$

and the intuitive argument applies strictly to the first term. Suitable additional assumptions would need to be made if it were to apply to the expectation of (3.2-7). The argument does apply to $\partial^2 E(U_i)/\partial v_{ji}^2 = E(U_i'' r_j^2)$.

In contrast to the final terms of (3.2-6), the coefficients a_{ki} of the γ_{jk} do not depend primarily on the amount of jth security chosen. Instead, $a_{ki} = E(f_k U_i')/\lambda_i$ depends on how $U_i'(R_i)$ varies with f_k. If we assume that the f_k are independent of each other,[4] this correlation depends primarily on $\Sigma_{h=1}^{J} \gamma_{jk} v_{hi}$, which may be regarded as the total exposure of the portfolio to f_k, and not just on the exposure coming from the jth security $\gamma_{jk} v_{ji}$. Indeed, continuing to assume that the f_k are independent of each other, if $\Sigma_{h=1}^{J} \gamma_{hk} v_{hi} = 0$, then U_i' does not vary with f_k and a_{ki} is zero in (3.2-5) even though $\gamma_{jk} v_{ji}$ may be nonzero. This contrast between the last two terms of (3.2-5) reflects the assumption that the investor is only vulnerable to fluctuations in e_j through his holding of security j while he is vulnerable to the common factors through all his holdings.

Our central hypothesis about the extent of diversification states that the first of the risk terms in (3.2-5), $\Sigma_{k=1}^{K} \gamma_{jk} a_{ki}$, is large relative to the last term. Indeed, we assume that v_{ji} is so small that the last term can be ignored, but that the exposures to risk from the common factors, $\Sigma_{j=1}^{J} \gamma_{jk} v_{ji}$, are not so small that the a_{ki} can be overlooked. With such diversification, equation (3.2-5) can be approximated by

$$p_j = \mu_j a_{0i} + \sum_{k=1}^{K} \gamma_{jk} a_{ki}. \tag{3.2-8}$$

The price of the security may be treated as a linear function of its expected return and its own coefficients for each of the common factors. As expressed in equation (3.2-8), it might seem that the coefficients of the linear function are specific to the investor i, but this cannot be the case if diversification is widespread.

Suppose that the number of securities being held by some investor i for which (3.2-8) holds exceeds $(K + 1)$, the number of factors plus one. Then the price of these securities must be determined in such a way that they obey approximately a linear equation, namely, (3.2-8). Similarly, suppose that another investor, say, h, has extensively diversified holdings (that is, values of v_{jh} which are sufficiently small that [3.2-8] gives a close

4. The specification does not require independence of the f_k, and this assumption is used here only to aid heuristic interpretation of the model. Note that if the f_k are not independent, the a_{ki} will not necessarily be zero when $\Sigma_{h=1}^{J} \gamma_{hk} v_{hi} = 0$. The assumption is usually made in the statistical factor-analysis model to obtain identifiability of the γ_{jk}; but as we explain in section 3.4, our empirical implementation of the model will be along somewhat different lines.

approximation to [3.2-5]) of at least $(K + 1)$ of these securities. Then the coefficients for investor h, a_{kh} $(k = 0, \ldots, K)$, must be the same as those for investor i, and the prices of all the securities of which his holdings are extensively diversified must also obey the same linear equation (3.2-8). Furthermore, if a third investor has extensively diversified holdings of at least $(K + 1)$ securities of which at least one of the other two investors has extensively diversified holdings (even if neither holds $[K + 1]$ of these securities himself), then the coefficients for the third investor must be the same as those for the other two, and any other securities of which his holdings are extensively diversified will have prices which exhibit the same linear structure.

More generally, then, if investors diversify extensively and if there is sufficient overlap among investors in the securities of which their holdings are extensively diversified, we can proceed sequentially in the way outlined to build up a set of securities found in the extensively diversified parts of some portfolios, all of which must (approximately) obey equations of the form

$$p_j = \mu_j a_0 + \sum_{k=1}^{K} \gamma_{jk} a_k . \tag{3.2-9}$$

Equation (3.2-9) expresses the main conclusion of the diversification model about security prices. Implicitly it claims that the market determines prices for the characteristics of expected return and association with the common factors but gives a zero value to specific risk. That this pattern of prices must exist is inferred from the "stylized fact" of reasonably wide diversification by some investors, given the other assumptions. It is worth noting that the argument inferring the pattern of prices does not turn on the process of equilibrium price determination, but on the pattern of demand of some investors that is assumed to be observed.

Returning to equation (3.2-5), we can say that it will be the case that any investor having extensively diversified holdings of $(K + 1)$ securities to which (3.2-9) applies must have extensively diversified holdings of all of them because the final terms of (3.2-5) must be negligible for all these securities. Therefore the argument concerning the overlapping sequences of diversified holdings may seem needlessly complicated. However, the argument remains applicable even when other considerations prevent investors from having extensively diversified but nonzero holdings of all securities to which (3.2-9) might apply. It is indeed the ability of this theory to withstand changes of asumptions and restrictions that makes it appealing.

3.2.2 Variations in Investors' Opportunities

The diversification argument is insensitive to many institutional considerations to which other valuation arguments are vulnerable. It holds up well to a prohibition on negative holdings of securities, to significant costs of short-selling, or to costs that limit the number of securities that it is practical or desirable to hold.

It is highly questionable whether allowing negative holdings on the same terms as positive ones well approximates the situation that investors do face. Although short-selling goes part way to providing for the assumed opportunities, it does not really offer them fully. As Miller (1977) and Jarrow (1980) point out, there are significant costs and foregone opportunities in short-selling as well as institutional impediments that are not trivial.

A prohibition of negative holdings does not alter the form of the first-order conditions for securities of which a positive amount is held. Formally the investor may be regarded as maximizing (3.2-3) subject to (3.2-2) and the constraint

$$v_{ji} \geq 0; \qquad j = 1, \ldots, J. \tag{3.2-10}$$

The first-order conditions (3.2-4) now become

$$\mu_j E(U_i') + \sum_{k=1}^{K} \gamma_{jk} E(f_k U_i') + E(e_j U_i) + \psi_{ji} - \lambda_i p_j = 0. \tag{3.2-11}$$

The additional term ψ_{ji} is the non-negative Kuhn-Tucker multiplier arising from constraint (3.2-10). It has the usual property that $\psi_{ji} v_{ji} = 0$. Hence, for securities for which $v_{ji} > 0$, (3.2-10) reduces to (3.2-4) and the diversification argument continues to hold for small positive holdings. The prices of these securities therefore conform with the relationship (3.2-9). Thus, if utility-maximizing decisions yield diversified portfolios when negative holdings are not allowed, the earlier conclusion about prices applies: the prices of securities which investors do hold in diversified portfolios must exhibit the linear structure shown in (3.2-9).

We saw earlier that $E(e_j U_i') = 0$ when $v_{ji} = 0$. For securities for which (3.2-10) is binding, we can rewrite (3.2-11) as

$$\mu_j a_{0i} + \sum_{k=1}^{K} \gamma_{jk} a_{ki} = p_j - \psi_{ji}/\lambda_i \leq p_j; \tag{3.2-12}$$

that is, securities which are not held by investor i are considered to be overvalued by him. If the investor does have extensively diversified holdings of at least $(K + 1)$ securities for which (3.2-9) holds, then his values of a_{ki} equal the a_k for these prices and ψ_{ji} will be zero for all securities to which the diversification argument applies. Constraint (3.2-10) therefore only becomes interesting when other changes in the model are made.

A similar argument to the one advanced when short sales are prohibited applies when they are allowed but involve additional costs not incurred with positive holdings. Equation (3.2-5) remains appropriate for securities of which the investor holds positive amounts. If these amounts are also small, the final term is again insignificant, and so (3.2-9) still must hold approximately.

Costs of various sorts may be associated with the number of different types of securities held. If these diversification costs are sufficiently small, the investor will be able to diversify extensively enough to eliminate virtually all systematic risk. Given the list of securities held by a utility-maximizing investor, the first-order conditions (3.2-4) apply to these securities and the structure of prices (3.2-9) must also obtain approximately if extensive diversification occurs. The argument developed in the previous section about overlapping sets of securities becomes particularly pertinent in this situation. With costs for the number of securities held, we would not expect investors actually to hold all securities for which (3.2-9) might hold. Instead, they will only proceed to the point where the specific risk of securities held can be regarded as small. Nevertheless, securities connected to each other through some sequence of extensively diversified holdings of the sort described in section 3.2-1 will all obey the same equation even though any particular portfolio may include only a small proportion of them. As diversification costs become important, however, the specific-risk terms will also increase in importance in (3.2-5) as diversification becomes less extensive and so the appropriateness of the approximation (3.2-9) becomes less apparent.

We have been considering investors' utility functions to be defined over end-of-period return. Investors undoubtedly do have wider concerns. It is hard to believe that investors' investment horizons are not multiperiod or that their utility functions depend only on the returns of their portfolio of securities. Fortunately, our conclusions are not sensitive to allowing the utility function also to depend on other circumstances or variables that may be correlated with the values taken on by "systematic" factors. In this respect, the model easily overcomes the difficulties that Fama (1970) showed could arise for valuation models from their being embedded in a wider and multiperiod context. We may indeed think of the factors as representing the general economic conditions that will affect investors' investment and consumption opportunity sets and their other income opportunities as well as the returns to securities.

We can readily expand the investor's utility function in (3.2-3) to include other choice variables and dependence on other random aspects of the state of the world. What matters is that the utility function depends on security selection only through the effect that the amount of the security held has on R_i, because the argument relies on assuming that $\partial E(U_i)/\partial v_{ji} = E[r_j \partial U_i/\partial R_i]$. Similarly the budget constraint may be a great deal more complicated than (3.2-2) provided that v_{ji} only enters it linearly, multiplied by p_j. When these conditions are met, equations (3.2-4) continue to express the first-order conditions for the security choices of a utility-maximizing investor. The other critical assumption needed for the diversification theory is that for some investors e_j is correlated with U_i' only through its inclusion in R_i so that the final term of (3.2-5) is zero when v_{ji} is zero for such investors who do include security j in the extensively diversified part of their portfolios. In other words, the second critical condition is that

$$E(e_j U_i') = 0 \quad \text{when} \quad v_{ji} = 0. \tag{3.2-13}$$

This condition will hold if e_j is independent of the other random variables that affect the investor since the $E(U_i' \mid v_{ji} = 0)$ does not depend on e_j and hence (3.2-13) will hold. Such a condition seems a reasonable interpretation of the model, with the f_k capturing correlations of returns not only among themselves but also with other things that generally affect investors. The model does not preclude some investors' being affected by variations in e_j directly, but it would then predict that such investors would not include security j in the extensively diversified part of their portfolios.

The diversification model is not insensitive to variations in the assumption (3.2-1) made about returns, though some alterations of this assumption can be made. While the assumption in (3.2-1) concerns the levels of return per share, the model could easily be expressed in terms of rates of return. Indeed, formally, if we divide equation (3.2-9) by $p_j a_0$ and rearrange, we obtain

$$\mu_j/p_j = 1/a_0 - \sum_{k=1}^{K} (a_k/a_0)(\gamma_{jk}/p_j). \tag{3.2-14}$$

Letting $\rho_j = \mu_j/p_j - 1 = E(\pi_j)$ and defining α_{jk} as (γ_{jk}/p_j), we can rewrite this equation as

$$\rho_j = b_0 + \sum_{k=1}^{K} b_k \alpha_{jk}. \tag{3.2-15}$$

The parameter $b_0 = 1/a_0 - 1$ could be considered to be the risk-free rate of interest, $\bar{\rho}$. Stated differently, $a_0 = 1/(1 + \bar{\rho})$. With the prices p_j being treated as fixed, either way of writing the model is appropriate.

We might have assumed from the outset that, instead of equation (3.2-1), the appropriate specification is

$$\pi_j = \rho_j + \sum_{k=1}^{K} \alpha_{jk} f_k + e_j. \tag{3.2-16}$$

Equation (3.2-15) might then appear to be a more appealing way of writing the valuation conclusion of the model than equation (3.2-9). Equation (3.2-14) can now be used to translate the model to equations (3.2-1) and (3.2-9). Specification (3.2-1) has the advantage over (3.2-16) that with it one can think readily of the determination of current prices by expected further return. Moreover, it makes clear that prices, rather than investors' expectations, can adjust to bring about the structure of prices and expected rates of return that is characteristic of the model. However, over time equation (3.2-16) may show more stability than expression (3.2-1) because, for example, unexpected capital gains might tend to arise primarily through proportional changes in μ_j and γ_{jk}. Equation (3.2-16) may therefore be a better basis for empirical work than assumption (3.2-1), and will be so used later.

The assumption we cannot easily relax is the linear structure assumed in (3.2-1) or (3.2-16). For instance, if r_j depends on a few factors through some nonlinear function, say, $r_j(f_1, \ldots, f_K, e_j)$, then the first-order conditions corresponding to (3.2-4) are

$$\text{(3.2-17)} \qquad \int r_j(f_1, \ldots, f_k, e_j) U_i'(R_i) dG(f_1, \ldots, f_K, e_1, \ldots, e_j) - \lambda p_j = 0.$$

These conditions do not in general seem to impose a useful structure on prices even though diversification does occur.

This limitation of the model may seem unfortunate. Valuation, as section 3.3 examines at more length, involves discounting returns over time. This discounting will tend to introduce rather complicated nonlinearities into returns even though the underlying stochastic structure of more basic elements is linear. The presumption of the diversification model is that a linear structure still yields an adequate approximation, or at least that investors act as if they believe it does.

The model can accommodate some forms of nonlinearity in (3.2-1). For example, a polynomial dependence of returns on the factors would simply introduce the coefficients of the polynomial separately into the vaulation equation (3.2-9). A more interesting case arises when the relationship is bilinear.

As we shall find in section 3.3, systematic variability of returns can be considered to arise from two sources. The dividends that the investor can expect will vary with general economic conditions as will the general market appreciation of the company, which may produce capital gains. We might consider the γ_{jk} to reflect this sensitivity. The second source of variability is that the market may change the value it places on these characteristics. Such changes would result in price changes and produce capital gains. That is, the a_k coefficients in (3.2-9) may change, and there is nothing in the diversification model to prevent such change. We might then assume two sets of K factors (now without mean zero), f_k and m_k, with means $\bar{f}_k$ and $\bar{m}_k$, respectively, where the second set of factors represents potential random changes in market valuation. Then we might replace (3.2-1) with

$$(3.2\text{-}18) \qquad r_j = \mu_j + \sum_{k=1}^{K} \gamma_{jk} f_k m_k + e_j$$

$$= \mu_j + \sum_{k=1}^{K} \gamma_{jk} \bar{f}_k \bar{m}_k + \sum_{k=1}^{K} \gamma_{jk} (f_k - \bar{f}_k)(m_k - \bar{m}_k)$$

$$- \sum_{k=1}^{K} \gamma_{jk} (f_k - \bar{f}_k)\bar{m}_k - \sum_{k=1}^{K} \gamma_{jk} (m_k - \bar{m}_k)\bar{f}_k + e_j.$$

The first-order conditions corresponding to (3.2-4) yield

$$(3.2\text{-}19) \qquad p_j = (\mu_j + \sum_{k=1}^{K} \gamma_{jk} \bar{f}_k \bar{m}_k) a_{0i}$$

$$+ \sum_{k=1}^{K} \gamma_{jk} \{E[(f_k - \bar{f}_k)\bar{m}_k U_i'] + E[(m_k - \bar{m}_k)\bar{f}_k U_i']$$

$$+ E[(f_k - \bar{f})(m_k - \bar{m})U_i']\}/\lambda_i + E(e_j U_i')/\lambda_i$$

$$= \mu_j' a_{0i}' + \sum_{k=1}^{K} \gamma_{jk} a_{ki}' + E(e_j U_i')/\lambda_i.$$

This is of the same form as (3.2-5), and diversification has the same implications for the final term as it had in (3.2-5). Therefore, while the exact interpretation of the a_{ki}' coefficients has changed, the structure (3.2-18) implies the same sort of price structure as does (3.2-1) when there is diversification. Equations (3.2-18) and (3.2-19) might suggest that we should be able to detect $2K$ factors in returns but that only K factor coefficients enter valuation. However, the factors are actually the $f_k m_k$, of which there are only K. The general form of (3.2-9) is also preserved when some factors in (3.2-18) enter in a bilinear way and other linearly, which would happen if some factors were not random, or if the same m factor applied to more than one of the γ_{jk} coefficients.

3.2.3 Diversity of Expectations among Investors

We have presumed in developing the diversification model that all investors know the values of the key parameters μ_j and γ_{jk}, or at least that they are all in agreement about what these values are.[5] This may not be a good approximation, especially when the model is to be used with data such as we examined earlier where differences among individual predictors were prominent. Diversity of perceptions about the values of μ_j and γ_{jk} would seem to strike at the basis of the theory, which has relied on there being a few common values for these parameters in order to obtain the structure of prices. However, diversity of expectations causes us difficulties mainly when negative holdings are prohibited or costly.

5. This assumption plays the role of the homogeneous expectations assumption in the usual CAPM. It is not strictly speaking necessary that all investors have the same subjective distribution of returns, as long as they do agree on the expected return and on the coefficients of the common factors in each security's return.

Suppose that each investor's perception of the process generating returns corresponds to equation (3.2-1), but that different investors do not use the same parameters so that investor i perceives return to be generated by

$$r_j = \mu_{ji} + \sum_{k=1}^{K} \gamma_{jki} f_k + e_{ji}. \tag{3.2-20}$$

The individual first-order conditions now give rise to

$$p_j = \mu_{ji} a_{0i} + \sum_{k=1}^{K} \gamma_{jki} a_{ki} + E(e_{ji} U_i') / \lambda_i. \tag{3.2-21}$$

If the portfolio is extensively diversified—again in the sense that the investor's holdings of at least some securities are sufficiently small that for these securities the last term of (3.2-21) is negligible—then the prices of such securities must obey a linear relationship in the parameters that the investor perceives.

Suppose another investor, m, also has extensively diversified holdings of more than $(K+1)$ of the securities that investor i holds in small quantities. Then the last term of (3.2-21) is negligible, and also for this investor we obtain

$$\mu_{ji} a_{0i} + \sum_{k=1}^{K} \gamma_{jki} a_{ki} = \mu_{jm} a_{0m} + \sum_{k=1}^{K} \gamma_{jkm} a_{km} = p_j. \tag{3.2-22}$$

By hypothesis, equation (3.2-22) applies to more than $(K+1)$ securities so that the relationship in (3.2-22) cannot come about simply by adjustment of the marginal utility coefficients a_{ki}. The investors' expectations must themselves be related to each other through their approximately obeying a linear relationship. To illustrate, if there is agreement on the γ_{jk}, equation (3.2-22) becomes

$$\mu_{jm} = \mu_{ji} a_{0i} / a_{0m} + \sum_{k=1}^{K} \gamma_{jk} (a_{km} - a_{ki}) / a_{0m} \tag{3.2-23}$$

for all securities for which the holdings of both investors are well diversified. This restriction on the μ_{jm} was not part of our original hypothesis about investors' expectations. What we have found is that the hypothesis of diversification across the same securities imposes a strong restriction on the extent to which the expectations of different investors can differ, at least with respect to the securities they hold.

The difficulty produced by diversification and diversity of expectations can also be expressed in terms of prices. Suppose that two investors have extensively diversified holdings of $(K+1)$ of the same securities. Then we can express the prices of any securities of which the second investor has an extensively diversified holding in terms of his expectations for that security and the valuation coefficients of the first investor. To see this, let V_i and V_m be the $(K+1) \times (K+1)$ matrices with rows made up of vectors

$\{\mu_{ji},\gamma_{j1i},\ldots,\gamma_{jKi}\}$ and $\{\mu_{jm},\gamma_{j1m},\ldots,\gamma_{jKm}\}$, for any $(K+1)$ securities held in extensively diversified portfolios by each. We can then solve (3.2-22) to obtain

$$\text{(3.2-24)}\qquad \{a_{0m},\ldots,a_{Km}\}' = V_m^{-1}V_i\{a_{0i},\ldots,a_{Ki}\}' = D_{mi}\{a_{0i},\ldots,a_{Ki}\}'.$$

The difficulty we noted is that D_{mi} must be the same no matter which securities are selected for the V matrices provided both investors do have extensively diversified holdings of them. Then we can write (3.2-21) as

$$\text{(3.2-25)}\qquad p_j = \{\mu_{jm},\gamma_{j1m},\ldots,\gamma_{jKm}\}D_{mi}\{a_{0i},\ldots,a_{Ki}\}'.$$

The overlapping portfolios argument is also applicable in this situation. If there is a matrix D_{nm} connecting investors n and m as in (3.2-24), then we can let $D_{ni} = D_{nm}D_{mi}$ even though n and i do not have diversified holdings of the same securities. Similarly, partial overlap of n with i and m could produce the constraint.

The problem we now face is to account for diversification in the face of diverse expectations. To do so, let us suppose that diversity of expectations among investors is captured by

$$\text{(3.2-26)}\qquad \mu_{ji} = \mu_j + \epsilon_{ji} \quad\text{and}\quad \gamma_{jki} = \gamma_{jk} + \eta_{jki},$$

where the additional terms ϵ_{ji} and η_{jki} can be treated as random variables having mean zero and being independent across investors. Condition (3.2-22) now becomes

$$\text{(3.2-27)}\qquad (\mu_j + \epsilon_{ji})a_{0i} + \sum_{k=1}^{K}(\gamma_{jk} + \eta_{jki})a_{ki} = (\mu_j + \epsilon_{jm})a_{0m} + \sum_{k=1}^{K}(\gamma_{jk} + \eta_{jkm})a_{km}.$$

The parameters μ_j and γ_{jk} may represent the true parameters of the process that is generating returns, but need only be the average of the investors' perceptions. The most straightforward way for (3.2-27) to hold is for the random terms to be equal to each other so that diversity of expectations between investors about security parameters is actually negligible for the diversified parts of their portfolios. The most reasonable way for this to occur is for the random quantities to be zero. With random differences among investors which have mean zero, we can presume that an investor's expectations for many securities will be close to the mean expectation and that these securities will be the ones of which his holdings will be extensively diversified. In that case, the coefficients a_{0i} and a_{ki} must also be the same for all investors. In turn, prices will then have to obey (3.2-9).

To develop the argument more fully, rewrite (3.2-21) as

$$E(e_j U_i')/\lambda_i = p_j - (\mu_j + \epsilon_{ji})a_{0i} - \sum_{k=1}^{K} (\gamma_{jk} + \eta_{jki})a_{ki}. \tag{3.2-28}$$

As noted earlier, the left-hand side of this equation is zero when $v_{ji} = 0$ and otherwise has sign opposite to v_{ji}. If diversification is expected to occur in typical investors' portfolios so that the expected value (taken across investors) of the left-hand side of (3.2-28) is zero, then

$$\begin{aligned} p_j &= E(\mu_j + \epsilon_{ji})a_{0i} + \sum_{k=1}^{K} [E(\gamma_{jk} + \eta_{jki})a_{ki}] \\ &= \mu_j a_{0i} + \sum_{k=1}^{K} \gamma_{jk} a_{ki}, \end{aligned} \tag{3.2-29}$$

for all i. Equation (3.2-29) is the same as (3.2-8), which produced the valuation formula developed earlier.[6] Hence, even in the face of diverse expectations among investors, the model of price determination continues to be valid if diversification is a good approximation to observed behavior. The additional argument is that, because diversification implies that investors' expectations obey restriction (3.2-22), we may safely treat an investor's having extensively diversified holdings of securities as indicating that his expectations about them are typical.

The diversification model with heterogeneous expectations does predict that investors will have certain holdings that are not small because each investor's opinions may not be typical for some stocks. Investors will have relatively large holdings of the securities they deem especially attractive, involving them in a nontrivial exposure to the specific risks they perceive these securities to bear. One would also expect that some atypical holdings would be negative.[7]

This consideration suggests that the argument for the diversification model may not stand up well to a prohibition on negative holdings (or substantial costs of short-selling). The weakness is especially pressing if most portfolios, even extensively diversified ones, do not contain most securities. This pattern of investor behavior implies, given the other assumptions, that most investors perceive most securities to be over-

6. If ϵ_{ji} and η_{jki} are such that their medians and modes are the same as their means, then the assumption that "typical" behavior in these senses is diversification produces the same result.

7. Extensive short-selling does not occur, but we would attribute this phenomenon to the significant transactions costs involved rather than to investors' having homogeneous expectations. It is probably also the case that professional investment managers concentrate their efforts on finding securities that appear especially attractive. They do not seek out securities that may be overpriced.

valued or, at least, not more attractive than the securities in their portfolios. Restriction (3.2-27) still must hold for the expectations of any pair of investors with respect to securities in which each has extensively diversified holdings. It is now less reasonable, however, to suppose that this occurs because these expectations represent typical opinions and that one can infer from diversification that the coefficients a_{ki} need to be the same for all securities involved.

Suppose that, with negative holdings prohibited, investors hold only a scattering of securities. These holdings represent the few securities that they do not perceive to be overvalued. The Kuhn-Tucker conditions (3.2-12) can be written for securities of which the investor has a zero holding as

$$(\mu_{ji}a_{0i} + \sum_{k=1}^{K} \gamma_{jki}a_{ki})/p_j \leq 1. \tag{3.2-30}$$

The inequality is reversed for ones he does hold.

The values of $(\mu_{ji}a_{0i} + \Sigma_{k=1}^{K} \gamma_{jki}a_{ki})$, which represent the marginal values of securities to investors ignoring specific risk, are random variables. Since, by assumption, most investors do not hold any particular security, the value of this quantity for an investor who does hold the security must be unusually high. It must therefore be in the upper part of the distribution of this random variable. If we assume in the usual way that the density function is smooth and unimodal, the density decreases as this quantity increases.

For securities that an investor does hold,

$$-E(e_{ji}U_i')/\lambda_i p_j = (\mu_{ji}a_{0i} + \sum_{k=1}^{K} \gamma_{jki}a_{ki})/p_j - 1. \tag{3.2-31}$$

Diversification corresponds to small values of the right-hand side of (3.2-31). But these values, being in the upper tail of a unimodal distribution, are more likely to fall in an interval of given width with lower bound zero than to fall in any other positive interval of the same width. We may therefore conclude that an extensively diversified holding of any particular security is more probable than a larger one. The diversification argument used the observation of diversification to conclude that prices must exhibit certain patterns. Now, however, observing diversification only arises from the random nature of expectations and the observation that any positive holding is unusual. Hence the apparently systematic tendency to diversification may have no implication for security prices.

We have seen that equations (3.2-27) could happen to describe many of the securities that the opinions and utilities of each investor lead him to select without there being any systematic connection between the opinions of different investors or between prices and the μ_j and γ_{jk} parameters. Any nondiversification tendencies that would arise from prices not following the linear relationship specified earlier are hidden and frustrated by the non-negativity constraint which prevents the typical (and less optimistic) opinion from finding expression in the market. While we may properly conclude that all investors who do hold small quantities of security j have the same value of the sum in (3.2-27), this in turn does not directly tell us about valuation relationships expressed in terms of more objective—or agreed upon—parameters.

Nevertheless, this argument against the diversification theory may not be as devastating as it may first appear. To be sure, the costs of short-selling are substantial (if not strictly speaking prohibitive), and some major market participants are directly prohibited from selling short. However, failure to hold most securities may arise from other sources.

There are several good reasons for portfolio managers to limit the number of securities they hold. As is well known, adequate diversification may easily be achieved with holdings of as few as twenty or thirty securities and the addition of further securities to the portfolio will probably reduce specific risk to an insignificant degree. Thus, for all practical purposes, there is little additional benefit from full diversification. On the other hand, there may be significant costs to increased numbers of holdings. "Prudent man" rules are often interpreted as requiring that someone from the investment company staff periodically visit the managements and facilities of all companies held in the portfolio. At the very least, these rules would require that periodic reports be prepared for review by an investment committee. In addition, custodial fees and audit charges, as well as the fees of collecting dividends and other distributions, are partly a function of the number of types of securities held. Thus there may be small but significant costs to more extensive diversification, which after some point has only very small or indeed negligible benefits. These costs may provide the reason for any particular investor not holding most securities.

An investor contemplating adding another security to his portfolio has to balance the advantages of the added holding against the increased costs. The advantages come as increased expected returns or decreased risk exposure. With extensive diversification, further decreases in risk coming from still smaller holdings of those securities already included are negligible. Against this decreased risk must be set the increase in risk coming from the positive holding of the new security. The risk from the new security increases with the size of the holding. Even if an unusually high rate of return is expected, investment will occur only if the quantity held can be large enough for the increased return to overcome the transactions costs without so increasing specific risk that no advantage accrues to the investor. Although we have treated the investor as certain about his own appreciation of the parameters of the distributions of return, it would be more reasonable to presume that he also treats them as random and so as giving rise to risk. Without (costly) investigation by the investor, we might suppose that the dispersion of his subjective distribution is such that it rules out pursuing any small perceived above-normal return. At the same time, we may assume that the expected return from investigation to bring the posterior dispersion down is less than the cost. If all this is the case, failure by most investors to hold a particular security cannot be taken to indicate that only those with unusually favorable opinions of the security do hold it. Any investor's not holding a particular security suggests only that she does not perceive it to offer very unusually favorable opportunities, not necessarily that she considers it a poor investment.

These arguments make it reasonable to suppose that an investor's opinions about the securities he does hold in small quantities represent average or typical opinion rather than extreme ones. Given that an investor may hold a fairly large fraction of the securities about which he has well-informed opinions, it is then the case that many of these opinions will tend to be near the means of their distributions. For securities for which his opinions are unusually favorable, an investor will, of course, have a large holding while he will hold none of the securities he judges a relatively poor investment.

This behavior does not upset the basic pricing of the diversification model, although its derivation does rely on an investor's small holdings tending to represent an average opinion. That this condition may obtain is rendered more likely by the model's being potentially self-fulfilling. Suppose that the diversification model does apply to the expected values across investors of the parameters. Suppose also, as we have been arguing, that a large part of an investor's beliefs and portfolio holdings correspond reasonably closely to these values (so that his values of the parameters a_{ki} correspond to the market coefficients). Then this investor may reasonably assume that any security which he has not investigated in detail will be found on investigation to offer only an expected return and coefficients for the factors that would not make it a clearly compelling candidate for inclusion or exclusion in a portfolio. That is, he can base his expectations about securities about which he does not have reliable, independent information on the market's valuation. As a result, the market parameters do represent the average of informed opinion, which is then joined by the typical uninformed opinions of those who do not hold the security. Such expectations are exactly what we need to preserve the validity of the diversification model.

When we consider whether heterogeneous expectations and prohibition (or large costs) of short-selling destroy the validity of the diversification argument, the crucial question is why a typical investor does not hold a particular security. If the reason is that she does not really know much about the security, or that she thinks its value is about right but that it offers only opportunities roughly equivalent to ones available from securities she already holds, no problems arise. If the omission can be explained only by her perceiving that the security offers a poor return relative to the ones she holds, then diversity of expectations and restrictions on short-selling would destroy the basis of the argument.

Our discussion has concentrated on diversity of evaluations of the firm-specific parameters μ_{ji} and γ_{jki}. Another cause of diverse expectations raises far less difficulty, as Ross (1977) has noted. We have treated the common factors f_k as having mean zero—with μ_j absorbing any nonzero means that actual factors may have. Instead, we may regard the factors as quantities requiring forecast, such as economic activity, interest rates, inflation, or even "regulatory climate." Differences of opinion might result from different forecasts of these quantities. Specifically we might assume that

$$\mu_j = \gamma_{j0} + \sum_{k=1}^{K} \gamma_{jk}\mu_k. \tag{3.2-32}$$

Suppose that diversity of expectations arises from investors' having different values for μ_k, say, μ_{ki}, so that

$$\mu_{ji} = \gamma_{j0} + \sum_{k=1}^{K} \gamma_{jk}\mu_{ki}. \tag{3.2-33}$$

Equation (3.2-5) now becomes

$$\begin{aligned} p_j &= (\gamma_{j0} + \sum_{k=1}^{K} \gamma_{jk}\mu_{ki})a_{0i} + \sum_{k=1}^{K} \gamma_{jk}a_{ki} \\ &\quad + E(e_jU_i')/\lambda_i \\ &= \mu_j a_{0i} + \sum_{k=1}^{K} \gamma_{jk}[(\mu_{ki} - \mu_k)a_{0i} + a_{ki}] \\ &\quad + E(e_iU_i')/\lambda_i \\ &= \mu_j a_{0i} + \sum_{k=1}^{K} \gamma_{jk}a'_{ki} + E(e_iU_i')\lambda_i. \end{aligned} \tag{3.2-34}$$

This is of the same form as (3.2-5). Investors adjust their exposure to the f_k so as to alter their marginal utilities to offset forecast differences with the result that the linear objective structure of valuation is preserved.

3.2.4 Differential Tax Treatment of Investors

This discussion has centered around random differences in the perceived returns. There may also be systematic differences among individual investors arising from different tax treatments of various parts of return streams. The tax laws do make distinctions among different types of returns associated with a particular security. Specifically, the dividend and capital gains components of total returns may be taxed differently. In addition, different investors are taxed at different rates on the two components. Incorporating this feature of actual investment situations in the model introduces some serious empirical implications about the extent to which diversification does occur.

Suppose two types of returns are associated with a security, r_j^1 and r_j^2, taxed in the case of investor i at the rates t_i^1 and t_i^2, respectively. Let $s_i^1 = (1 - t_i^1)$ and $s_i^2 = (1 - t_i^2)$ be the after-tax proportions of returns that come to investor i from the two sources. Equation (3.2-1) now becomes

$$(3.2\text{-}35) \qquad \begin{aligned} r_j &= r_j^1 + r_j^2 \\ &= \mu_i^1 + \mu_j^2 + \sum_{k=1}^{K} (\gamma_{jk}^1 + \gamma_{jk}^2) f_k + e_j^1 + e_j^2 . \end{aligned}$$

The relevant return for investor i is

$$(3.2\text{-}36) \qquad \begin{aligned} r_{ji} &= s_i^1 r_j^1 + s_i^2 r_j^2 \\ &= s_i^1 \mu_j^1 + s_i^2 \mu_j^2 + \sum_{k=1}^{K} (\gamma_{jk}^1 s_i^1 + \gamma_{jk}^2 s_i^2) f_k \\ &\quad + e_j^1 s_i^1 + e_j^2 s_i^2 . \end{aligned}$$

The equation to be derived from the first-order conditions, corresponding to (3.2-5), is

$$(3.2\text{-}37) \qquad \begin{aligned} p_j = (s_i^1 \mu_j^1 + s_i^2 \mu_j^2) a_{0i} + \sum_{k=1}^{K} (s_i^1 \gamma_{jk}^1 + s_i^2 \gamma_{jk}^2) a_{ki} \\ + E(e_j^1 s_i^1 U_i' + e_j^2 s_i^2 U_i') / \lambda_i . \end{aligned}$$

Extensive diversification would imply that the p_j obey the linear equations obtained by dropping the last term in equation (3.2-37). In general, this will not be possible when there are many investors unless tax rates do not vary across them or the proportions of the two types of returns are constant across securities. Neither condition seems very plausible. For example, extensive diversification would require inter alia that $(\mu_j^1 s_i^1 + \mu_j^2 s_i^2) a_{0i}$ be the same for all investors. This will not be the case unless we can write the term $(\mu_j^1 s_i^1 + \mu_j^2 s_i^2)$ as the product of parameters ϕ_j and θ_i in the form

$$(3.2\text{-}38) \qquad \phi_j \theta_i = (\mu_j^1 s_i^1 + \mu_j^2 s_i^2)$$

for all j and i. The same type of requirement applies to other terms in (3.2-37).

Differential tax treatments raise a very serious problem for the diversification theory if it is presumed that the diversified portfolios contain securities with significantly different splits between the dividend and the capital-gain components of return and are held by individuals with different rates of taxation. Whether this presumption is justified is less clear. There are many major institutional investors, such as pension funds and endowment funds, for which the distinction between income and capital gains is immaterial because no differences arise in their tax treatment. Such investors dominate the market for many securities in the sense that they are between them the principal holders. If these investors have extensively diversified portfolios, then the model will be applicable for the securities that they do hold in small quantities. Insofar as other investors receive different returns due to taxes, one would expect them to diversify over other securities, especially if restrictions on negative holdings[8] apply or there are costs associated with the number of securities held. The diversification argument does not preclude different valuation equations for securities held in different portfolios provided that they are not connected through portfolios in which there are diversified holdings of the different types of securities.

The existence of differential tax treatments of dividends and capital gains raises problems for other valuation theories, including the CAPM. It also raises a fundamental question concerning the justification for a firm's paying dividends at all. Our argument may account for dividend-paying behavior. If common-stock valuation is dominated by institutional investors not affected by tax differences, then the valuation formula of the diversification model holds without distinguishing the forms that returns may take. The possibility of dividend policy significantly affecting valuation through tax effects is therefore removed and only other reasons for retaining earnings (including their effects on expected returns) remain relevant.

3.2.5 Comparison with Other Valuation Models

The arguments for the applicability of the diversification model are hardly conclusive when heterogeneity of expectations is recognized. The validity of the other main candidates is unfortunately no more clear. These are the CAPM and Ross's (1976, 1977) strikingly original APT.

Like Ross's argument, our reasoning relies on diversification to eliminate risks specific to the security and produces the same conclusions. Ross's original argument is easily summarized. Provided that negative holdings are permitted on the same terms as positive ones, an investor can hold a portfolio with zero net cost to himself. That is, he can form a portfolio with holdings v_{ja} such that

$$\sum_{j=1}^{J} v_{ja} p_j = 0. \tag{3.2-39}$$

8. Negative holdings—that is, short sales—in which the different types of return receive the same tax treatment as positive holdings certainly are not feasible. It may be worth noting that Blume and Friend (1975) found that the portfolios of individuals are not extensively diversified.

Second, he can choose this portfolio in such a way that the effects of each systematic factor cancel out and the portfolio return is not affected by the values taken on by these factors; that is, the portfolio is such that

$$\sum_{j=1}^{J} v_{ja}\gamma_{jk} = 0\,, \qquad k = 1, \ldots, K. \tag{3.2-40}$$

If there are many securities, then the portfolio can also be diversified so that the holding of any one security is small. The return to the portfolio is $(\sum_{j=1}^{J} \mu_j v_{ja} + \sum_{j=1}^{J} e_j v_{ja})$. With adequate diversification, the second term can be treated as zero. The first term then can be taken as yielding the certain return on a costless portfolio. It "should" be zero since otherwise profitable opportunities would exist to invest costlessly and receive a riskless, positive return. The condition that $\sum_{j=1}^{J} \mu_j v_{ja} = 0$ for all v_{ja} obeying both (3.2-39) and (3.2-40) implies that there are coefficients a_k, $k = 0, \ldots, K$, such that

$$p_j = \mu_j a_0 + \sum_{k=1}^{K} \gamma_{jk} a_k. \tag{3.2-41}$$

This is also the conclusion of the diversification argument.

Arbitrage in the sense used in this argument clearly requires that negative holdings be possible costlessly. The relevance of a pure arbitrage argument when this is not the case may seem slim.[9] The advantage of the diversification argument is that it is not vulnerable to a constraint against negative holdings. The nature of the argument is also different. The APT relies on a property of market equilibrium and the law of large numbers to infer the structure of equilibrium prices. The diversification argument uses a stylized "fact" about utility-maximizing investors to infer the structure of security prices they face.

Diversity of expectations about expected return or factor coefficients also makes a pure arbitrage argument implausible. Different investors presumably would perceive different arbitrage possibilities, and it would

9. Surprisingly, Roll and Ross (1980) try to argue that negative holdings are not needed for the model because it must also "work" in differential form around the investors' (positive) holdings. However, this requires diversification to be valid. Specifically, consider a differential portfolio change, dv_{ji}, $i = 1, \ldots, J$ such that $\sum_{j=1}^{J} p_j dv_{ji} = 0$ and $\sum_{j=1}^{J} \gamma_{jk} dv_{ji} = 0$, $k = 1, \ldots, K$ (so that the change has the properties of an arbitrage portfolio). The argument then appears to be that $\sum_{j=1}^{J} \mu_{ji} dv_{ji} = 0$ because no increase in risk will be introduced into the portfolio by the change. But this argument does require diversification. Let σ_j^2 be the variance of e_j. Then the differential change in portfolio variance will be $\sum_{j=1}^{J} v_{ji}\sigma_j^2 dv_{ji}$. This will be of the same order of magnitude as $\sum_{j=1}^{J} \mu_j dv_{ji}$ unless the $v_{ji}\sigma_j^2$ for which dv_{ji} are nonzero are themselves infinitesimal. Thus the portfolio's holdings of these securities must be extensively diversified. While this differential argument does not itself support the arbitrage argument, it does indicate that if diversification occurs in the portfolios of investors who would exploit potential arbitrage opportunities, then the APT pricing must exist even if these investors are not expected utility maximizers.

not be possible for the market to eliminate them for all investors. Hence, with diverse expectations within the context of the model (3.2-1), investors must be presumed to perceive arbitrage possibilities that they are not able to exploit extensively for some reason. As Ross (1977) noted, however, the arbitrage argument is robust to diversity of expectations over the expected value of the factors of the sort we considered in (3.2-32), just as is the diversification argument.

The other alternative to the diversification model is the CAPM. The CAPM is a general-equilibrium theory based on the hypothesis that all investors' utility functions can be defined over the means and variances of their portfolios. Letting ω_{jh} be the covariance of return j with return h, the first-order conditions yield

$$p_j = \mu_j \theta_i + \psi_i \sum_{h=1}^{J} \omega_{jh} v_{hi}. \tag{3.2-42}$$

Aggregation and the imposition of the equilibrium condition $\sum_{i=1}^{I} v_{ji} = n_j$, with n_j being the total number of shares of security j outstanding and I the number of investors, produce

$$p_j = \mu_j \theta + \psi \sum_{h=1}^{J} \omega_{jh} n_h. \tag{3.2-43}$$

Equation (3.2-43) indicates that the market adjusts prices so that the market portfolio—that is, a portfolio with holdings proportional to n_j—is mean-variance efficient. Equation (3.2-43) can be manipulated to give the standard expression for the CAPM:

$$\rho_j = \bar{\rho} + (\rho_m - \bar{\rho})\beta_j, \tag{3.2-44}$$

where β_j is the regression coefficient of the jth rate of return on the rate of return of the market portfolio, ρ_m is the expected rate of return on the market portfolio, and $\bar{\rho}$ is the expected rate of return on a portfolio uncorrelated with the market and is the rate of return on the risk-free security if it exists. Equation (3.2-44) corresponds to (3.2-15). However, as Roll (1977) emphasized, it is a direct implication of (3.2-43) and can only be tested using exactly the market portfolio.

The impracticality of the CAPM disappears when the return assumption (3.2-1) is also made. Letting ξ_{km} be the covariance of the factors f_k and f_m, the variances and covariances can be expressed as

$$\begin{aligned} \omega_{jj} &= \sum_{k=1}^{K} \sum_{m=1}^{K} \gamma_{jk}\gamma_{jm}\xi_{km} + \sigma_j^2, \\ \omega_{jh} &= \sum_{k=1}^{K} \sum_{m=1}^{K} \gamma_{jk}\gamma_{hm}\xi_{km}, \qquad j \neq h. \end{aligned} \tag{3.2-45}$$

Here σ_j^2 is the variance of the specific factor e_j. The price equation (3.2-42) can now be written

$$(3.2\text{-}46) \qquad p_j = \mu_j/(1+\bar{\rho}) + \psi \sum_{k=1}^{K} \gamma_{jk} \sum_{m=1}^{K} \sum_{h=1}^{I} \xi_{km}\gamma_{hm}n_h + \psi\sigma_j^2 n_j$$

$$= \mu_j/(1+\bar{\rho}) + \sum_{k=1} \gamma_{jk}\phi_k + \psi\sigma_j^2 n_j.$$

If $\sigma_j^2 n_j$ is small relative to the middle terms (which arise from the covariances of return j with the common factors), then we can safely approximate (3.2-46) by (3.2-9), the valuation equation that is implied by diversification. If this is the case, it will also be the case that the portfolios chosen by investors can be considered extensively diversified in the sense used in section 3.2.1. Similarly, (3.2-15) approximates (3.2-44) when the specific risk term in (3.2-46) is small. In other words, when the return assumption (3.2-1) is made and when specific risks are small in the economy relative to systematic risks, the prices arising from the CAPM are well approximated by the APT formula.

When (3.2-1) does not hold, the CAPM encounters what is at present the insurmountable problem of measuring β_j in a way which would make the approach valid, since the market portfolio cannot be measured in the appropriate way. This removes a large part of the attraction of the theory since the mean-variance utility-maximizing hypothesis and the uniform nature of all investors' behavior are not very appealing. While the CAPM can be widened to allow nonmarketable returns as Mayers (1972) showed, the sum of these nonmarketable incomes would also need to be measured accurately. The conditions needed to obtain the CAPM in a multiperiod setting are very strong as Fama (1970) and Merton (1971) showed, and it is questionable if a wider model can be given empirical content without a factor type specification. As we shall see in section 3.4, the assumption (3.2-1) is very convenient for giving empirical content to the valuation theory, quite apart from its allowing conclusions to be drawn from the weaker assumptions about investor behavior of the diversification model.

The CAPM is quite robust to some forms of diversity of expectations if short-selling is costless. However, it is vulnerable to restrictions on short-selling and to other restrictions on investors' holdings when there are heterogeneous expectations or when other conditions prevent investors' optimal risky portfolios from being proportional to the market portfolio. Analysis then becomes difficult as Levy (1978) and Jarrow (1980) demonstrated, and useful valuation formulae cannot be derived. The measurement difficulties and the implications of the CAPM about portfolio holdings that are contrary to observation make the diversification model appear a more promising valuation approach on which to base empirical work.

The tests we shall conduct later will use the sort of regression coefficients that are normally considered to test the CAPM. However, our justification for their use does not require the validity of the CAPM and the corresponding appropriateness of market indexes. As we explain in section 3.4, these coefficients can be treated as reliable proxies for the factor coefficients.

3.3 Expected Returns and Earnings Growth

A number of tests of the diversification model are possible. The main prediction of the theory is that the prices of securities will be linearly related to their (*ex ante*) expected returns and their coefficients, γ_{jk}, of the factors in equation (3.2-1). Equivalently, expected rates of return will be linear functions of these coefficients. When *ex ante* expectations are equated to the mathematical expectations of the returns that can be observed, the most direct approach would be to estimate the model (3.2-1), expressed in rate of return form, using factor analysis, and test the hypothesis of the diversification model as a restriction on the parameters. Similarly, using data on r_j, μ_j, and γ_{jk} might be estimated and related to price data.

Such investigations are largely beyond the capabilities of the data set being used in this study. Furthermore, by concentrating on *ex post* returns, this approach presumes that expectations held by the investors correspond to the actual parameters of the distribution of returns. In consequence, it would throw no light on the role and nature of expectations of the sort we have been studying. Our objective in the present section is to express the expectations of returns in terms of the growth rates we have collected and to consider to what extent the return structure of the diversification model remains reasonable within this framework. The hypothesis about investors' expectations is very simple because of the fact that the useful expectations we were able to collect concerned only growth rates of earnings per share.

The quantity about which expectations are formed in the valuation model is the total return to a security, denoted by r_j. Presumably, the gross return to the holder of a share is the dividend paid in the period plus the price of the share at the end of the period. Let t represent the time subscript, which will always be the last subscript. The current period is denoted by zero. Let d_{jt} be the dividend paid in the period immediately preceding t so that the prices p_{jt} are ex dividend. For simplicity, we assume that dividends are paid at the ends of periods. Then the presumption is that

$$r_{jt} = d_{jt+1} + p_{jt+1}, \tag{3.3-1}$$

so that

$$\mu_{jt} = E(r_{jt}) = E(d_{jt+1}) + E(p_{jt+1}). \tag{3.3-2}$$

Expression (3.3-2) depends on the price of the security in the next period. It does not take a very perceptive investor to realize that next period's price is relevant to his returns or that other investors' attitudes (even irrational ones) toward the securities in the next period will affect the prices in that period. But it is not clear what the investor should assume about these matters. Indeed, the problems involved in Keynes's (1936) celebrated beauty contest may easily manifest themselves in the market. Self-fulfilling forecasts, based on no sensible valuation of the assets of a company, could in some periods even dominate the determination of the price of its shares. The history of speculative bubbles or the problems involved with the valuation of assets such as Picasso paintings must cause some doubts about formulations which tie the expectations of investors to features of securities in which they invest. But some such formulation must be used if the theory is not to be an empty bit of logic.

Suppose that utility-maximizing investors with extensively diversified portfolios believe that prices in the future will be formed in the manner suggested by the diversification theory in equation (3.2-9). That is, assume investors believe that

$$p_{jt} = \mu_{jt} a_{0t} + \sum_{k=1}^{K} \gamma_{jkt} \alpha_{kt}, \tag{3.3-3}$$

where the time subscripts recognize that, in principle, all quantities in (3.3-3) may change over time. We may then presume that investors form their expectations of p_{jt} by taking expectations of the right-hand side of (3.3-3) so that (3.3-2) becomes

$$\begin{aligned} \mu_{jt} = {} & E(d_{j,t+1}) + E(\mu_{j,t+1} a_{0,t+1} \\ & + \sum_{k=1}^{K} \gamma_{jk,t+1} a_{k,t+1}). \end{aligned} \tag{3.3-4}$$

We can see from this expression that risk arises from two sources. First, the valuation coefficients a_{kt} may change; second, the evaluation of the firm-specific parameters μ_{jt} and γ_{jkt} may change. In the latter case, probably the most important source of change is a potential reevaluation of the earnings (and hence dividend) prospects of the company.

Repeated substitution of (3.2-4) for μ_{jt} in itself yields

$$\begin{aligned} \mu_{j0} = {} & E(d_{j1}) + E\left[\sum_{t=2}^{\infty} \left(\prod_{s=1}^{t-1} a_{0s}\right) d_{jt}\right] \\ & + E\left(\sum_{k=1}^{K} a_{k1} \gamma_{jk1}\right) \\ & + E\left[\sum_{t=2}^{\infty} \left(\prod_{s=1}^{t=1} a_{0s}\right) \sum_{k=1}^{K} a_{kt} \gamma_{jkt}\right]. \end{aligned} \tag{3.3-5}$$

This involves a formidable number of terms and expectations, including the products of the a_{0s} coefficients. Considerable simplification is needed if anything useful is to be said about valuation. This can be achieved by assuming that the expected values of the valuation coefficients formed at time zero (conditional on any intervening terms) are all equal in the sense that

(3.3-6) $$E(a_{kt}) = a_{k0}, \qquad k = 1, \ldots, K,$$

(3.3-7) $$\gamma_{jkt} = \gamma_{jk},$$

and

(3.3-8) $$E\left(\prod_{s=1}^{t} a_{0s}\right) = a_{00}^t.$$

Assumption (3.3-6) is needed mainly for notational convenience and can be easily relaxed. The same is less true for (3.3-7) only in the sense that we shall need constant values to give empirical content to the model. Assumption (3.3-8) can be relaxed less readily and only partially without the danger of introducing possibly serious nonlinearities into the diversification model. Assumptions (3.3-6) to (3.3-8) and the additional assumption that the a_{0s} are independent of d_{jt}, of γ_{jkt}, and of a_{kt} allow us to write (3.3-5) as

(3.3-9) $$\mu_{j0} = \sum_{t=0}^{\infty} a_0^t E(d_{jt+1}) = \sum_{t=1}^{\infty} a_0^t \sum_{k=1}^{K} a_k\gamma_{jk}.$$

Determination of μ_{j0} now depends on the future path of dividends. The easiest assumption to make about dividends is that they are expected to grow (indefinitely) at a constant rate g_j so that $E(d_{jt}) = d_{j0}(1 + g_j)^t$. When it is recalled that we can express a_0 as $1/(1+\bar{\rho})$, where $\bar{\rho}$ is the risk-free interest rate, (3.2-9) becomes

(3.3-10) $$\begin{aligned} p_{j0} &= d_{j0}\sum_{t=1}^{\infty}(1+g_j)^t/(1+\bar{\rho})^t \\ &\quad + \sum_{k=1}^{K}\sum_{t=0}^{\infty} a_k\gamma_{jk}/(1+\bar{\rho})^t \\ &= d_{j0}(1+g_j)/(\bar{\rho}-g_j) + \sum_{k=1}^{K} a_k\gamma_{jk}(1+\bar{\rho})/\bar{\rho}. \end{aligned}$$

This equation can be written as

(3.3-11) $$\begin{aligned} g_j + (1+g_j)d_{j0}/p_{j0} &= \bar{\rho} - (1+\bar{\rho})\sum_{k=1}^{K} a_k\gamma_{jk} \\ &\qquad \times (g_j - \bar{\rho})/(p_{j0}\bar{\rho}) \\ &= \bar{\rho} - \sum_{k=1}^{K} b_k\beta_{jk}. \end{aligned}$$

The expression on the left-hand side of (3.3-11) can be interpreted by supposing for the moment that the price of a security can be expressed as the present value of the expected stream of dividends discounted at the (constant) expected rate of return to the security, ρ_j. That is, if we assume that

$$(3.3\text{-}12) \qquad p_{j0} = \sum_{t=1}^{\infty} E(d_{jt})/(1+\rho_j)^t$$

$$= d_{j0} \sum_{t=1}^{\infty} (1+g_j)^t/(1+\rho_j)^t$$

$$= d_{j0}(1+g_j)/(\rho_j - g_j),$$

then we can solve for ρ_j as

$$(3.3\text{-}13) \qquad \rho_j = g_j + (1+g_j)\, d_{j0}/p_{j0}.$$

We can therefore interpret (3.3-11) as being an equation for the expected rate of return to security j, and so it corresponds to the expected rate of return equation (3.2-15). If dividends in the next period, or their expected values, are known, then equation (3.3-13) becomes

$$(3.3\text{-}14) \qquad \rho_j = g_j + d_{j1}/p_{j0},$$

which also becomes the expression in (3.3-11). With d_{j0} in place of d_{j1}, (3.3-14) gives the formula when a continuous time model is used.

Instead of expressing the model in terms of expected rates of return in (3.3-11), we could divide (3.3-10) by earnings to obtain

$$(3.3\text{-}15) \qquad p_{j0}/e_{j0} = [(1+g_j)/(\bar{\rho} - g_j)]\, d_{j0}/e_{j0}$$

$$+ \sum_{k=1}^{K} (a_k \gamma_{jk}/e_{j0})(1+\bar{\rho})/\bar{\rho}.$$

This formulation brings out the relationship between our valuation model and more traditional approaches based on earnings multiples such as that by Williams (1938). There the risk term is ignored so that the valuation expression is given by only the first term of the right-hand side of (3.3-15).

These traditional formulations have a number of formidable drawbacks which apply here too. Their derivations are inapplicable in cases where no dividends are paid. They lead to an infinite value for the security when $g_j \geq \bar{\rho}$. They require projecting differential growth rates from here to kingdom-come at a constant rate. This last drawback would seem to render the model particularly inappropriate for use with growth-rate projections such as ours that were specifically made for limited periods of time.

Such difficulties have led several writers to formulate finite-horizon models for share prices. The basic idea is that dividends and earnings are assumed to grow at rate g_j for T periods. They then either stop growing or else proceed to grow only at some normal rate, such as the general rate of growth of the economy. In some models, the growth rate is assumed to revert to the final growth rate in stages, or according to a smooth decay function. Correspondingly, dividends may return in any one of several ways to some standard payout ratio consistent with a continuation of normal growth corresponding to that of the economy as a whole.

There is a large family of explicit models that can be developed from these approaches. They tend to share two features. First, they are apt to be nonlinear in the important quantities. Second, they all depend on the same rather limited list of variables. Thus the various formulations based on the discounted-dividends approach all suggest that there is an implicit function of the form

$$(3.3\text{-}16) \qquad f(\rho_j, g_j, d_{j0}, e_{j0}) = 0.$$

In the context of the diversification model, ρ_j includes μ_{j0}, p_{j0}, and also the terms $\Sigma_{k=1}^{K} \gamma_{jk} a_k$ which enter μ_j by being logically part of future prices. In principle we can solve (3.3-16) for μ_{j0}. Suppose that we can write this solution as

$$(3.3\text{-}17) \qquad \mu_{j0} = e_{j0}\,\mu(g_j, d_{j0}/e_{j0}, \sum_{k=1}^{K} \gamma_{jk} a_k / e_{j0}).$$

Then, dividing (3.2-9) by e_{j0} and assuming a linear approximation for μ in (3.3-17), we obtain

$$(3.3\text{-}18) \qquad p_{j0}/e_{j0} = c_0 + c_1 g_j + c_2 d_{j0}/e_{j0} + \sum_{k=1}^{K} c_{k+2} \gamma_{jk}/e_{j0}.$$

A still better approximation might arise from using quadratic terms from (3.3-17) as well. However, some experiments using the explicit limited-horizon model of Malkiel (1963) indicated that a linear equation such as (3.3-18) fitted very well over a wide range of parameter values and growth rates and was a good approximation for the true nonlinear model. These same formulations also arise as approximations when the price-formation equations are taken as the starting point for the derivation of expected return ρ_j.

The conclusion of these arguments is that, although the constant growth-rate and constant expected-return models may seem implausible, they may serve as an adequate linear approximation. The specific formulation we use is geared to the data available, concentrating on the growth rates of earnings. However, we do have both short- and long-term growth rates available, and we found that they were not closely related. There is a large variety of two-parameter models for earnings that could be expressed in terms of long- and short-term growth rates. Rather than adopt any specific such model, we shall assume that the short-term growth rate may appear as another argument in (3.3-17) and so as an additional variable in (3.3-18).

Our subsequent empirical work is based on relying on the adequacy of this linear approximation without explicitly assuming a particular process for earnings growth and retention. In addition, however, we shall investigate one explicit formulation for ρ_j based on the simple model (3.3-14) to which the valuation model (3.3-11) is supposed to apply.

Having adopted a more specific (though approximate) hypothesis about expected returns, we now need to examine to what extent the critical aspects of (3.2-1) needed for the diversification model seem reasonable. The question is whether the assumptions that prices follow (3.2-9), that expectations are formed in the way discussed, and that diversification is observed are all internally consistent.

Equations (3.2-9) and (3.3-1) allow us to write the actual return as

$$(3.3\text{-}19)\qquad \begin{aligned} r_{j0} &= d_{j1} + p_{j1} \\ &= d_{j1} + \mu_{j1}a_{01} + \sum_{k=1}^{K} \gamma_{jk1}a_{k1}\,. \end{aligned}$$

Assume that the γ_{jk} are constant. Then using (3.3-4) we obtain

$$(3.3\text{-}20)\qquad \begin{aligned} r_{j0} - \mu_{j0} = d_{j1} - E_0(d_{j1}) + \mu_{j1}a_{01} - E_0(\mu_{j1})a_{00} \\ + \sum_{k=1}^{K} \gamma_{jk}(a_{k1} - a_{k0})\,, \end{aligned}$$

where E_0 indicates the expectation as of time zero. Now assume that the factor model applies to unexpected changes in dividends and expected returns (or capital gains if a_0 is constant) in the sense that

$$(3.3\text{-}21)\qquad d_{j1} - E_0(d_{j1}) = \sum_{k=1}^{K} \delta_{jk}f_{k1} + \epsilon_{j1}$$

and

$$(3.3\text{-}22)\qquad \mu_{j1} - E_0(\mu_{j1}) = \sum_{k=1}^{K} \phi_{jk}f_{k1} + \eta_{j1}\,.$$

Using specification (3.3-17) for μ_j, equation (3.3-22) suggests that the factors reflect any systematic influences that affect dividends or earnings of firms and their growth prospects as well as future discount rates. If we assume temporarily that the a_k coefficients are all constant, we obtain

$$(3.3\text{-}23)\qquad r_{j0} - \mu_{j0} = \sum_{k=1}^{K} (\delta_{jk} + a_0\phi_{jk})f_k + \epsilon_{j1} + a_0\eta_{j1}\,.$$

It is natural then to equate the coefficients $(\delta_{jk} + a_0\phi_{jk})$ with γ_{jk} and $(\epsilon_{j1} + a_0\eta_{j1})$ with e_1. The main difficulty here is that a_0 may well change so that the γ_{jk} would also change and do so by more than a factor of proportion constant across companies. Although the model could be developed without further difficulty using the coefficients δ_{jk} and ϕ_{jk} as (separate) γ_{jk} coefficients, we would then need separate ways of obtaining estimates of these coefficients. This would require development of models for dividends and earnings growth going far beyond the scope of the present study. Furthermore, we have already suggested that simple formulations provide adequate approximations to the more complicated (and unknown) specific models. In order to achieve empirical simplicity therefore we shall assume that $(\delta_{jk} + a_0\phi_{jk}) = \gamma_{jk}$. We can then express (3.3-20) as

$$
\begin{aligned}
(3.3\text{-}24) \qquad r_{j0} - \mu_{j0} = {} & a_{00} \sum_{k=1}^{K} \gamma_{jk} f_k + (a_{01} - a_{00}) E_0(\mu_{j1}) \\
& + (a_{01} - a_{00}) \sum_{k=1}^{K} \gamma_{jk} f_k \\
& + \sum_{k=1}^{K} \gamma_{jk}(a_{k1} - a_{k0}) + e_{j0}.
\end{aligned}
$$

This specification is close to the bilinear form hypothesized in section 3.2.2 which, though nonlinear in factors, does give rise to the valuation model in equation (3.2-9). This valuation equation itself was used in obtaining equation (3.3-23) so that the formulation we have adopted is self-consistent, given the various approximations assumed. The key one among them is that γ_{jk} can be regarded as constant.[10] This might be relaxed while maintaining the spirit of the diversification model, but only by producing major additional programs to obtain useful empirical measures of risk.

3.4 Specification of Risk Measures

Our problem in the previous section concerned relating the data on expected earnings growth to the valuation model. Relevant data are not available on the perceptions of market participants about risk, and we shall have to presume that parameters estimated from *ex post* data correspond to investors' perceptions. The resulting variables of necessity are mainly *ex post* measures derived from realized data rather than true *ex ante* data representing the views actually expressed by investors. We shall find, however, that our data do permit the development of one *ex ante* risk measure that proves quite serviceable.

The approach we take to obtain risk measures is more easily explained by considering the coefficients γ_{jk} in (3.2-1) than those in the more complicated equation (3.3-23). We shall, however, assume that over time the coefficients are more stable if we estimate in rate of return form. That is, we shall assume that (3.2-16) holds over time and consider estimating the parameters α_{jk}. This choice was made partly to increase comparability with other valuation studies.

The data set available to us makes estimation of the α_{jk} coefficients by factor analysis impractical. We therefore adopt a different procedure,

10. As equation (3.3-23) makes clear, one of the γ_{jk} coefficients should be $E_0(\mu_{j1})$. Like the problem encountered with combining (3.3-19) and (3.3-20), this raises the problem that the γ_{jk} are not likely to be constant over time rather than directly attacking the valuation formulation. The assumption might be made more palatable by presuming something like (3.2-32).

which may not be inferior and which bears an immediate connection to most recent investigations of valuation models. This involves regressing experienced rates of return of each company on various market or economy-wide indexes. The standard empirical approach to the CAPM uses such regression coefficients; but if (3.2-1) is not correct, this procedure is subject to the criticism of Roll (1977) that the market portfolio must be correctly specified before useful measures can be derived. Fortunately, the assumptions of the diversification model permit use of such regression coefficients, though not without some econometric difficulties.

To bring out the essence of the argument, assume for a moment that there is only one factor so that we can write (3.2-16) as

$$\pi_{jt} = \rho_j + \alpha_j f_t + e_{jt}. \tag{3.4-1}$$

Since f is unobservable, we can normalize it by assuming that $E(f_t^2) = 1$. Let $\sigma_j^2 = E(e_{jt}^2)$. We can try to represent f_t by an "index" g_t, a weighted sum of individual security returns with weights consisting of v_{jg} ($j = 1, \ldots, J$) units of each security. (With actual indexes, many v_{jg} are zero.) The rate of return to this index is

$$\begin{aligned} \pi_{gt} &= \sum_{j=1}^{J} \rho_j v_{jg} = \sum_{j=1}^{J} f_t \alpha_j v_{jg} = \sum_{j=1}^{J} e_{jt} v_{jg} \\ &= \rho_g + f_t \alpha_g + e_{gt}. \end{aligned} \tag{3.4-2}$$

Thus the index also is a linear function of the common factor f_t and of a random term. The covariance of this index with the return to security j is given by

$$\begin{aligned} E(\pi_{gt} - \rho_g)(\pi_{jt} - \rho_j) &= E(f_t \alpha_g + e_{gt})(f_t \alpha_j + e_{jt}) \\ &= \alpha_g \alpha_j + \sigma_j^2 v_{jg}, \end{aligned} \tag{3.4-3}$$

since, by assumption, f_t not only has unit variance but also is independent of all e_{jt}, which are assumed to be independent of each other. If the returns are normally distributed,[11] the regression of the rate of return of each security, π_{jt}, on the rate of return of the index, π_{gt}, has population regression coefficient

$$\begin{aligned} \beta_{gj} &= E[(\pi_{gt} - \rho_g)(\pi_{jt} - \rho_j)]/E[(\pi_{gt} - \rho_g)^2] \\ &= (\alpha_g \alpha_j + \sigma_j^2 v_{jg})/(\alpha_g^2 + \sigma_g^2) \\ &= \theta_1 \alpha_j + \theta_2 \sigma_j^2 v_{jg}. \end{aligned} \tag{3.4-4}$$

11. The normality assumption is required to ease expression of the expected value of the regression coefficients. Without it, analogous expressions hold for the expectations though with different explicit interpretations of some coefficients and all formulae hold asymptotically when we take the limits in probability.

Equation (3.4-4) indicates that the regression coefficient is a linear combination of the systematic risk and of the specific risk associated with security j. The second term depends on the size of v_{jg} relative to the total holdings of securities incorporated in the index and on the variance of σ_j^2 of the specific risk of the security. If v_{jg} can be considered to be small, β_{gj} may be taken to be proportional to the coefficient α_j needed for specifying the valuation equation. The factor of proportion θ_1 is the same for all securities. As a result, this measure would be quite adequate for investigations of the valuation model since that model does not provide hypotheses about the numerical values of the coefficients a_k in (3.2-9) or b_k in (3.2-15).

Whether one can safely ignore the specific-risk term in equation (3.4-4) depends partly on the index being used. Major indexes are heavily weighted with precisely the leading securities that we have been studying. Indeed, one might feel that an index which included none of these securities was a rather odd construction. However, if the index can be taken to represent a fairly extensively diversified portfolio, then the reasoning that suggests that the security's own variation can be considered trivial for purposes of valuation also suggests that the term $\sigma_j^2 v_{jg}$ in (3.4-4) may be ignored. Thus, if the diversification argument is appropriate, the use of the regression coefficient β_{gj} calculated from a broadly based index is appropriate. It is not required, as it would be if only the general CAPM were valid, that the index represent the true "market."

The essence of the argument is that the index may itself be regarded as a measure of f_t which is subject to an error of observation. However, with an adequately diversified index, the measurement error does not produce a serious bias since, with the second term in (3.4-4) negligible, $\beta_{gj} = \theta_1\alpha_j$ with θ_1 having the same value for each security.

The variables on which the regressions are run need not be limited to one index, or indeed even to market indexes. Instead, we can also use other quantities such as national income or inflation with which we expect earnings or dividends to be correlated or with which the market-valuation parameters might vary. We can also use rates of return of other securities or other common factors. Extension to more variables may cause some problems of interpretation, however.

Suppose that an H-element vector of variables x_t is available on which we can regress the rate of return of each security. Let these variables be such that we can describe them as

$$x_t' = \phi_0' + f_t'\Phi + \eta_t', \tag{3.4-5}$$

where f_t is now the K-element vector of true factor values, Φ is the $K \times H$ matrix of (population) coefficients for the (hypothetical) regression of x_t on f_t, and η_t is the vector of residuals from this regression. We shall treat these residuals as random variables with the usual regression assumptions, namely, that they have mean zero and variance/covariance matrix S. The assumption that (3.2-1) fully describes all relevant correlations of returns with other quantities, which is the key to the diversification model, leads to the assumption that all residuals of the equations (3.4-1) for the rates of return of the individual securities, e_{jt}, are independent of the vector η_t.

Our suggested procedure is to regress the separate rates of return π_{jt} on the variables x_t (augmented by a constant). The regression is thus of the form

$$\pi_{jt} = \hat{\delta}_{j0} + x_t'\hat{\delta}_j + \hat{w}_{jt}, \tag{3.4-6}$$

where $\hat{\delta}_j$ is the vector of estimated regression coefficients and $\hat{w}_{jt}$ are the residuals. Since we will want to use $\hat{\delta}_j$ in place of α_j, the K-element vector of factor loadings α_{jk} from (3.2-16), we are concerned with the relationship of $\hat{\delta}_j$ to α_j.

Suppose that X is the $T \times H$ matrix whose rows are the observations on $(x_t - \bar{x})$ to be used in calculating the regression, $\bar{x}$ being the vector of average observations of x_t. Let F be the $T \times K$ matrix whose rows are the corresponding f_t' vectors and π_j be the corresponding $T \times 1$ vector with typical element $(\pi_{jt} - \Sigma_{t=1}^{T}\pi_{jt}/T)$. For the moment let e_j designate the $T \times 1$ vector with typical element e_{jt}. Then the estimates of $\hat{\delta}_j$ of (3.4-6) can be taken as estimating δ_j, given by

$$\begin{aligned} \delta_j &= E(\hat{\delta}_j) \\ &= E(X'X)^{-1}X'\pi_j \\ &= E[(X'X)^{-1}X'F]\alpha_j + E[(X'X)^{-1}X'e_j] \\ &= \Xi\alpha_j . \end{aligned} \tag{3.4-7}$$

The last equality in (3.4-7) arises from assuming that e_j is independent of X. Note that Ξ is the expected value of the regression coefficients in the (multivariate) reverse regression of F on X. The supplementary specification (3.4-5) was not used in obtaining (3.4-7), and it serves only to make it reasonable to assume that the estimates $\hat{\delta}_j$ have the usual properties. With (3.4-5), the "true" residual of (3.4-6) becomes

$$\begin{aligned} w_{jt} &= \pi_{jt} - x_t'\delta \\ &= e_{jt} + (f_t' - x_t'\Xi)\alpha_j \\ &= e_{jt} + f_t'[I - \Phi\Xi]\alpha_j + \eta_t'\Xi\alpha_j . \end{aligned} \tag{3.4-8}$$

While this is not necessarily independent of x_t, it can be considered orthogonal to it since e_{jt} is and the remaining terms are the (population) residuals of the reverse regression.[12]

12. The difficulties here are overcome if we assume that all variables are normally distributed. Then the reverse regression is as appropriate as (3.4-5). If we let $E(f_t f_t') = M$, a

Equation (3.4-7) expresses δ_j as a linear transformation of the coefficients α_j, with the same transformation being used for each company. Several aspects of this transformation are worth noting. First, even if there are as many variables as factors, i.e., if $H = K$, and there is a one-to-one correspondence of factors to variables in the sense that $\Phi = I$, δ_j is a biased estimate of α_j. This bias results in each element of δ_j being a linear combination of all the elements of α_j. To avoid this, we would need $E(F'F)$ and S as well as Φ to be diagonal, so that $X'X$ could also be expected to be diagonal. This does not happen to be a condition that appears to be met in our data. Without these conditions, Ξ will not be diagonal so that any one δ_j coefficient will combine the effects of several factors. One must therefore be cautious in interpreting the elements of δ_j as giving the effects of the factors to which the corresponding elements of x_t correspond most directly.

More thorny, but still tractable, problems arise when the number of x-variables is not the same as the number of factors. In the event that there are more x-variables than factors, i.e., if $H > K$, it will not be the case that only K of the δ_j coefficients will be nonzero; instead, normally all these population coefficients will differ from zero. One can therefore expect their estimates also to be nonzero. What will be the case is that there will be a linear relationship among the population coefficients. When a set of J companies is considered, with the δ_j vector for each being estimated, then the cross-product matrix of the true regression coefficients would be singular in the sense that

$$(3.4\text{-}9) \qquad \sum_{j=1}^{J} (\delta_j \delta_j') = \Xi \sum_{j=1}^{J} (\alpha_j \alpha_j') \Xi',$$

and so the matrix in (3.4-9) is of (at most) rank K. Thus, if we were able to use the true coefficients δ_j in place of α_j (or γ_j) to estimate the valuation model, we would be confronted with perfect multicollinearity if we had more δ_j coefficients than there were α_j coefficients. This problem would also arise if there were fewer α_{jk} coefficients than factors, as (3.2-18) suggests might be the case, and we had as many δ_j as there were factors.

The true values of δ_j are no more available than are the true α_j coefficients. All we can use are estimated coefficients $\hat{\delta}_j$, and these coefficients do not share the property of the population parameters revealed in (3.4-9). These estimated coefficients, of course, differ from the true population parameters by the usual estimation error, say, u_j. That is, what we have available are

---

little manipulation gives us that $\Xi = [\Phi' M \Phi + S]^{-1} \Phi' M$ and w_{jt} is both independent of x_t and normally distributed. Similarly, if we have a large sample and if $M = p \lim F'F/T$, then again we can express $p \lim \Xi = [\Phi M' \phi + S]^{-1} \Phi' M$, and the orthogonality of w_j to X raises no difficulties for large samples since $p \lim w_j' X/T = 0$.

$$\hat{\delta}_j = \delta_j + u_j = \Xi\alpha_j + u_j. \tag{3.4-10}$$

Let $h_j = E(w_{jt}^2)$, with w_{jt} being the residual in (3.4-8). Presuming that it does have the usual regression properties yields the usual formula

$$E(u_j u_j') = E(\hat{\delta}_j - \delta_j)(\hat{\delta}_j - \delta_j)' = h_j(X'X)^{-1}. \tag{3.4-11}$$

As a result, corresponding to (3.4-9), we obtain

$$E\left[\sum_{j=1}^{J} (\hat{\delta}_j\hat{\delta}_j')/J\right] = \Xi \sum_{j=1}^{J} \alpha_j\alpha_j'/J\Xi' + \left(\sum_{j=1}^{J} h_j/J\right)(X'X)^{-1}, \tag{3.4-12}$$

which is of full rank. We would then apparently be able to use all $\hat{\delta}_j$ coefficients in estimating a valuation equation even though there are more of them than factors.

The estimation errors present in the $\hat{\delta}_j$ coefficients act as measurement errors when we turn to estimation of the valuation equation. The major consequence of this is that all $\hat{\delta}_j$ coefficients can be expected to have nonzero estimated effects in the valuation equation even though there are more of them than of the α_j coefficients. More disconcerting, other risk variables, provided they do have some correlation with the α_j, will appear to enter the valuation equations even though (3.4-5) is correct and $H \geq K$. The reason is the same as the reason why the δ_j coefficients can all be nonzero even when $H > K$. Indeed, expression (3.4-10) relates δ_j to α_j in the same sort of way that (3.4-5) related to x_t to f_t. This would also be the case for having still other risk measures. It is therefore fortunate that solutions are available to the problems caused by $\hat{\delta}_j$ being subject to estimation error.

The structure of the cross-product matrix of $\hat{\delta}_j$ given in (3.4-12) corresponds closely to the one assumed by the model used in statistical factor analysis. However, $\Sigma_{j=1}^{J}\ (h_j/J)(X'X)^{-1}$ is not a diagonal matrix, as is usually required for factor analysis. Balanced against this difficulty is the fact that $(X'X)$ is known. As a result, the applicability of the factor-analysis model can be investigated (through using a different metric for δ_j) and we are able to investigate from the regression coefficients, all of which may differ significantly from zero, how many underlying common-factor coefficients are actually present in security returns.

Determining the number of linearly independent δ_j coefficients indicates the number of γ_j coefficients that the diversification model suggests should be present in the valuation equation, but it does not solve the errors-in-variables problem also presented by the δ_j. The structure of the cross-product matrix shown in (3.4-12) and the fact that h_j can be estimated along with δ_j mean that we can correct for the estimation errors in $\hat{\delta}_j$. The procedure, developed in Cragg (1982), yields consistent estimates not of the a_k coefficients of (3.2-9) since α_j or γ_j are not estimated directly, but it will consistently estimate the vector $a'\Phi^{-1}$, where a is the vector of valuation coefficients. More important, other coefficients, including those of other possible explanatory variables, will now be estimated consistently. Thus, testing whether these coefficients are zero does test whether additional variables affect valuation and not simply whether they may help to proxy the inaccurately measured α_{jk} coefficients. This is fortunate since two further measures of risk suggest themselves from the data we shall be using.

The first additional risk variable is the variance of the individual predictions of long-term growth for each company, referred to as s_g^2. This variance may be taken to indicate the extent of uncertainty about the future rate of growth of the company, which may well be related to the uncertainty about the future returns of the company's securities. The variable s_g^2 may capture some combination of both specific and systematic risk, since disagreement about the future growth of earnings may come from different perceptions about the individual companies, or it may arise from different opinions about the values to be taken on by variables that affect all companies, such as economic activity or interest rates.

This common-variable interpretation of the variance of the predictions would conform with our earlier discussion of these forecasts. We found when examining the forecasts in section 2.2.4 that a common-factor model might well be appropriate. In section 3.2.3 we noted that differences of opinion might stem from two sources, one of which has a common-factor interpretation. Combining the specifications considered in section 3.2.3 might suggest that

$$\text{(3.4-13)} \qquad \begin{aligned} \mu_{ji} - \mu_j = \gamma_{j0i} - \gamma_{j0} &+ \sum_{k=1}^{K} (\gamma_{jki} - \gamma_{jk})(\mu_{ki} - \mu_k) \\ &+ \sum_{k=1}^{K} \gamma_{jk}(\mu_{ki} - \mu_k) + \sum_{k=1}^{K} \mu_k(\gamma_{jki} - \gamma_{jk}). \end{aligned}$$

For simplicity in describing the possible nature of s_g^2, assume that all individual differences in (3.4-13) are independent of each other so that we can consider s_g^2 to measure

$$\text{(3.4-14)} \qquad \begin{aligned} E(s_{gj}^2) &= E\left[\sum_{i=1}^{I} (\mu_{ji} - \mu_j)^2 / I\right] \\ &= E\left\{\left[\sum_{i=1}^{I} (\gamma_{j0i} - \gamma_{j0})^2 + \sum_{k=1}^{K} \mu_k^2(\gamma_{jki} - \gamma_{kj})^2 \right.\right. \\ &\quad \left.\left. + \sum_{k=1}^{K} \gamma_{jk}^2(\mu_{ki} - \mu_k)^2\right] / I\right\}, \end{aligned}$$

where I is the number of predictors. Variation in this quantity across firms arises partly from the variations in the evaluation of the individual firms but also partly from the differences in the values of the vectors of predicted values captured in the last term of equation (3.4-14). If this latter source of disagreement dominates the expression, then the variance (or the standard deviation) of the predictions of growth might be considered a measure of systematic risk. If differences of opinion about individual firms dominate, then it is more a measure of a type of specific risk. If only systematic risk in the sense of the γ_{jk} coefficients matters as the diversification theory suggests, then s_g^2 would only be important if the x_t variables used in estimating $\hat{\delta}_j$ failed to catch some type of common factor.

The second risk measure we consider is the estimated residual variances from the regressions of π_{jt} on x_t used to calculate the coefficients $\hat{\delta}_j$. More than the previous measures, this one could be taken to be a measure of specific risk. However, as shown in expression (3.4-8), systematic risk also enters this variable through the errors-in-variables problem. Nevertheless, if we first deal with the estimation errors in $\hat{\delta}_j$ coefficients, we might then consider use of this residual variance with these coefficients as a test indicating whether specific risk matters.

Arguments developed in this chapter allow us now to proceed to investigate the relationships between the predictions of earnings growth which we collected and the valuation of securities. The approach taken by the APT, especially using our diversification argument, to establish share prices provides a valuation equation which is robust to many variations in the specific assumptions used in its derivation. Given the hypothesis that many portfolios can be considered extensively diversified, a particular type of relationship should exist that relates security prices to expected returns and to risk measures. Reasonable arguments suggest that the predicted growth rates which we collected are closely related to and can be used to derive expected returns. Finally, in the present section, we have indicated how variables corresponding to the relevant risk measures may be obtained.