## Chapter 2: THE MARKET MODEL, INDEX MODELS AND FACTOR MODELS

### 2.1 INTRODUCTION

In this chapter, the central focus is on models that specify the return on a security i over some period t, $\underline{r}_{it}$, as a function $\Phi_{it}(\cdot)$ of a set of stochastic variables $\{\underline{\delta}_{jt}\}_{j=1,..,k}$:

$$(2.1) \quad \underline{r}_{it} = \Phi_{it}(\underline{\delta}_{1t},...,\underline{\delta}_{kt}) + \underline{\xi}_{it}$$

where the random error term $\underline{\xi}_{it}$ denotes the unmodelled part of the return. We could take the viewpoint that functions like (2.1) postulate how security returns are generated, hence we can call these models 'return generating processes' (RGPs for short). Note that (2.1) specifies a contemporaneous relationship between $\underline{r}_{it}$ and $\{\underline{\delta}_{jt}\}_{j\in k}$.

Many of these models are presented in the literature. Relating security returns to some variables allows a decomposition of these returns (and their variability) into a systematic part and an unsystematic part. Thus the fundamental concept of systematic risk emerges, central to portfolio theory. However, we emphasize that systematic risk is a relative concept and can materialize in various ways, depending a.o. on the explanatory variables that are considered.

In this chapter, we concentrate on distinguishing various forms of return generating processes by analyzing the implicitly or explicitly underlying assumptions. Apparently, there is a need for a typology of these models as well as for a consistent terminology. Many of the insights readily carry over to the theoretical analyses in chapter three and the empirical analyses in chapter four.

In this context we pay attention to the so called 'market model' and the 'single index model'. Section 2.2 and 2.3 discuss these models in both a theoretical and an empirical context. Then we step over to models in which the RGP is of a multi-dimensional nature: 'multi-index models' and 'multi-factor models'. Theoretical aspects of these models are discussed in section 2.4, whereas section 2.5 is devoted to their empirical context.

Section 2.6 summarizes the chapter.

## 2.2 THE MARKET MODEL (MM)

The 'market model' (henceforth abbreviated as MM) specifies a linear relationship between the return on individual securities and the return on the aggregate portfolio of all securities in the market. This model is frequently employed in financial research and various interpretations are attached to the model and its parameters.

In section 2.2.1 we first define the MM and discuss some of its general properties. Section 2.2.2 is devoted to the MM as a theoretical concept. We consider the case in which the MM is imposed on the relationship between returns on individual securities and the aggregate portfolio, and we investigate the conditions on the joint probability distribution of security returns under which the MM is implied. Section 2.2.3 finally is devoted to the MM as an empirical concept. The insights that are obtained are relevant for the discussion of index models and factor models later in this chapter, and serve as starting point for our theoretical discussions in chapter three.

### 2.2.1 The definition of the market model

The MM relates the return on a security to the return on the aggregate m of all securities in the market (the 'market portfolio') and is defined as (Fama [1976, p.69]):[1,2]

$$(2.2) \qquad \underline{r}_{it} = \alpha_i + \beta_i \underline{r}_{mt} + \underline{\varepsilon}_{it} \qquad \forall\ i \in m$$

where $\underline{r}_{it}$ and $\underline{r}_{mt}$ denote the return on security i and the 'market portfolio' in period t, respectively, and where $\alpha_i$ and $\beta_i$ are constants. Stochastic variables are underlined. The random error or disturbance term $\underline{\varepsilon}_{it}$ is usually assumed to satisfy:

$$(2.3) \qquad E(\underline{\varepsilon}_{it}) = 0$$

$$(2.4) \qquad E(\underline{\varepsilon}_{it} \cdot \underline{r}_{mt}) = 0$$

for $\forall\ i \in m$, where $E(\cdot)$ denotes the expectations operator.

---

1) Fama [1976, pp.66,73] assumes that the 'market portfolio' is equally weighted. Without loss of generality, we prefer to consider the 'market portfolio' as the capitalization weighted portfolio of all securities. In essence, the market portfolio is the composition of all existing assets and thus consists of all wealth whose return is uncertain.

2) The term 'market model' was first used by Fama [1968, p.37] to describe what we will call a single index model or a single factor model; see section 2.3.

In the MM, it is usually assumed that variances and covariances exist. Taking (unconditional) expectations of eq. (2.2) and using (2.3), we have:

$$(2.5) \qquad \alpha_i = E(\underline{r}_{it}) - \beta_i E(\underline{r}_{mt})$$

Also, using eq. (2.4), it follows that

$$\mathrm{Cov}(\underline{r}_{it}, \underline{r}_{mt}) = \mathrm{Cov}(\alpha_i + \beta_i \underline{r}_{mt} + \underline{\varepsilon}_{it}, \underline{r}_{mt})$$

$$= \beta_i \mathrm{Var}(\underline{r}_{mt})$$

so

$$(2.6) \qquad \beta_i = \frac{\mathrm{Cov}(\underline{r}_{it}, \underline{r}_{mt})}{\mathrm{Var}(\underline{r}_{mt})}$$

where $\mathrm{Cov}(\cdot,\cdot)$ and $\mathrm{Var}(\cdot)$ denote the covariance and variance operator, respectively.

This expression for $\beta_i$ follows from applying (2.4). Conversely, when $\beta_i$ is defined by (2.6), eq. (2.4) follows. At first sight, the assumption that $\underline{\varepsilon}_{it}$ and $\underline{r}_{mt}$ are uncorrelated may seem surprising. After all, since the 'market portfolio' m is a convex combination of all securities, the disturbance term $\underline{\varepsilon}_{it}$ is part of the return on m. Using eq. (2.2) we have:

$$(2.7) \qquad \underline{r}_{mt} = \sum_{i \in m} m_i \underline{r}_{it} = \sum_i m_i \alpha_i + \sum_i m_i \beta_i \underline{r}_{mt} + \sum_i m_i \underline{\varepsilon}_{it}$$

$$= \alpha_m + \beta_m \underline{r}_{mt} + \underline{\varepsilon}_{mt}$$

where $m_i$ denotes the relative (market value) weight of security i in m, satisfying $\sum_i m_i = 1$. Using eqs. (2.5) and (2.6) and the linearity of the covariance operator, we have $\alpha_m = 0$ and $\beta_m = 1$. Thus:

$$\underline{r}_{mt} = \underline{r}_{mt} + \sum_i m_i \underline{\varepsilon}_{it}$$

which in turn implies that

$$(2.8) \qquad \sum_i m_i \underline{\varepsilon}_{it} = \underline{\varepsilon}_{mt} = 0$$

So $\underline{\varepsilon}_{it}$ and $\underline{r}_{mt}$ are uncorrelated, since there exist (linear) dependencies between the securities' disturbances $\underline{\varepsilon}_{it}$ guaranteeing that their (weighted) average in m is always zero.[3] This implies that the residual covariance matrix is singular. We therefore refer to eq.(2.8) as the **singularity próperty** of the MM.

[3] Fama [1973, p. 1183; 1976, p.74], in contrast to Fama [1968, p.38].

In the next subsections we first examine the theoretical background of the MM. We pay special attention to the assumptions and to the relation with capital asset pricing theory. We then review the MM in an empirical context. For the sake of brevity, we will often suppress the time index t in the following.

### 2.2.2 The market model as a theoretical concept

#### 2.2.2.A The MM as a projection

As already noted by Beja [1972, p.38], the MM as specified by eqs. (2.2-4) does not involve any assumption at all. In fact, in terms of linear algebra the MM can be viewed as an orthogonal projection of the vector $\underline{\mathbf{r}}_i$ into the vector subspace that is spanned by the vector $\underline{\mathbf{r}}_m$ (Graybill [1983, p.84]). This construction is always possible. The vector $\underline{\boldsymbol{\epsilon}}_i$ with the disturbance terms then represents the orthogonal complement; in geometric terms, the vectors $\underline{\mathbf{r}}_m$ and $\underline{\boldsymbol{\epsilon}}_i$ are perpendicular. This means that $E(\underline{\epsilon}_{it} \cdot \underline{r}_{mt}) = 0$ (eq.(2.4)).

In statistical terms, the orthogonal projection is the least squares estimation to eq.(2.2).[4] Indeed, the structure of eqs.(2.2-4) can always be specified by forcing a linear least squares regression of $\underline{r}_{it}$ on $\underline{r}_{mt}$. The MM can then be interpreted as a linear regression function where $\alpha_i$ and $\beta_i$ are the intercept and slope coefficient.[5] A plot of (2.2) in $(r_{it}, r_{mt})$-space results in the "characteristic line" of security i (Treynor [1965, p.64]).

Forcing a linear least squares regression between $\underline{r}_{it}$ and $\underline{r}_{mt}$ is always possible and as a result of the orthogonality, $\underline{\epsilon}_i$ and $\underline{r}_m$ are uncorrelated (eq.(2.4)). In this way, without making any further assumptions, linearity is **imposed** on the MM. In the next subsections, we investigate the conditions under which the linear MM eq.(2.2) is **implied** by the stochastic properties of $\underline{r}_i$ and $\underline{r}_m$.

#### 2.2.2.B Bivariate normality

Fama [1973; 1976, Ch.3] shows that the MM arises as an implication of the assumption that the joint distribution of the security returns is multivariate normal.[6] From this assumption it follows that the returns

[4] Cf. Graybill [1983, p.437].

[5] Cf. Fama [1973, p.1183 fn.7]. Note that apart from the assumptions concerning 'spherical disturbances', the assumption of normally distributed error terms is necessary for the OLS-model only to make statistical inferences.

[6] The assumption that the returns on **portfolios** are normally distributed may at first sight seem a less stringent assumption. However, when returns on **all** linear combinations of individual

on individual securities as well as on portfolios are normally distributed. Hence, the joint distribution of the returns on any security i and any portfolio is bivariate normal. The bivariate normality of security and portfolio returns is the foundation of Fama's theoretical and empirical work on the MM.[7]

The bivariate normality of the joint distribution of $\underline{r}_i$ and $\underline{r}_m$ implies that the conditional distribution of $\underline{r}_i$, given that the return on the 'market portfolio' $\underline{r}_m$ takes a particular value $r_m$, is also normal. The mean of this conditional distribution is:

$$(2.9) \qquad E(\underline{r}_i|r_m) = \alpha_i + \beta_i r_m \qquad \forall\ i \in m$$

where $\alpha_i$ and $\beta_i$ are as defined by (2.5) and (2.6). This conditional mean is linear in $r_m$ and known as "the regression function" (Chow [1983, p.12]), the "linear regression function" (Spanos [1986, p.122]) or "the regression curve" (Mood, Graybill & Boes [1974, pp. 158,168] and Maddala [1977, p.98]).[8] The terminology on this point is not clear (cf. Fama [1973, p.1183 fn.7]), but it is important to make a sharp distinction between the **linear** regression function as described in 2.2.2.A and **the** regression function eq. (2.9).

The variance of $\underline{r}_i$ given $r_m$ is called the skedasticity function and is given by:

$$(2.10) \qquad Var(\underline{r}_i|r_m) = Var(\underline{r}_i) - \beta_i^2 Var(\underline{r}_m) = Var(\underline{r}_i)[\ 1 - \rho^2_{im}\ ]$$

where $\rho_{im}$ is the (product moment) correlation coefficient between $\underline{r}_i$ and $\underline{r}_m$. Eq. (2.10) shows that the conditional variance is homoskedastic, i.e. free of the conditioning variable $r_m$.

From the normality and homoskedasticity of $\underline{r}_i|r_m$, it follows then for any value of $r_m$, that the deviation of $\underline{r}_i$ from its conditional expected value,

$$(2.11) \qquad \underline{\varepsilon}_i = \underline{r}_i - E(\underline{r}_i|r_m) = \underline{r}_i - (\alpha_i + \beta_i r_m)$$

has a normal distribution with zero mean and variance given by (2.10). As the distributional results of $\underline{r}_i|r_m$ hold for any value of $r_m$, we can rewrite (2.11) as (2.2).[9]

---

securities (i.e. portfolios) are normally distributed, the joint distribution of all securities' returns must be multivariate normal. In a general context, however, Hamedani [1984] qualifies a statement like this.

7) Fama [1976, p.66]; see also Roll [1969, p.272 fn.4].

8) Spanos [1986, p.122] reserves the term "regression curve" for the graph $(r_m, E(\underline{r}_i|r_m))$.

9) Note that $E(\underline{r}_i|r_m)$ is non-stochastic. Alternatively, we could directly consider the conditional expectation $E(\underline{r}_i|\underline{r}_m) = \alpha_i + \beta_i \underline{r}_m$, which is a random variable. $E(\underline{r}_i|\underline{r}_m)$ is to be interpreted as the expectation of $\underline{r}_i$, conditional to the $\sigma$-field relative to which $\underline{r}_m$ is defined. See Spanos [1986, pp.125-126,317-318] on this point.

Combining these results: if the joint distribution of $\underline{r}_i$ and $\underline{r}_m$ is bivariate normal, the relation between $\underline{r}_i$ and $\underline{r}_m$ can statistically be described by the bivariate regression function eq.(2.2). This means that the (linear) MM is **implied** by the assumption of bivariate normality. In this multivariate framework, it is not necessary to make assumption eq.(2.4) that $\underline{\varepsilon}_i$ and $\underline{r}_m$ are uncorrelated. As Fama [1976, pp.69-70] shows, it results from the bivariate normality of $\underline{r}_i$ and $\underline{r}_m$. Using eq. (2.2) and the definition of $\beta_i$, we have:

$$(2.12) \quad \mathrm{Cov}(\underline{\varepsilon}_i, \underline{r}_m) = \mathrm{Cov}(\underline{r}_i - \alpha_i - \beta_i \underline{r}_m, \underline{r}_m)$$

$$= \mathrm{Cov}(\underline{r}_i, \underline{r}_m) - \beta_i \mathrm{Var}(\underline{r}_m) = 0.$$

A stronger statement is possible. Because $\underline{\varepsilon}_i$ is a linear combination of $\underline{r}_i$ and $\underline{r}_m$, the joint distribution of $\underline{\varepsilon}_i$ and $\underline{r}_m$ is also bivariate normal. The zero covariance in eq. (2.12) thus implies that $\underline{\varepsilon}_i$ and $\underline{r}_m$ are independent. Hence, the expected value and the variance of $\underline{\varepsilon}_i$ do not depend on $\underline{r}_m$:

$$(2.13) \quad E(\underline{\varepsilon}_i | \underline{r}_m) = E(\underline{\varepsilon}_i)$$

$$(2.14) \quad \mathrm{Var}(\underline{\varepsilon}_i | \underline{r}_m) = \mathrm{Var}(\underline{\varepsilon}_i)$$

### 2.2.2.C Linearity of MM

Forcing a linear least squares regression between $\underline{r}_{it}$ and $\underline{r}_{mt}$ is always possible and as a result of the orthogonality, $\underline{\varepsilon}_i$ and $\underline{r}_m$ are uncorrelated. In this way, no restrictions are placed on the joint distribution of $\underline{r}_i$ and $\underline{r}_m$ and the linearity between $\underline{r}_i$ and $\underline{r}_m$ is imposed. In the terminology of Stapleton & Subrahmanyam [1983, p.1638], this is the first-level specification of the MM.

Linearity is implied in the second-level specification, where it is assumed that the conditional expectation of the disturbance term with respect to $\underline{r}_m$ equals the unconditional expectation of $\underline{\varepsilon}_i$:[10)]

$$(2.13') \quad E(\underline{\varepsilon}_i | \underline{r}_m) = E(\underline{\varepsilon}_i) = 0$$

Substituting $\underline{\varepsilon}_i$ from the linear MM eq.(2.2) in (2.13'), we get:

$$(2.15) \quad E(\underline{r}_i | \underline{r}_m) = \alpha_i + \beta_i \underline{r}_m$$

When variances and covariances exist, eq. (2.13') implies that $\underline{\varepsilon}_i$ and $\underline{r}_m$

---

10) Note that (2.13') implies that $E\{\underline{\varepsilon}_i \cdot g(\underline{r}_m)\} = 0$ for any continuous function $g(\cdot)$. This situation lies between the independence and the uncorrelatedness of $\underline{\varepsilon}_i$ and $\underline{r}_m$ and can be characterized as 'semi-independence'.

are uncorrelated; hence, the definitions of the parameters $\alpha_i$ and $\beta_i$ in eqs.(2.5) and (2.6) readily follow.[11] So, whenever the conditional expectation of $\underline{r}_i$ with respect to $\underline{r}_m$ is linear in $\underline{r}_m$, the actual relationship between $\underline{r}_i$ and $\underline{r}_m$ is linear. The conditions under which the linear MM is **implied** by the stochastic properties of $\underline{r}_i$ and $\underline{r}_m$ thus coincide with the conditions under which the conditional expectation $E(\underline{r}_i|\underline{r}_m)$ is a linear function. Whenever the conditional expectation of $\underline{r}_i$ with respect to $\underline{r}_m$ is not linear in $\underline{r}_m$, the actual relationship between $\underline{r}_i$ and $\underline{r}_m$ is non-linear.

In the last two specifications of the MM, further restrictions are placed on the joint distribution of $\underline{\varepsilon}_i$ and $\underline{r}_m$. In the third-level specification, the homoskedasticity assumption of the disturbance term is added: eq. (2.14). The fourth-level specification finally entails the independence of $\underline{\varepsilon}_i$ and $\underline{r}_m$. In that case, the density of $\underline{\varepsilon}_i$ conditional on $\underline{r}_m$ equals the unconditional density of $\underline{\varepsilon}_i$.[12] Only in the case of bivariate normality, the zero covariance between $\underline{\varepsilon}_i$ and $\underline{r}_m$ does imply independence.

**Normality revisited**

As shown in section 2.2.2.B, the bivariate normality of $\underline{r}_i$ and $\underline{r}_m$ implies a third-level specification of the MM. Conversely, linearity and homoskedasticity (eqs. (2.13') and (2.14)) imply the bivariate normality of $\underline{r}_i$ and $\underline{r}_m$ (and hence the independence of $\underline{\varepsilon}_i$ and $\underline{r}_m$ according to the fourth level specification).[13]

The motivation for assuming that security returns are jointly normally distributed lies in the central limit theorem.[14] The continuous compounded return (the logarithm of the price relative, assuming either no dividends or dividends reinvested) on a security over some period is the sum of the changes in the log prices from one trading point to the next during the entire period under consideration. When these subperiod returns are (sufficiently) independent and their variance exists, then the distribution of the continuous compounded period returns will approach the normal distribution. In applying the central limit theorem, the time additive property of log returns is

---

[11] As eq. (2.13') implies (2.12), we actually do not need the definitions of $\alpha_i$ and $\beta_i$ to derive (2.12). This contrasts Fama [1976, p.78].

[12] For the random variables $\underline{x}$, $\underline{y}$ and $h(\underline{y})$, where $h(\cdot)$ is a continuous function, we have $E\{h(\underline{y})|\underline{x}\} = E\{h(\underline{y})\}$ if $\underline{x}$ and $\underline{y}$ are independent (cf. Spanos [1986, p.125]).

[13] Discussing theoretical aspects of the estimation of the MM, Fama [1976, p.78] states that the normality of the conditional distribution $\underline{r}_i|\underline{r}_m$ with mean (2.15) and homoskedastic variance is a less restrictive assumption than the bivariate normality of $\underline{r}_i$ and $\underline{r}_m$. The insistence of adding the homoskedasticity condition, however, invalidates this statement. See footnote 20 below.

[14] Cf. Fama & Miller [1972, p.260] and Fama [1976, pp.17-20].

used: log returns can be aggregated over time. Log returns, however, cannot be aggregated in cross-section over different securities. The log return $\underline{R}_p$ on a portfolio p with investment weights $\{x_i\}_{i\in p}$ can be computed as:

$$(2.16) \qquad \underline{R}_p = \ln(1+\underline{r}_p) = \ln\{ \textstyle\sum_{i\in p} x_i \cdot \exp(\underline{R}_i) \}$$

which clearly is a non-linear function of the log returns of the component securities.[15] So, in order to use the normality that is implied by the central limit theorem, we must assume that discretely compounded and continuous compounded returns are approximately equal. This approximation will only be reasonable when returns are limited to the range $-15\% \le \underline{r} \le 15\%$. We also refer to Appendix 4.B to chapter four for a discussion.

However, when focussing on linearity, bivariate normality is a sufficient and not a necessary condition, since there are several other conditions which imply a linear relationship between $\underline{r}_i$ and $\underline{r}_m$. As remarked before, these conditions coincide with the conditions under which the conditional expectation (2.15) is a linear function.

Under what conditions is the conditional expectation a linear function?

**Stable (Paretian) distributions**

The generalized version of the central limit theorem[16] states that if a (weighted) sum of random variables has a limiting distribution, the limiting distribution is a member of the stable (Pareto-Lévy) class. In the case in which variances are assumed to exist, the limiting distribution is the normal distribution. This is the case discussed above. When the variances do not exist, it follows from the generalized central limit theorem that the limiting distribution is stable (Paretian).[17] By definition, stable distributions are invariant or

[15] See also Tsiang [1973, p.751] on this point.

[16] See Mandelbrot [1963,1972], Fama [1963] or Fama & Miller [1972, p.264].

[17] This corrects Alexander & Francis [1986, p.144], who can only resort to a 'conceptual argument'.

The class of stable (Paretian) distributions is completely characterized by four parameters: the characteristic exponent $\theta$, which determines the tailedness and the type of a stable distribution; the skewness or symmetry parameter; the location parameter and the scale or dispersion parameter. For a thorough exposition of multivariate stable distributions, we refer to Press [1972].

In the financial literature, analyses are confined to symmetric stable distributions (Simkovitz & Beedles [1980] being an exception). For $\theta<2$, the variance does not exist and for $\theta<1$, the closed ('stable') under (weighted) addition.

Because of the stability property, stable (Paretian) distributions are desirable for portfolio models. When returns on individual securities follow stable (Paretian) distributions with the same characteristic exponent θ, the returns on all portfolios (including m) follow stable distributions with the same characteristic exponent.[18] Models like eq.(2.2) have been estimated assuming stable (Paretian) distributions (Fama [1965b], Blume [1970]). When the bivariate distribution of $\underline{r}_i$ and $\underline{r}_m$ is symmetric stable and when expectations exist (θ>1), then the conditional expectation of $\underline{r}_i$ with respect to $\underline{r}_m$ is linear in $\underline{r}_m$; this is motivated below. Hence, when $\underline{r}_i$ and $\underline{r}_m$ are bivariate symmetric stable (Paretian), the linear MM is **implied**. (See also section 2.3.2 below.) Except for the normal case, however, the distribution of the conditional expectation depends on the conditioning variable, so we encounter here the analogon of 'heteroskedasticity' in the finite variance case.

**Elliptical distributions**

(Symmetric) stable distributions are popular for modelling security returns. For θ<2, they can account for the observed 'fat tails' of empirical return distributions, and the stability property would ensure that (given the same θ) the distributions of individual securities and portfolios are of the same two-parameter type. There is however, another family of two-parameter distributions that allows for fat tails and that is closed under addition: the symmetric elliptical (or spherical) distributions.[19]

When the returns on the individual securities follow symmetric elliptical distributions, the returns on portfolios (including m) also follow symmetric elliptical distributions. The bivariate symmetric elliptical distribution of $\underline{r}_i$ and $\underline{r}_m$ then implies that the conditional expectation of $\underline{r}_i$ with respect to $\underline{r}_m$ is linear in $\underline{r}_m$, so the regression

---

mean does not exist. For θ=2, we have the normal distribution, for θ=1 the Cauchy distribution and for θ=½ the arc sine distribution. These are the only cases for which closed-form density functions are known.

[18] See Tsiang [1973, p.751] for a criticism. Because in general log returns are involved (in Mandelbrot [1963, pp.404ff] and Fama [1965a] e.g.), we encounter the same aggregation problem as in the normal case before.

Regarding the stability property, we note that one must distinguish between two cases: stability in a **time series** context (in the spirit of the generalized central limit theorem) and stability in a **cross-section** context (in the spirit of portfolio analysis). In a strict sense, both forms of stability are mutually exclusive.

[19] Owen & Rabinovitch [1983] discuss elliptical distributions in portfolio theory. A useful review and bibliography of this class of distributions is provided by Chmielewski [1981] (not cited by Owen & Rabinovitch [1983]).

function is linear (provided that expectations exist). The linearity of the conditional mean for the class of symmetric elliptical distributions is proven by Kelker [1970, p.424]; see also Chmielewski [1981, p.72]. Hence, when $\underline{r}_i$ and $\underline{r}_m$ are bivariate symmetric elliptical, the MM is **implied**. However, the conditional variance is heteroskedastic (provided that second moments exist). Indeed, homoskedasticity characterizes the normal distribution within the class of elliptical distributions.[20]

The symmetric stable distributions belong to the class of (symmetric) elliptical distributions (Press [1972, p.455]). This motivates the linearity of the MM in the context of symmetric stable (Paretian) distributions as postulated above.

**Symmetrical distributions**

All this implies that the linear MM is a very general model. Roll [1969, p.272 fn.4] already hinted that the linearity of the MM "probably holds for any bivariate **symmetric** distribution" (emphasis by Roll). However, it is unknown to what extent an approximate fit to the symmetric (stable or) elliptical distributions carries through to the approximate linearity of the conditional expectation. In chapter three, we will discuss the linearity of models in more detail.

#### 2.2.2.D Decomposition of security returns

The MM enables a decomposition of the return $\underline{r}_i$ on any security into the market or systematic component $\beta_i \underline{r}_m$ and the residual or unsystematic component $(\alpha_i + \underline{\varepsilon}_i)$.

Given the return decomposition (and assuming finite variances), a decomposition of the return variance readily follows:

$$(2.17) \qquad \mathrm{Var}(\underline{r}_i) = \beta_i^2 \mathrm{Var}(\underline{r}_m) + \mathrm{Var}(\underline{\varepsilon}_i) \qquad \forall\ i \in m$$

The first component, related to the variance of 'the' market portfolio's return, is the systematic variance; the second component is the unsystematic or residual variance. When variance is a valid measure for risk, these components relate to:

| | | |
|---|---|---|
| systematic risk | and | unsystematic or residual risk, |
| market risk | and | non-market risk, |
| non-diversifiable risk | and | diversifiable risk, |

respectively. As $\mathrm{Var}(\underline{r}_m)$ is a constant for all securities in m, $\beta_i$ is measure of systematic (market, non-diversifiable) risk.

[20] See Kelker [1970, pp.424-426] for details and Chmielewski [1981] for further references. See also Spanos [1986, p.458].

function is linear (provided that expectations exist). The linearity of the conditional mean for the class of symmetric elliptical distributions is proven by Kelker [1970, p.424]; see also Chmielewski [1981, p.72]. Hence, when $\underline{r}_i$ and $\underline{r}_m$ are bivariate symmetric elliptical, the MM is **implied**. However, the conditional variance is heteroskedastic (provided that second moments exist). Indeed, homoskedasticity characterizes the normal distribution within the class of elliptical distributions.[20]

The symmetric stable distributions belong to the class of (symmetric) elliptical distributions (Press [1972, p.455]). This motivates the linearity of the MM in the context of symmetric stable (Paretian) distributions as postulated above.

#### Symmetrical distributions

All this implies that the linear MM is a very general model. Roll [1969, p.272 fn.4] already hinted that the linearity of the MM "probably holds for any bivariate **symmetric** distribution" (emphasis by Roll). However, it is unknown to what extent an approximate fit to the symmetric (stable or) elliptical distributions carries through to the approximate linearity of the conditional expectation. In chapter three, we will discuss the linearity of models in more detail.

### 2.2.2.D Decomposition of security returns

The MM enables a decomposition of the return $\underline{r}_i$ on any security into the market or systematic component $\beta_i \underline{r}_m$ and the residual or unsystematic component $(\alpha_i + \underline{\varepsilon}_i)$.

Given the return decomposition (and assuming finite variances), a decomposition of the return variance readily follows:

$$(2.17) \qquad \mathrm{Var}(\underline{r}_i) = \beta_i^2 \mathrm{Var}(\underline{r}_m) + \mathrm{Var}(\underline{\varepsilon}_i) \qquad \forall\ i \in m$$

The first component, related to the variance of 'the' market portfolio's return, is the systematic variance; the second component is the unsystematic or residual variance. When variance is a valid measure for risk, these components relate to:

| | | |
|---|---|---|
| systematic risk | and | unsystematic or residual risk, |
| market risk | and | non-market risk, |
| non-diversifiable risk | and | diversifiable risk, |

respectively. As $\mathrm{Var}(\underline{r}_m)$ is a constant for all securities in m, $\beta_i$ is measure of systematic (market, non-diversifiable) risk.

---

[20] See Kelker [1970, pp.424-426] for details and Chmielewski [1981] for further references. See also Spanos [1986, p.458].

We emphasize, however, that 'systematic risk' and 'non-diversifiable risk' are **relative concepts** and can materialize in various ways. Although these expressions are often used in a general context, they are meaningless until is specified relative to what they are defined. As in the MM the return on 'the' market portfolio is the conditioning variable, systematic risk is synonymous with market risk and diversification is evaluated with respect to $\underline{r}_m$. From eq. (2.8) it follows that a portfolio is perfectly diversified when the investment weights of all securities coincide with the relative weights of these securities in m. Should in contrast a random variable other than $\underline{r}_m$ be chosen on which the security returns are conditioned, then the terms 'systematic' and 'non-diversifiable' are to be interpreted with respect to that variable.

Finally, the MM allows a decomposition of the covariance between the returns on any two securities as:

$$(2.18) \qquad Cov(\underline{r}_i,\underline{r}_j) = \beta_i\beta_j Var(\underline{r}_m) + Cov(\underline{\varepsilon}_i,\underline{\varepsilon}_j)$$

This implies that the comovement between the returns on any two securities can be attributed either to their joint relationship with $\underline{r}_m$ or to the covariance between their residual returns. The latter part of the covariance is termed 'extra-market covariance' (cf. Rosenberg [1974]). In sections 2.3 and 2.4, the interrelationship between security returns will be discussed in more detail.

#### 2.2.2.E Relation to CAPM

After the discussion of the various aspects of the MM, it is worthwile to consider its relationship with the Capital Asset Pricing Model (CAPM) of Sharpe [1964], Lintner [1965a,b] and Mossin [1966]. Basically, the CAPM assumes a perfect and competitive capital market where expected utility-maximizing risk averse investors have identical *ex ante* beliefs over a one-period horizon, say period t. For any of several reasons, investors choose $(E,\sigma^2)$-efficient portfolios of securities.[21] As the market portfolio is simply the sum of the individual portfolios, it is also efficient. Mathematically equivalent to the $(E,\sigma^2)$-efficiency of the market portfolio is the linearity of the risk premium of any security i in $\beta_i$, the well-known Security Market Line (SML):[22]

$$(2.19) \qquad E(\underline{r}_{it}) = r_{ft} + \beta_{it}\,[\,E(\underline{r}_{mt}) - r_{ft}\,]$$

[21] More generally, the CAPM deals with capital assets, i.e. all terminal wealth producing assets. Cf. footnote 1 above.

[22] The equivalence between the efficiency of the reference portfolio and the linearity of the risk premia in the betas with respect to this portfolio was already hinted at by Sharpe [1964, pp.440-441] and analyzed more elaborately by Merton [1972], Roll [1977] and Ross [1977a].

where $r_{ft}$ is the risk free rate of return over the period t and $\beta_{it}$ equals $\beta_i$ as defined as in eq.(2.6).

The major distinction between the MM and the CAPM is that no general equilibrium conditions are expressed in or implied by the MM. Whereas the CAPM (or the SML) is formulated in an *ex ante* context in terms of expectations, the MM is defined in an *ex post* context in terms of actual (i.e. stochastic) returns. This is the most distinctive feature between both models.

**The MM and the derivation of the CAPM**

In the standard derivations of the CAPM, there is no role whatsoever for the MM. This implies that the validity of the one model does not depend on the validity of the other.[23)]

Stapleton & Subrahmanyam [1983] and Kwon [1985], however, have shown that the usual sufficient conditions for $(E,\sigma^2)$-analysis, i.e. quadratic utility or multivariate normality, can be replaced by the linearity assumption of the MM (eq.(2.13')) in deriving the CAPM.[24)] As this linearity is implied by the broader class of symmetric elliptical distributions, we note that their methodology forms an alternative route to derive the CAPM-type pricing results of Owen & Rabinovitch [1983]. The same argument applies to the pricing results of Fama [1971] for a symmetric stable (Paretian) market.

**The MM and the interpretation of the CAPM**

Aside from the beforementioned alternative derivations of the CAPM, there exists no formal relationship between the MM and the CAPM. However, as both models share the parameter $\beta$, the MM is generally used to provide an intuitive basis for the risk measure $\beta$ that is appropriate in the theoretical framework of the CAPM. In this context, Sharpe [1964, p.438] refers to the "economic meaning" of $\beta$: it measures the degree of responsiveness of a security's return to changes in the market portfolio's return (or 'the level of economic activity'). Unfortunately,

---

23) The point that the MM is no element of the CAPM is stressed by Jensen [1967,p.83; 1972,p.364 fn.23; 1979,p.24], Fama [1968, pp.37-38] and Beja [1972, p.38], among many others. However, for those who may think this is an historic issue, we refer to the recent paper by Treynor [1993]. In addition, the discussion is blurred by an almost general confusion between the MM and the single index model; see section 2.3.

24) Ross [1982, pp.67ff] also shows that the validity of CAPM does not critically depend on quadratic utility or multivariate normality. However, when homoskedasticity is imposed on the conditional distribution $\underline{r}_i|\underline{r}_m$, multivariate normality is implied and nothing is gained over the standard assumptions of the CAPM (as noted by Stapleton & Subrahmanyam [1983, p.1641]); see section 2.2.2.C.

Sharpe [1964, p.438 fn.23] explicitely refers to the single index model. The single index model differs from the MM as it includes the additional assumption that the residual returns of all securities are mutually uncorrelated (see Sharpe [1963] and our section 2.3). The insistence of adding this assumption to the MM assumptions caused a lot of confusion in interpreting the pricing equations[25] and led many researchers to escape to a single factor model, where the return on the market portfolio is substituted by the return on a 'market factor'.[26]
This confusion was eventually resolved by Beja [1972] and Fama [1973]. More recently, Stapleton [1980], Rosenberg [1981] and Markowitz [1984] distinguish between $\beta$-coefficients from different models and discuss their relation to the CAPM.[27]

Synonymous to the 'economic meaning' of $\beta$ in the CAPM is the interpretation of $\beta$ as a measure of systematic risk. Some of the beforementioned confusion in interpreting pricing relationships can be traced to the use of the term 'systematic' without explicit reference to the underlying variable with respect to which the dichotomy between systematic and unsystematic is defined. As the decomposition of risk in a systematic and an unsystematic part follows from the MM and not from the CAPM (see section 2.2.2.D), we argue that an intuitive exposition of the CAPM can better be based on the notion of $\beta$ as a measure for the marginal contribution of a security to the market portfolio's risk (see chapter three, section 3.4.1). When following that line of reasoning, a misconception of the MM as the appropriate univariate or 'one-factor' RGP for the CAPM is averted.[28]

**The MM as a RGP for the CAPM**

The most general form of a RGP is:

$$(2.20) \quad \underline{r}_{it} = E(\underline{r}_{it}) + \underline{\eta}_{it}, \qquad \forall\ i \in m,$$

[25] See for instance Lintner [1965b], Sharpe [1966] and Fama [1968, pp.37ff].

[26] See for instance Fama [1968, p.39], Jensen [1968, pp.391ff] and Jensen [1969, pp.179ff].

[27] Pettit & Westerfield [1974, p.581] and Jensen [1979, p.14], and more recently, Fogler & Ganapathy [1982, p.18], Alexander & Francis [1986, pp.75,116-117], Smith, Stephens & Proffitt [1992, p.122] and Reilly [1994, p.278], to name a few, insist on adding the assumption of mutually uncorrelated residuals to the MM in the context of $\beta$-estimation and the CAPM. Only in the fourth edition of their book, Elton & Gruber [1991, p.123] discuss the MM and differentiate it from the single index model (cf. Elton & Gruber [1987]).

[28] In fact, the MM's singularity property, discussed below eq.(2.8), invalidates the often used (naive) diversification argument for indicating the relevance of 'systematic' risk and beta in a CAPM context. Examples are Shapiro [1991, p.107] and Ross, Westerfield & Jaffe [1993, pp.292ff]. This argument implicitly refers to the use of the single index model.

where $\underline{\eta}_{it}$ is the deviation of the actual (*ex post*) return from the expected value.[29] In the context of the CAPM, it is tempting to place some restrictions on the RGP. More specific, by substituting eq.(2.5) in (2.2), we get:

$$(2.21) \quad \underline{r}_{it} = E(\underline{r}_{it}) + \beta_i [\underline{r}_{mt} - E(\underline{r}_{mt})] + \underline{\varepsilon}_{it}$$

The error process $\underline{\eta}_{it}$ is now decomposed into a systematic component $\beta_i [\underline{r}_{mt} - E(\underline{r}_{mt})]$ and a non-systematic part $\underline{\varepsilon}_{it}$.

The CAPM, however, does not **assume** a specific model that generates the security returns over the period t. Instead, as noted by Connor [1984, pp.28-29], the CAPM uses security returns to **construct** a RGP. As the security returns determine the value of the 'return generating factor' $\underline{r}_{mt}$, rather than the factor value affecting the returns of the securities, the MM can be seen as an ***'ex post* factor model'**. This implies that the constructed MM in eq. (2.2) or (2.21) **as if** the risky returns are generated by just one market 'factor'[30] is a completely valid representation, irrespective the process that actually generates the returns.[31] As the market portfolio by definition is well-diversified with respect to itself, the singularity property eq. (2.8) is always satisfied.

As the MM does not entail any restriction on the RGP and the CAPM does not assume any specific RGP, it is not surprising that the MM in eq.(2.2) or (2.21) is fully consistent with the CAPM.

Summarizing, the MM in eqs.(2.2) or (2.21) does not entail any restriction on the one-period RGP and is fully consistent with the CAPM. Reversely however, as discussed in chapter one, section 1.5, it can be shown that the CAPM does place a constraint on the parameters of the constructed RGP. Substituting the SML eq. (2.19) in (2.21), we get:

$$(2.22) \qquad \underline{r}_{it} = r_{ft} + \beta_{it} [\underline{r}_{mt} - r_{ft}] + \underline{\varepsilon}_{it}$$

(where the time subscript for $\beta_i$ is added to stress the one-period context). This expression can be regarded as an '*ex post* SML' and was first developed by Jensen [1968, pp.392-393] along a somewhat different route. Comparing eq.(2.22) with (2.2), the constraint that the CAPM places on the RGP is:

---

29) As $E(\underline{\eta}_{it}) \equiv 0$, this specification is indeed completely general and can be contrasted with the 'fair game' model (Fama [1970, p.385]), which specifies that the expected return, conditional on some begin-of-period information set $\phi_{t-1}$, $E(\underline{r}_{it}|\phi_{t-1})$, equals the unconditional expected return $E(\underline{r}_{it})$. This implies $E(\underline{\eta}_{it}|\phi_{t-1}) = E(\underline{\eta}_{it}) = 0$.

30) Or two orthogonal factors in Black's [1972] zero-beta formulation of the CAPM.

31) Sharpe [1977], for example, discusses the CAPM in the context of a multi-factor RGP.

(2.23) $\alpha_{it} = [1-\beta_{it}]\cdot r_{ft}$

To avoid confusion, it is appropriate to make a distinction between the 'simple MM' eq.(2.2) and the 'MM in excess return form', where the risk free rate is subtracted from the returns:[32)]

(2.24) $\underline{r}_{it} - r_{ft} = \alpha'_i + \beta_i[\underline{r}_{mt} - r_{ft}] + \underline{\varepsilon}_{it}$

The restriction eq.(2.23) can now be translated as $\alpha'_i = 0$.

### 2.2.3 The market model as an empirical concept

#### 2.2.3.A The MM in practice: the market index model

Sofar, the discussion of the MM was purely theoretical. As 'the' market portfolio is not observable in practice, eq. (2.2) has no empirical content. In empirical applications of the model, the market portfolio is therefore proxied, usually by some stock market index. The first question that arises is what index would be appropriate to use as a conditioning variable. According to the population of the index, the method of weighting and the method of construction, several stock market indices can be distinguished (cf. Frankfurter [1976, p.950]). In practice, a capitalization weighted stock market index is mostly used as a proxy for the market portfolio. However, the index choice depends on the specific application of the MM. Using a proxy index in empirical tests of the CAPM leads to well-known problems (Roll [1977,1978]).

Using a (proxying) market index, the MM reduces to a 'market index model', which is the relevant model in an empirical context. Having noted this, we will use the terms MM and market index model as synonyms.

In the following, we briefly discuss some applications of the MM, starting with the CAPM (thus extending section 2.2.2.E to an empirical context). We then consider some issues that are relevant in estimating the model. It is not unwarranted that we cover this topic in some greater detail, because the insights gained from studying empirical problems related to the MM readily carry over to the other models treated in this chapter.

---

32) As the risk premium equals the **expected** excess return, we prefer the term 'excess return version' over the term 'risk premium version' (as used by Harrington [1987, pp.117, 104 fn.2], among others). This argument is not inconsistent with our use of the term 'risk premia' in chapter one, section 1.4.2 (footnote 48). Although in that context no expected value is considered, instead a cross-section restriction is placed upon the coefficients.

### 2.2.3.B The MM as a bridge between *ex ante* CAPM and *ex post* tests

From a theoretical point of view, the MM is popular in that it provides an intuitive basis for explaining the relevant risk measure $\beta$ in the CAPM. From an empirical point of view, the MM is popular as a beta-generator. As the SML has an *ex ante* character and tests of the CAPM necessarily have an *ex post* character, the MM serves as a bridge between the *ex ante* context and the *ex post* formulation in that it specifies how the *ex post* returns are 'generated' (eq.(2.22).

The issue of an appropriate RGP for the CAPM arises in empirical testing because the CAPM is casted in terms of *ex ante* expectations while the empirical tests are generally confined to *ex post* realized return data.[33] As Sharpe [1978, p.918] notes: "Any test using *ex post* values is a joint test of a theory of asset pricing and a model of a return generating process" (italics by Sharpe). In a general way, one could assume that the probability distributions generating the *ex post* returns are stationary over time and that expectations are on average and on the whole correct, so that sample average returns can be substituted for *ex ante* expected returns.

The question remains how to estimate the *ex ante* beta coefficients. General aspects of this empirical issue will be pointed out below. At this point, we only mention that the estimation of the MM in excess return form, eq.(2.24), must satisfy certain assumptions in order to serve as an operational counterpart of CAPM. In this context, Roll [1969] was the first to recognize the biases that can occur, notably from measurement errors in the risk free rate $r_{ft}$ and from covariance between $r_{ft}$ and the index return $r_{mt}$. In addition, we note that it is not always justified to adjust the regression parameter $\beta_i$ with the aim to improve on the specification of the model. For example, a correction for (autoregressive) residual heteroskedasticity (in a time series formulation) or the use of robust estimators (like least absolute deviation estimators) can yield an unbiased estimate of **a** regression slope parameter, but this is no longer **the** risk measure $\beta$ as specified by the standard CAPM.[34]

### 2.2.3.C Other uses of the MM

In the MM, security returns are conditioned on the market (index) return. This conditioning aspect is the reason for two other uses of MMs. In application of MMs in event studies[35], the noisy security returns are filtered for fluctuations that cannot be attributed to the event under study, thus facilitating the study of residual returns and

[33] Friend, Westerfield & Granito [1978] is an exception: they use *ex ante* data from a sample of financial institutions.
[34] See also chapter 5, section 5.2 (footnote 5), for a related issue.
[35] Fama [1976, p.77] and Brown & Warner [1980].

the event's influence. No pricing argument can be invoked to accentuate on the 'normal' (i.e. explained) return, and there is no need for the residual return component to be security specific.[36] Instead, the underlying argument is that conditioning reduces variance.

The same argument implies another use of the MM. As indicated by Sharpe [1964, p.439], the (CAPM-) $\beta$ indicates the predicted response of $r_i$ to changes in $r_m$. This is the basis for using the MM for forming conditional predictions: given some market return and the beta, a prediction for the security return can be specified. This prospective use of the MM underlies timing strategies (cf. Treynor & Black [1973] and Klemkosky & Maness [1978]). In a more general context, we will return to this issue in chapter three, section 3.4.2, and chapter five, section 5.4.2.

### 2.2.3.D Estimating the MM

Estimating the MM involves a number of decisions, for example with respect to the particular form of the model, the incorporated index, the estimation procedure, the length of the sample estimation period and the length of the observation intervals within the sample period.

#### Stationarity and stability

The first three issues have been touched on in the foregoing and will here be taken for granted. The latter two choices are less arbitrary when we recognize that in estimating the parameters of the MM, the researcher is confronted with the issues of stationarity and stability. Both issues carry over to the estimation of various return generating processes, including factor models in which we have special interest.

We have **stationarity** when the model parameters are invariant under a change in the length of the sample period T or under a change from one sample period $T_1$ to another period $T_2$. Intertemporal stationarity is relevant for prospective uses of the MM as it determines the value of *ex post* information (historical return data) in an *ex ante* context (the *ex ante* value of beta). We have **stability** when the model parameters are invariant under changes in the length of the observation interval, given the length of the sample period. The observation (or measurement) interval is the interval over which the returns are

[36] This contrasts a fairly general opinion; see for example Bowman [1983, p.563].

compounded and refers to the appropriate time horizon.[37] In the early days of portfolio theory, the stationarity of return variances and covariances was commonly accepted (cf. Baumol [1966, pp.98-99], e.g.). Joyce & Vogel [1970], however, first stressed the ambiguity in estimating return variances as a result of different lengths of the sample period and different lengths of the observation interval.

The seminal studies of Altman, Jacquillat & Levasseur [1974] and Pogue & Solnik [1974] first pointed at the latter intervalling aspect in the context of the MM. Since then, a large body of literature has been build around the stability problem.[38] Especially when high-frequency (daily) data are used, non-synchronous security trading gives rise to an errors-in-variables problem and causes biased estimators. Several corrected estimators are proposed in order to cope with this problem and the induced residual autocorrelation.[39] It is important to recognize the stability issue and the problems it can generate. When using the model for conditional predictions, it is then important to equate the length of the estimation intervals and the prediction interval.

For prospective uses of the MM (and of course, any other model), the problem of non-stationarity of its parameters is more severe and much more difficult to cope with. The stationarity of beta is usually measured by the correlation or the mean absolute deviation between a security's or portfolio's beta estimated in one period and the beta of that security or portfolio estimated in a subsequent period. The former beta might be regarded as an assessment of the future beta and the latter beta can be regarded as the realized beta (Blume [1971, p.7]). The correlation or the mean absolute deviation between these two betas can then be interpreted as a measure of the accuracy of the beta assessments. It is important to realize that the forecasted beta is compared with an estimated beta, and not with the unobservable true beta.[40] Errors in the beta estimates may thus cloud the test results.

---

[37] Different meanings are attributed to the concept of 'time horizon', for example: "the *anticipated holding period* of a security or portfolio (...), the *period over which investment results are to be measured* and the *time in which a specific objective is expected to be attained*" (Block [1969, p.124], his italics). Compare with the definitions in chapter one, section 1.1.2.

Empirical studies on the examination of beta stationarity typically use a prediction period of length identical to the estimation period, irrespective of the observation interval.

[38] For example Smith [1978a], Scott & Brown [1980], Hawawini [1983] and, more recently, Handa, Kothari & Wasley [1989,1993], to whom we refer for further references.

[39] For example Scholes & Williams [1977] and Dimson & Marsh [1983]. More recently, Shanken [1987] discusses this topic in a more general estimation context.

[40] Blume [1971, p.8 fn.22] and Garbade & Rentzler [1981, p.578].

Below we present several characteristic aspects of parameter non-stationarity. Although the discussion is primarily cast in terms of the MM and beta, we emphasize those aspects that can be expected to arise for any other (security return generating) model. Of course, any specific conclusions cannot automatically be transposed to other models.

**The portfolio size effect**

The pioneering study in examining the statistical properties of beta-estimates is Blume [1971]. He estimated security betas over non-overlapping seven-year periods and allocated the securities to various sized portfolios according to their beta-rank. Comparing the portfolio betas estimated in one period and the betas of the same portfolios estimated in the next period, he found that the (rank order and product moment) correlations of the portfolio betas were quite low for individual securities and small portfolios but quite high for large portfolios.

Porter & Ezzell [1975], however, use randomly selected portfolios instead of beta-ranked portfolios and find no relationship between portfolio size and beta correlations. Alexander & Chervany [1980, pp.131-133] resolved this 'portfolio size controversy' by noting that the betas of randomly selected portfolios will be scattered very tightly around one and that the correlation can no longer detect the degree in which betas change. Using the mean absolute deviation of successive portfolio betas instead of correlation, they find that this measure decreases with portfolio size, regardless the way in which portfolios are formed. Tole [1981] confirms this for randomly selected portfolios but argues that the portfolio size effect is not exhausted after including 10 or 25 securities in the portfolio (as the former studies concluded) but that substantially more (i.e. ±100) securities are needed.[41]

Clearly, by forming portfolios, measurement errors in individual securities' betas are averaged out. Increasing portfolio size then implies an increasing diversification of estimation errors, which could explain the portfolio size effect. In addition, portfolio aggregation smoothes out any changes in the security betas which are due to firm specific circumstances, thus improving the ability to perceive a common non-stationarity in the betas. In this light, it is then important to note that the studies on the portfolio size effect do not involve explicit tests of beta stationarity. It is even more important to note that biases can occur when forming portfolios on the basis of beta-(parameter-) ranks. Although this may seem an attractive procedure to form portfolios that exhibit maximum differences in beta (or any other ranking parameter), one may actually maximize estimation error instead of minimizing it. This leads us to the next aspect.

[41] See also the discussion by Christofi [1983] and Tole [1983].

**The regression tendency**

When comparing successive estimated betas of the ranked portfolios over time, Blume [1971] discovered a consistent 'regression tendency towards the mean' (portfolios with a high or low beta estimated in one period tended to have less extreme betas estimated in the next period) and suggested a simple correction procedure to enhance *ex ante* beta estimates.

A conventional explanation for the regression tendency is the 'selection bias' or 'order bias'[42] which, given measurement errors in the estimated betas, is caused by the way in which portfolios are constructed. More specific, when portfolios are constructed using the ranked beta values, the securities in the high (low) beta portfolio tend to have positive (negative) estimation errors in their betas. This implies that positive and negative sampling errors are concentrated in different portfolios, undermining the intention to diversify sampling errors by forming portfolios and leading to less extreme portfolio betas in the next period. This would imply that estimated betas would regress towards the mean, even if true betas were constant over time.

In a subsequent paper, Blume [1975] formalized the regression tendency by using properties of jointly normally distributed random variables, and decomposed the effect in an order bias and a true regression tendency. He showed that the order bias will also be present in the estimated betas for individual securities and is not induced by portfolio construction: the order bias arises from the existence of **estimation errors**. Thus even if the cross-sectional distribution of true betas is stationary over time and true betas are perfectly correlated, the estimated betas can exhibit the regression tendency. This tendency then is a statistical artifact. However, after adjusting for the order bias generated by measurement errors, Blume [1975] found that much of the original regression tendency remained. This true regression tendency indicates that true betas are **imperfectly correlated** over time. Elgers, Haltiner & Hawthorne [1979] find that the true regression tendency of betas (i.e. after correcting for the order bias) works not only forward but also backward in time, thus confirming the imperfect correlations between betas over time.[43,44]

---

[42] Black, Jensen & Scholes [1972, p.84], Fama & MacBeth [1973, p.615] and Blume [1975, pp.785-788].

[43] More recently, Kolb & Rodriguez [1989] critically discuss the regression tendency and point out that a common interpretation of the regression tendency in the literature as a 'beta drift' ('betas exhibit a long term drift toward one') is not justified by Blume's [1975] theoretical analysis and rests on the additional assumption that the distribution of betas collapses towards one.

[44] As an explanation for the true regression tendency, Blume [1975, p.786] gives "unstated economic or behavioral reasons", or more specific (p.795): "the risk of existing projects may tend to become less extreme over time (...) [or] new projects taken on by firms may tend to have less extreme risk characteristics than existing

Considering the above, we expect this regression tendency to be a fairly general phenomenon. Both from estimation errors in estimated parameters (which is unavoidable) and from imperfect correlation in true parameter values (which is very likely, whether the underlying parameter distribution is stationary or not), we may expect extreme parameter estimates to revert to the mean.

In order to cope with estimation errors, increasing the number of observations is an alternative for the cross-section solution of portfolio aggregation. This leads to the choice of the length of the estimation period.

**Length of the estimation period**

Baesel [1974, p.1493][45] and Altman, Jacquillat & Levasseur [1974, p.1499] (for the French stock market) find that the non-stationarity of individual security betas declines considerably as the length of the sample period increases.[46] However, increasing the length of the sample period in order to minimize sampling error has its counterpart in increasing the bias in parameter estimates due to possible structural changes. This launched a quest for the 'optimal estimation interval'. Alexander & Chervany [1980], for example, empirically find that 4 to 6 years of monthly data show the greatest conditional predictive accuracy. Even today, we experience the heritage of this result, in that 5 years of monthly data is generally considered to be an optimal estimation interval. Theobald [1981] (who does not refer to Alexander & Chervany

projects". Inspired by these explanations, Elgers, Haltiner & Hawthorne [1979] argue that the distribution of betas collapses towards the mean value of one, thus explaining part of the regression tendency. The results of Blume [1979], however, do not support this hypothesis. Furthermore Kolb & Rodriguez [1989] find that the tendency of extreme betas to move toward the mean is offset by the tendency of betas near the mean to move toward the extremes. These two effects maintain the stationarity of the distribution of betas, thus confirming Blume's [1979] assertion that the distribution of betas exhibits no tendency to collaps toward one. Kolb & Rodriguez [1990] present additional evidence that the distribution of betas, estimated over 60 month periods, is approximately stationary and does not collaps.

45) His finding that the non-stationarity of betas depends inversely on the extremity of beta is shown to be a statistical artifact (see Alexander & Chervany [1980, p.124]). It arises from allocating securities to portfolios according to their beta-rank. As a result, there is a wide range of beta values in the extreme fractile portfolios and a narrow range in the middle fractile portfolios.

46) Levy [1971b, p.60] finds similar results for short term (13, 26 and 52 week) estimated betas. The results of Roenfeldt, Griepentrog & Pflaum [1978, p.120], however, indicate that an estimate of a 'one-year beta' can better be based on an estimation period longer than one year.

[1980]) develops an analytical framework in which he demonstrates that, although beta-stationarity is an increasing function of the length of the estimation period, the stationarity will not increase indefinitely with estimation period length. For UK data, he suggests an average sample period of 120 to 210 months.

In this respect, the 'optimal estimation interval' shows resemblence with the holy Grail: everybody looks for it and expects to find it in other places. Instead of seeking an infallible recipe, it is in our opinion better to recognize the existence of non-stationarity and the changing effects it can have on parameter estimates. Next one could try to model the parameter changes or —preferably— try to figure out what possible underlying mechanisms of non-stationarity may be.

**Modelling beta non-stationarity**

Relaxing the assumption that beta is stationary over the estimation period ($\beta_t = \beta$) implies that standard estimation procedures (ordinary least squares, OLS) are no longer adequate. When there is evidence that beta is stochastic rather than fixed, it becomes relevant whether beta is purely random or autocorrelated. That is, in order to apply an adequate estimation procedure, it is necessary to know what process governs the time series behavior of beta.

There are various attempts to explicitly model the behavior of the MMs parameters in general and betas in particular, over time.[47]

According to the random coefficient model, beta varies randomly around the mean value ($\underline{\beta}_t = \beta + \underline{u}_t$). Fabozzi & Francis [1978] introduce this model and present evidence that that this model is valid for a significant minority of NYSE stocks. Lee & Chen [1980] improve their estimation method and also find support for the model, as does Sunder [1980]. Conflicting evidence is presented by Alexander & Benson [1982] and Ohlson & Rosenberg [1982].

In the varying parameter model, beta follows a random walk ($\underline{\beta}_t = \underline{\beta}_{t-1} + \underline{u}_t$; Kantor [1971], Garbade & Rentzler [1981]). Dotan & Ofer [1984] test an amended version of this model and conclude that there is no justification for using this model instead of the simple OLS model. Simonds, LaMotte & McWorther [1986], however, find considerable evidence in favor of this model.

Another class of models are the models in which beta follows an autoregressive process over time ($[\underline{\beta}_t-\beta] = \gamma[\underline{\beta}_{t-1}-\beta] + \underline{u}_t$). Schaefer, Brealey, Hodges & Thomas [1975] find slight evidence for this model compared to the random coefficients and the random walk model, but Bos & Newbold's [1984] results are inconclusive. Ohlson & Rosenberg [1982]

[47] Another route of research does not look for gradual changes but investigates whether structural changes occur in the process generating returns. Kon & Lau [1979], McDonald [1983] and Lee [1985], for example, confirm the existence of structural changes in the MM's parameters.

generalized this model and find that there is both a tendency for betas to converge slowly towards their mean and a purely random perturbation in beta. Collins, Ledolter & Rayburn [1987] test the Ohlson & Rosenberg [1982]-model against the random coefficients model and find that portfolio aggregation strengthens the autoregressive tendencies of beta. D'Souza, Brooks & Oberhelman [1989] in turn expand Ohlson & Rosenberg's [1982] model and support their findings that the stochastic components of betas may have a more complicated time-dependent structure than a simple random process.

Chang & Weiss [1991] finally, do not make strong *a priori* assumptions about the time series behavior of betas and empirically find that the behavior of betas conforms to an ARMA(1,1)-process.

When considering this parade of research results, it is difficult to draw any conclusions, except perhaps for slight autoregressive tendencies of beta. McDonald [1983, p.175] compares different models of beta non-stationarity and concludes that "attempts to measure continuous changes in beta do not appear to have any empirical significance." He contributes the lack of statistical significance of the random coefficients model and the random walk model to two factors: the models appear to have some difficulty in partitioning the variation in beta into its various components based on the sample information, and the assumption of a stationary stochastic process implicit in these models. Considering this argument, as well as the notion that the identification of a process governing betas over time largely is an empirical question, it is interesting to analyze explanations that are suggested for the non-stationarity of betas over time.

**Explaining beta non-stationarity**

When true beta is a random coefficient (or exhibits non-stationarity otherwise) but its value is forced to be constant in a regression, part of the market (index) return will show up in the error term, thus inducing heteroskedasticity. This heteroskedasticity, in turn, reduces the efficiency of the least squares estimators of the MM parameters by inflating their variances. The increased sampling error, finally, could in part explain the observed beta non-stationarity (cf. the portfolio size effect, discussed before).

The presence of (residual) heteroskedasticity in the MM and its implications are well documented; see for example Schwert & Seguin [1990] and the references cited therein. At this point, however, we stress that when the MM is **implied** by the (stationary) joint distribution of security returns and when this distribution is non-normal (which is commonly accepted), then homoskedasticity is **automatically** ruled out (cf. section 2.2.2.C). As in this case a security's expected return, conditional on the market (index) return, is a linear function of the market return, the true beta is constant. This in turn implies that the implied heteroskedasticity cannot be attributed to a non-constant beta in this case. As we should not be very surprised by the presence of heteroskedasticity in the MM, we warn against drawing inferences with respect to beta non-stationarity.

From non-linear dependencies like heteroskedasticity, it is a little step to linear dependencies: autocorrelated MM residuals (for example because of non-synchronous trading; see before). When autocorrelation patterns change over time because of irregular trading activity or changes in market thinness, we expect to observe changes in ordinary least squares estimates of the MM parameters. Also in this case, true betas may or may not be constant.

Considering the definition of beta in eq.(2.6), we can allocate changes in beta to changes in either the variance of the market return or the correlation with the market return, or both. Meyers [1973a], for example, finds significant changes in the explanatory power of the 'market factor' (the first principal component) from period to period; Francis [1979] also reports beta changes as a result of changes in the correlation with the market return.

In this context, a number of studies concentrates on the identification of micro- or macro-economic variables that influence beta-stationarity. Officer [1973] investigated changes in the variability of the market factor and found that these changes appear to be related to changes in the Index of Industrial Production (a measure of business activity in the US) and changes in the M2 money supply. Levy [1971a] related shifts in beta to alternating bull and bear market conditions. Fabozzi & Francis [1977] reject the influence of bull-bear market forces on beta, but conclude later that the intertemporal non-stationarity of beta appears to result from changes which are associated with business cycles (Francis & Fabozzi [1979]).

Using the concept of duration, Bildersee & Roberts [1981] relate changes in beta to changing interest rates. More recently, DeJong & Collins [1985] start from the joint Option Pricing Model / CAPM framework and find a significant relation between beta non-stationarity and the degree of leverage of the respective firm and changes in the risk-free interest rate. In a more general context, McDonald [1985] considers major structural changes in the economy. These changes were accompanied by large changes in interest rates and yielded shifts in betas. Further evidence concerning these changes in the structure of the economy and the variance-covariance structure of security returns is presented by Bollerslev, Engle & Wooldridge [1988], Harvey [1989] and Van Der Meulen [1987, 1989]. Analyzing the covariance structure of deflated returns of general classes of assets (including currencies), Van Der Meulen detects several large shifts in the covariances between these assets over time. As these changes affect the composition of the assets' covariance matrix, the betas of these asset classes (and of the individual assets therein) presumably will not be constant. Theoretical beta non-stationarity resulting from changes in underlying return generating processes (mimicking structural changes in the economy) is explored by Hallerbach [1992].

Aside from these macro-economic considerations, attention was paid to micro-economic determinants of beta. The seminal study linking beta with fundamental firm characteristics (like accounting variables) is Beaver, Kettler & Scholes [1970]. Since then, many studies pursued to uncover the relationship between beta and firm fundamentals.[48] In particular, we mention Rosenberg & Guy [1976a,b]; elaborating on these results, BARRA developed a predictive model for investment risk embracing both beta and residual risk (Rosenberg [1984,1985]; see also chapter one, section 1.4).

By linking beta to the corresponding firm and to the (firm's) economic environment, more insight can be gained into the characteristics of beta and its behavior over time. At the same time, it can be questioned whether beta is the appropriate (or relevant) vehicle to measure and analyze investment risk. More specifically, a more direct link between security returns and changes in the economic environment can be pursued by directly considering the influence of economic variables on investment returns. This is explored in more detail in sections 2.4 and 2.5 and the following chapters. A study of firm characteristics may then shed light on a firm's determinants of its interactions with the economic environment. We touched on this issue in chapter one, section 1.4. Vermeulen [1994] empirically investigates this issue for small and medium-sized business firms.

## 2.3 THE SINGLE INDEX MODEL (SIM) AND THE SINGLE FACTOR MODEL (SFM)

Crucial input for Markowitz' [1952,1959] quadratic mean-variance portfolio problem is the covariance matrix of the individual securities' returns. For the opportunity set of N securities under consideration, the underlying RGP can then be completely general, like eq.(2.20). For this opportunity set, the covariance matrix consists of N variances and ½N(N-1) different covariances. Markowitz [1959, p.96-97] stressed that with an increasing number of securities under consideration, the estimation of the vast quantity of required covariances between the

[48] We refer to Callahan & Mohr [1989] for a recent overview and synthesis.

returns of all securities becomes an unsurmountable problem.[49]

For this reason, models have been developed to simplify the covariance relationships between individual securities. In this section, we will discuss the most simple forms of these covariance models: the 'single index model' and the 'single factor model'. Like the former section, we will distinguish between theoretical and empirical aspects. Extensions of these models are presented in section 2.4.

### 2.3.1 Definition

Following a suggestion of Markowitz [1959, pp.96-101], Sharpe [1963, p.281] simplifies the covariance structure of security returns by specifying a model, in which the returns on the securities in the opportunity set N are linearly related to the return of some index or factor:

$$(2.25) \qquad \underline{r}_{it} = a_i + b_i \underline{r}_{It} + \underline{e}_{it} \qquad \forall\ i \in N$$

with:

$$(2.26) \qquad E(\underline{e}_{it}) = 0$$

$$(2.27) \qquad Cov(\underline{r}_{It}, \underline{e}_{it}) = 0$$

$$(2.28) \qquad Cov(\underline{e}_{it}, \underline{e}_{jt}) = 0 \qquad \forall\ i \neq j$$

where: $\underline{r}_I$ = the random return on index or factor I
$a_i$ = a constant intercept
$b_i$ = the slope or sensitivity coefficient
$\underline{e}_i$ = a random zero-mean disturbance term.

The basic characteristic of this model is the assumption that the securities' returns are interrelated only through a common relationship with a single underlying variable. This relationship is expressed by the attributes $b_i$ of the securities. Sharpe [1963, p.281] suggests that this

[49] This is especially true for analysts' estimates of covariances. Given an opportunity set of 100 securities, 4,950 different covariances must be estimated. Using historical data instead, this implies that more than 100 time series observations on each return are needed, otherwise the covariance matrix will necessarily be singular. The singularity of the covariance matrix is not relevant when the optimization takes place under (binding) restrictions such as short sales constraints, because then no matrix has to be inverted. In that case, mathematical programming techniques are needed to solve the quadratic programming problem. See Perold [1984] and Sharpe [1987] for some techniques.

In an historical perspective, there was also the problem that the quadratic optimization would become very cumbersome; see footnote 51 below.

underlying variable may be "the level of the stock market as a whole, the Gross National Product, some price index or any other factor thought to be the most important single influence on the returns from securities." A broad, 'representative' stock market index, however, is mostly used in this context. Whenever an index, composed of securities, is used, we call (2.25-28) a SIM. The index I is then a proxy for the true common single factor that generates the returns. When instead some exogenous variable is used as factor, we call the system a SFM.[50)]

We note that the SIM cannot represent a causal relationship, since the index return consists of (some combination of) security returns. In principle, the SFM may be given causal content, but we may expect that many more factors may exert their influence on security returns.

In essence, the SIM/SFM is a regression model. The most distinctive characteristic of the SIM/SFM is the restriction (2.28) that the covariance matrix of the disturbances is diagonal. For this reason, the SIM/SFM is also called the **'diagonal model'**. This condition differentiates the SIM from a market index model, as defined in section 2.2.1. As there is no mechanism in a regression procedure to guarantee that (2.28) is satisfied, the diagonal restriction is **assumed** to hold. So, whereas in a market index model the security returns are conditioned on a given market index, the SIM assumes that this conditioning is **exhaustive**. This implies that the choice of the conditioning variable I in a SIM/SFM is directed at achieving the diagonality property.

By means of the SIM (or a SFM), the return on each security i in the opportunity set can be broken down into two components:

- the part of the return that is specific to security i, $a_i + \underline{e}_i$;
- the common part, $b_i \underline{r}_I$, due to movements of the return on index I.

In a regression context, these components are uncorrelated by construction, so eq.(2.27) is always satisfied. As a result, the total return variance of a security can be decomposed into index-related, factor-related or common variance, and residual variance. As the residual returns $\{\underline{e}_i\}_i$ are mutually uncorrelated by assumption, this residual variance consists of only **specific** or **idiosyncratic** variance. (Note the difference with the decomposition in section 2.2.2.D.) Consequently, the 'implied' or "derived" covariance (Markowitz [1959, p.96]) between any security i and j can easily be computed as:

$$(2.29) \qquad \mathrm{Cov}(\underline{r}_i, \underline{r}_j) = b_i b_j \mathrm{Var}(\underline{r}_I)$$

As the index has a sensitivity coefficient of $b_I = 1$, in total only 3N+2 input data are required, in contrast to ½N(N+3) for the full covariance

50) Sharpe [1963; 1970, p.119] does not make this explicit distinction between indices and factors.

procedure.

Aside from the relaxed input requirements, the computational efficiency of the mean-variance portfolio analysis problem also greatly increases. Sharpe [1963] treats the index as a (N+1)-th dummy security and lets its portfolio weight be equal to the weighted average of the securities' sensitivities. As a result, the relevant $(N+1)\times(N+1)$ covariance matrix has non-zero elements only along the diagonal (i.e. N variances of the specific security returns and the variance of the index return). With this structure of the covariance matrix, the quadratic optimization procedure is greatly enhanced.

In order to improve the prospects for practical applications, several efforts were made to translate the portfolio selection problem into a linear programming procedure.[51)] These methods make use of the specific properties of the SIM.

For the case in which relatively stringent constraints are placed on the portfolio weights, Sharpe [1967] developed a linear programming approximation to the original quadratic optimization procedure. The variance of the return on a portfolio p with portfolio weights $\{x_i\}_{i\epsilon p}$ is:

$$(2.30) \quad Var(\underline{r}_p) = b_p^2 Var(\underline{r}_I) + Var(\underline{e}_p)$$

where the specific portfolio variance is:

$$(2.31) \quad Var(\underline{e}_p) = \textstyle\sum_{i\epsilon p} x_i^2 Var(\underline{e}_i).$$

When p is an equally weighted portfolio ($x_i = 1/N$), the specific portfolio variance equals:

$$(2.32) \quad Var(\underline{e}_p) = \textstyle\sum_i 1/N^2 \; Var(\underline{e}_i) = \overline{Var(\underline{e}_i)}/N$$

where $\overline{Var(\underline{e}_i)}$ denotes the mean value of the securities' specific variances. Given the assumed properties of the specific return components, the law of large numbers can be invoked according to which the specific portfolio variance tends towards zero when more and more

51) In an historical perspective, there was the problem that the quadratic optimization would become very cumbersome; computational problems thus restricted considering a portfolio of a practical size. As these difficulties are nowadays largely overcome, computational advantages form alone no reason to apply the linear approximation procedures. In effect, as Rudd & Rosenberg [1979, p.23] remark: "In most instances, the rigorous quadratic programming solution may actually be less expensive than the approximated linear programming approach, for the reason that the LP approximation entails the addition of otherwise unnecessary activities and constraints, which increase the LP far beyond the minimal size necessary in the exact formulation."

securities are included:

$$(2.33) \quad \lim_{N\to\infty} \mathrm{Var}(\underline{e}_p) = \lim_{N\to\infty} \overline{\mathrm{Var}(\underline{e}_i)}/N = 0$$

As a result of this naive diversification process, the residual variance can be ignored for large and approximately equally weighted portfolios. With rather restrictive bounds on the portfolio weights (e.g. $x_i \le .05$), the variance of a portfolio's return can then be adequately approximated by:

$$(2.34) \quad \mathrm{Var}(\underline{r}_p) \sim b_p{}^2\mathrm{Var}(\underline{r}_I)$$

This implies that in a well-diversified portfolio, the absolute values of the securities' sensitivity coefficients, $|b_i|$, can serve as linear (but approximate) surrogates for the non-linear standard deviation.[52)]

Sharpe [1971b] provides yet another linear approximation to the quadratic portfolio optimization problem (which does not critically depend on bounds on the portfolio weights). He first performs a transformation on the portfolio weights that converts the general expression for portfolio variance into a separable quadratic function of the portfolio weights.[53)] He then applies a piecewise linear approximation to each of the terms in this diagonalized expression. This approach does not depend on the use of the SIM, but when the SIM is used, the diagonalizing transformation is no longer necesssary (only the piecewise linear approximation remains).

Finally we note that, under the assumption that the SIM adequately describes the covariance structure of the stock returns in the opportunity set, Elton, Gruber & Padberg [1976,1977b,1978] have developed 'simple rules' to determine optimal portfolios. Bawa, Elton &

---

52) Stone [1973, p.626] presents another linear re-formulation of the portfolio selection problem. He elegantly approximates the specific portfolio variance by $\sum_{i\in p} x_i x_{i,max}\mathrm{Var}(\underline{e}_i)$, where $x_{i,max}$ is the upper bound on the portfolio weight of security i in the portfolio p. However, he incorporates this specific **variance** into the linear objective function, together with the index-related **standard deviation** $b_p \cdot \sigma_I$. This is inconsistent.

53) We can trace the transformation of a real quadratic form to a sum of squares to the spectral decomposition theorem in matrix theory (Graybill [1983, pp.47-48], Mardia, Kent & Bibby [1979, p.469]). Given the (symmetric) covariance matrix $\mathbf{\Sigma}$ of security returns, there exists a (real-valued) orthogonal matrix $\mathbf{P}$ such that $\mathbf{P'\Sigma P} = \mathbf{D}$, where the diagonal matrix $\mathbf{D}$ contains the (real) eigenvalues $d_i$ of $\mathbf{\Sigma}$, and where $\mathbf{P}$ contains the eigenvectors. Transforming a portfolio weight vector $\mathbf{v} = [v_i]_i$ to $\mathbf{w} = [w_i]_i \equiv \mathbf{P'v}$ yields the expression $\mathbf{v'\Sigma v} = \mathbf{w'Dw} = \sum_i w_i{}^2 d_i$ for the portfolio variance.
Since the diagonal elements of an upper triangular matrix are its eigenvalues, we only have to convert the covariance martrix into upper triangular form.

Gruber [1979] furthermore showed that all the results carry over to the case where the SIM holds in a stable Paretian context.

### 2.3.2 The SIM and SFM as theoretical concepts

**An internal consistency problem**

As remarked before, the diagonality condition (2.28) differentiates the SIM from the market index model. Indeed, in a market index model a given market index is used, whereas for the SIM the choice of the conditioning variable $\underline{r}_I$ is directed at achieving the diagonality property. The question arises whether the SIM can be considered as a special case of the MM or of a market index model.

In the context of the MM (2.2), a diagonality restriction on the disturbance terms $\{\underline{\varepsilon}_i\}_{i\in M}$ would imply an internal inconsistency in the model.[54] After all, as the (weighted) average of the securities' disturbances in M is always zero (eq. (2.8)), there must exist (linear) dependencies between them, violating diagonality.[55] The same argument applies for a market index model, where the index is a (weighted) average of the N securities, for which the SIM is asssumed to hold. In this case, the SIM could in principle only hold for N-1 securities. So the SIM cannot be considered as a special case of the MM or the market index model with an **internal** index.

For the reason of the internal inconsistency that arises when an internal index is used, a SFM can be postulated in which the conditioning variable is no portfolio but a common underlying 'market factor' (reflecting 'business conditions'), or a more or less abstract random variable.[56] Beja [1972, p.39], however, has shown that even in a SFM the diagonality restriction is very restrictive with respect to the possibility of **any** random variable to serve as a candidate for factor.[57]

---

[54] This point was first raised by Fama [1968, p.39]. See also Fama [1973, p.1184], Stapleton [1980, p.16] and section 2.2.1.

[55] Another way to see this is to consider an index I with composition $\{x_i\}_{i\in I}$. The variance of the index return equals $\text{Var}(\underline{r}_I) = \text{Var}(\sum x_i \underline{r}_i) = b_I \text{Var}(\underline{r}_I) + \text{Var}(\underline{e}_I)$. As $b_I=1$ and $\text{Var}(\underline{e}_I)>0$, this cannot hold.

[56] Cf. Lintner [1965a, pp.23-24 fn 32], Lintner [1965b, p.601], Fama [1968, p.39] and Fama & Miller [1972, pp.313-319].

[57] Consider the securities i, j and k $\in$ N. Whatever the signs of their sensitivity coefficients $b_i$, $b_j$ and $b_k$ may be, the products $b_i b_j$, $b_i b_k$ and $b_j b_k$ are either all positive or one positive and two negative. From the expression for derived covariances eq. (2.29), it then follows that a necessary (but not sufficient) condition for a SFM to hold is that for **any** triplet of securities within N, **either** all three pairs within the triplet are positively correlated **or** two of the pairs are negatively correlated and the third pair is positively correlated. This severe restriction on the covariance structure of the securities' returns under a SFM is not widely

In the context of a SIM, no internal inconsistency is implied when an **external** index is employed. In an extreme case, the intersection between the set of securities of which the index is composed and the opportunity set of N securities for which the SIM is specified, is empty. A more realistic case of an external index is when the index used contains contains more than N securities. So, in principle, the SIM could hold for an opportunity set of securities when this opportunity set is a subset of the set of securities from which the index is composed. This more general index can then be regarded as a proxy for the (not necessarily observable) true common single factor that generates the returns. This index, however, can never measure this factor exactly, no matter how many securities are used in the index. After all, the index will always contain a return component that is specific with respect to the factor. Hence, the index will always contain specific variance.[58)]

### The quest for a 'best' index

Disregarding the discussed consistency problem, one could try to construct a 'best' index for the SIM, using the securities from the opportunity set. As an index is a portfolio, it should also be a linear function of the securities' returns. Gaviano, Sutti & Szegö [1972] and Szegö [1980, pp.203-206] show how a 'best linear index' can be constructed that is best in the sense that it minimizes the (unweighted) sum of the residual variances of all securities in the opportunity set. This index maximizes the weighted average of the securities' squared correlation coefficients with the index, where the total return variances of the securities serve as weights. It follows that the composition of the best index is given by the eigenvector that is associated with the largest eigenvalue of the covariance matrix of the security returns. Gaviano, Sutti & Szegö's [1972] and Szegö's [1980] analyses are quite laborious, but the problem can readily be cast in terms of principal components analysis. According to the spectral decomposition theorem in matrix theory, the covariance matrix $\mathbf{\Sigma}$ of the N security returns can be decomposed as $\mathbf{\Sigma} = \mathbf{PDP}'$, where the diagonal matrix $\mathbf{D}$ contains the N eigenvalues $d_i$ of $\mathbf{\Sigma}$, and where $\mathbf{P}$ contains the N eigenvectors.[59)] As $\mathbf{\Sigma}$ is symmetric, the eigenvalues and the elements of the eigenvectors are real and $\mathbf{P}$ is orthogonal. As the best linear index is the (normalized) linear combination of the security returns

---

recognized.

58) Cf. Markowitz [1984, pp.13-14]. Roll [1978, pp.1066-1067] also discusses the use of an index as proxy for a true return generating factor. As a result of errors in variables, the intercept and slope of the SFM (2.25) will be mis-estimated.

59) As $\mathbf{P}$ is not unique, its colums (the eigenvectors) can be standardized, yielding an orthogonal matrix. See footnote 53 above and Johnston [1984, pp.536ff].

with maximum variance, it is their first principal component. Its composition is given by the (normalized) eigenvector associated with the largest eigenvalue.[60)]

**Proxying the underlying factor**

Note that a 'best' linear index implies internal inconsistency in the SIM. Instead of using (variance oriented) principal components analysis, one could apply (covariance oriented) factor analysis.[61)] The latter technique explicitly assumes a linear underlying process like the SFM. This implies that the covariance matrix $\mathbf{\Sigma}$ of security returns can be decomposed as $\mathbf{\Sigma} = \mathbf{bb}' + \mathbf{\Psi}$, where the diagonal N×N matrix $\mathbf{\Psi}$ contains the specific variances and where $\mathbf{b}$ is the colum vector of N sensitivity coefficients. To cope with the problem of degrees of freedom, the factor is standardized (zero-mean and unit standard deviation). The spectral decomposition is now applied to the covariance matrix, corrected for the specific variances (that the model cannot explain), yielding $\mathbf{\Sigma} - \mathbf{\Psi} = \mathbf{bb}'$. In this case, there is only one non-zero eigenvalue, which illustrates again the restrictive nature of the SFM.[62)] The implied factor is now only an approximate linear combination of the securities in the opportunity set. As the factor is not an internal index, the internal inconsistency is avoided.

In practice, a more general linear underlying k-factor model is assumed, where (considering the degrees of freedom problem) the factors are not only standardized, but also assumed to be uncorrelated (orthogonal). This implies that the covariance matrix $\mathbf{\Sigma}$ of security returns can be decomposed as $\mathbf{\Sigma} = \mathbf{BB}' + \mathbf{\Psi}$, where $\mathbf{B}$ is now a N×k matrix of sensitivity coefficients, associated with the k factors. Losely stated, the procedure is first to estimate the specific variances, summarized in $\mathbf{\Psi}$. A spectral decomposition of the corrected covariance matrix now yields $\mathbf{\Sigma} - \mathbf{\Psi} = \mathbf{BB}'$. The first factor (associated with the largest eigenvalue) is now the 'best' factor in the sense that it accounts for as much as possible of the covariance between the security returns.

---

60) In addition, we remark that when all elements of the covariance matrix are positive, the elements of the eigenvector corresponding to the largest eigenvalue will all have the same sign (see Theil [1960, p.478]). This implies that the best index then contains no short positions.

61) For expositions of principal components analysis and factor analysis, as well as the differences between these techniques, we refer to Mardia, Kent & Bibby [1979, Chs.8&9].

62) For symmetric matrices, rank equals the number of non-zero eigenvalues (Graybill [1983, p.16]). In this case, there is only one non-zero eigenvalue (with associated eigenvector $\mathbf{b}$), indicating that the corrected covariance matrix is of rank one.

**A link with asset pricing theory**

As discussed in section 2.2.2.E, the link between the MM and the CAPM is more of an indirect-intuitive than of a direct-technical character. The SFM, in contrast, is directly related to another asset pricing theory. More specific, Ross [1976] first derived the Arbitrage Pricing Theory (APT) starting from the assumption that security returns are generated by a one-factor model. As the APT can be derived from a more general multi-factor model, we refer to section 2.4.

**Simplifying portfolio analysis**

As mentioned before, the use of the SIM facilitates the collection of input data required for the mean-variance portfolio selection problem. A nice example is Hodges & Brealey [1970, 1972], who monitored institutional investors who worked with the SIM: some of the necessary inputs were estimated statistically, others were provided subjectively by the participants.

Aside from this, the SIM is important for simplifying the structure of interactions between security returns beyond the $(E,\sigma^2)$-context (as discussed before). In particular, this applies to the case of stable (Paretian) distributions with characteristic exponent $\theta<2$, where variance is replaced by a scale parameter and where the familiar covariance is no well-defined concept. Because the SIM specifies the interrelationships between the security returns in terms of sensitivities for the index return, this model (or its multivariate extensions) provide "a natural framework for portfolio analysis" (Fama [1965b, p.408]).

In various applications, the SIM is imposed on the data. However, it must be realized that the MM is implied for a wide range of distributions (cf. section 2.2.2). The diagonality property (the distinctive characteristic of the SIM over the MM) cannot be imposed on the data, since it is anyway **assumed** to hold. This then implies that the linearity of the SIM is much less a problem than its diagonality assumption. Because of the implied linearity between security and market returns in a symmetric stable (Paretian) market, investigations into the linearity of the SIM in such a market (like performed by Blume [1970, pp.158ff]) are of a tautological nature.

The appropriateness of the diagonality assumption is briefly discussed in the next section.

### 2.3.3 The SIM and SFM as empirical concepts

In the empirical context of the SIM/SFM, we first comment on the use of indices. We then consider the empirical validity of the diagonality assumption.

### Indices and factors

In empirical work on the SIM, several indices are used. Existing stock market indices, extending beyond the opportunity set of securities (common stocks) under consideration, were mostly used. Some studies, in contrast, use all securities in their opportunity set to construct an index. Sharpe [1967, p.503], Wallingford [1967, p.103] and Frankfurter [1976, pp.951-952], for example, compose an index by averaging the securities in their opportunity sets, whereas Elton & Gruber [1973] use the first principal component from their opportunity set as index. This is not a very lucky choice, since the use of these internal indices gives rise to the internal inconsistency problem. By using the principal factor —extracted from the data by means of factor analysis— as index, King [1966] escapes this problem.

In the context of the SFM, Smith [1969, p.329] uses five economic indices or factors: Gross National Product, Consumer Price Index, Federal Reserve Board Index of Industrial Production, Business Week Index (also measuring industrial production) and Money Supply.

### The diagonality assumption and index / factor choice

It is now interesting to consider the question how well a SIM or SFM is capable of explaining the covariances between the securities in the opportunity set. In other words: how 'diagonal' is the SIM or SFM when some single index or underlying factor is used?

This question can be considered in a direct and an indirect way. In the **direct** approach, the differences between the 'true' and the implied covariance (or correlation) matrix are analyzed in a more or less statistical way. In the **indirect** approach, the $(E,\sigma^2)$-efficient frontiers generated by the SIM/SFM and the full covariance model are compared and the differences are evaluated from an economic point of view, for example in terms of a higher or lower expected return for a given level of variance or standard deviation.[63]

In addition, the question can be posed either in an ***ex post*** context (how well are historic covariances explained) or in an ***ex ante*** context (how do the historic true and implied covariance matrices perform in forecasting the future covariance structure?).[64] In applications of the model, the performance in an *ex ante* context is of

---

63) Elton & Gruber [1973, p.1211] use the terms 'statistical significance' and 'economic significance' to distinguish both approaches.

64) Somewhat confusingly, some authors (Elton & Gruber [1973] and Cohen & Pogue [1967], for example) use the term *'ex ante'* for historical analysis and *'ex post'* for forecasting evaluation.

paramount importance.

We now critically summarize some empirical evidence on the SIM/SFM.

Sharpe [1963, pp.291-292; 1967, pp.503-505] found that (given a level of expected return) the composition of efficient portfolios from the SIM and the full covariance method were very similar. Wallingford [1967, p.102], in contrast, found that the full covariance efficient frontier dominated the SIM efficient frontier. These indirect and *ex post* studies are only indicative (the more because of the very small sample size of only 20 securities) and the results are inconclusive with respect to the performance of the SIM.

The first extensive study in this field is by Smith [1969]. Using an opportunity set of 100 stocks and annual returns over a period of ten years, he computed (co-) variances and efficient sets. He found that the mean-square errors between the 4,950 derived and true covariances were almost the same for the three stock market indices he used. In addition, the three SIM efficient frontiers were virtually coincident and only slightly dominated by the full covariance efficient frontier. These findings are consistent with the results presented by Cohen & Pogue [1967, p.179] and Alexander [1978]. Note, however, that Smith [1969] used only 10 observations to estimate his full covariance matrices. As a consequence, these matrices can have a rank of at most 10, resulting in a serious bias in favor of the SIM.

Using the five economic factors (as indicated above) on the other hand, Smith [1969] found that the stocks' average coefficient of determination was about 20%, compared against about 50% when using stock market indices. This limited their ability to replicate the full covariance matrix by the derived covariances. Interesting is that the SFM efficient frontiers dominated the full covariance efficient frontier. This, of course, does not indicate attractive investment opportunities, but implies that the SFMs were incorrectly specified. As the economic factors only explained a small portion of the return variability, they did do a very poor job in accounting for the covariance between the securities. Hence, the residual covariance matrix was not diagonal at all, but the SFM assumed away a large part of the true covariability between securities. The 'dominant efficient frontiers' thus resulted from **spurious diversification** and hence were not attainable.

This phenomenon is not limited to SFMs but will occur whenever the diagonality assumption is seriously violated and strong residual covariance effects exist. For the SIM, the inverse relationship between the degree of portfolio diversification and the (average) correlation between index and security returns was hypothesized by Frankfurter [1976, p.953] and is explicitly shown by Frankfurter & Frecka [1979].

The latter study is especially interesting, since Frankfurter & Frecka [1979] constructed simulated indices which were *a priori* unrelated to the returns on the securities in the opportunity set.

Although the resulting portfolios show superfluous diversification, Frankfurter & Frecka [1979, pp.933-934] found that the *ex ante* (risk adjusted) performance of efficient portfolios, generated by real as well as simulated indices, exhibited no significant differences. Starting from the premise that the relative quality of an index in the SIM must be judged in terms of *ex ante* performance, this lead them to the conclusion that "practically any number can be used as a proxy for the Sharpe common factor..." (p.937). Considering the specific sample and time period studied by Frankfurter & Frecka [1979], we are very reluctant to generalize their conclusion.

Elton, Gruber & Urich [1978], finally, examined the ability of SIMs to forecast the future correlation structure between security returns. Their analysis can thus be classified as direct and *ex ante*. They compared the forecasting ability of the historic full correlation matrix, several SIMs and the constant correlation model. In the constant correlation model, every forecasted correlation coefficient equals the overall cross-section average of historic coefficients; this naive model serves as a benchmark. The SIMs are used to generate derived *ex ante* correlation matrices, and differ in the way the sensitivities $b_i$ are estimated. Actually, the SIMs are estimated as market index models, except for a naive SIM, in which all $b_i$'s are assumed to equal one. The forecast accuracy of one model relative to another model was judged in terms of the mean difference in absolute forecast errors as well as in (the more stringent) terms of dominance of the cumulative frequency distribution of the absolute forecast errors. The striking result was that the SIMs estimated against the index outperformed the historical correlation matrix at a statistically significant level (5%). This implies that the historical full correlation matrix contains a lot of random noise.[65] They also outperformed the naive '$b_i$=1' SIM. This in turn implies that differences between sensitivity coefficients conveyed real information about the future correlation structure.[66]

One insight these studies offer is that the strength of the relationship between the selected index and the securities in the opportunity set is crucial. Between the extremes of the SIM or SFM and the 'full covariance model' (where each security has its own index), there is continuum of

[65] This is relevant in the context of estimation risk. As shown by Frankfurter, Phillips & Seagle [1971], errors in estimating means and (co-) variances cause erratic performance of the full covariance model, especially when the sample size is small. In a comparative simulation study, Frankfurter, Phillips & Seagle [1976] show that the SIM is less subject to this erratic behavior when relevant historical data are limited.

[66] Quite discomforting, however, is that the overall mean correlation model outperformed all other models. As the overall mean models and the three SIMs can be viewed as smoothing techniques, the results indicate that smoothing across correlation coefficients eliminates random noise and improves forecasting ability.

covariance models. We will explore these models in the next section.

## 2.4 MULTI-INDEX MODELS (MIM) AND MULTI-FACTOR MODELS (MFM)

### 2.4.1 Definition

A multi-index model (MIM) is a straightforward extension of a SIM and can be specified as:

$$(2.35) \qquad \underline{r}_{it} = a_i + \sum_{j=1}^{k} b_{ij}\underline{r}_{Ijt} + \underline{e}_{it} \qquad \forall\ i \in N$$

with:

$$(2.36) \qquad E(\underline{e}_{it}) = 0$$

$$(2.37) \qquad Cov(\underline{r}_{Ijt}, \underline{e}_{it}) = 0 \qquad \forall\ j$$

$$(2.38) \qquad Cov(\underline{r}_{Ijt}, \underline{r}_{Ist}) = 0 \qquad \forall\ j \neq s$$

$$(2.39) \qquad Cov(\underline{e}_{ht}, \underline{e}_{it}) = 0 \qquad \forall\ h \neq i$$

where: $\underline{r}_{Ij}$ = the random return on index or factor $I_j$
$a_i$ = a constant intercept
$b_{ij}$ = the slope or sensitivity coefficient for index or factor j
$\underline{e}_i$ = a random zero-mean disturbance term.

Like its univariate counterpart, this model assumes that the securities' returns are interrelated only through a common relationship with the k underlying variables. Whenever indices, composed of securities, are used, we call (2.35-39) a MIM. The indices are then proxies for the true common factors that generate the returns. When instead exogenous variables are used as factors (like $\{\underline{\delta}_j\}_{j\in k}$ in chapter three), we call the system a MFM.

The assumption (2.38) that the indices or factors are uncorrelated is not restrictive at all. Given any set of correlated indices or factors, it is easy to transform them into uncorrelated ones.[67] This allows for using univariate calculations in a multivariate context.

Like the SIM, the distinctive feature of a MIM is the diagonality of the residual covariance matrix. Hence, the conditioning on the indices is assumed to be exhaustive. Again, the return on each security i can be broken down into two components:

[67] See e.g. Sharpe [1970, pp.125-127] or Elton & Gruber [1991, pp.148-149]

- the part of the return that is specific to security i, $a_i + \underline{e}_i$, and
- the common part, $\sum_j b_{ij}\underline{r}_{Ij}$, due to movements of the returns on the k indices.

The latter part of the return generates the index-related, factor-related or common variance. The former part generates the residual variance, which consists by assumption (2.39) of only specific or idiosyncratic variance. Consequently, the 'implied' or 'derived' covariance between any security h and i can now be computed as:

$$(2.40) \qquad Cov(\underline{r}_h, \underline{r}_i) = \sum_{j=1}^{k} b_{hj} b_{ij} Var(\underline{r}_{Ij}) \qquad \forall\ h,i \in N$$

In the mean-variance portfolio selection problem, in total (k+2)N+2k input data are required, in contrast to 3N+2 for the SIM or SFM. Still, when the number of indices or factors is not to large, the input requirements are less than for the full covariance model. On the computational side, Pang [1980], Kwan [1984] and Markowitz & Perold [1981], for example, show how to increase the efficiency of the quadratic optimization problem when the covariance matrix has a MIM or MFM structure.[68)]

For approximating the full covariance matrix of security returns, the diagonality assumption of the MIM/MFM is crucial. When the conditioning is exhaustive, we term the models **'strict'** or 'exact' MIMs/MFMs. For other purposes, however, like generating conditional return predictions, the conditioning need not be exhaustive. In this alternative case, security returns are conditioned on some indices or factors without assuming (or worrying about) diagonality.

### 2.4.2 The MIM and MFM as theoretical concepts

As MIMs and MFMs are natural extensions of SIMs and SFMs, we can be short on this point. Two issues, however, deserve special attention: the number of indices/factors, and the candidates for these conditioning variables.

---

[68)] As for the SIM, Elton, Gruber & Padberg [1977a;1979a,b] have developed simple rules to determine optimal portfolios under the assumption that the covariance structure of the stock returns in the opportunity set adequately can be described by Cohen & Pogue's [1967] diagonal form of MIMs. These results carry over to the case where MIMs hold in a stable (Paretian) context (Bawa, Elton & Gruber [1979, p.1042]).

#### 2.4.2.A The number of indices or factors

Obviously, more and more indices can be chosen until the residual covariance becomes neglegibly small. Indeed, by performing principal components analysis on the (full rank) covariance or correlation matrix of the returns on the N securities in the opportunity set, N internal orthogonal indices are obtained that account exactly for all return covariability. Allas, the resulting model does in no way reduce the dimensionality of the return generating and is of no practical use. As in practice a subset of the first principal components generally accounts for most of the variability in the data, a smaller set of indices could be used, eliminating virtually all residual covariability. The same applies to factor analytical techniques, which are a more obvious choice in this context.

Notwithstanding the actual existence of a k-dimensional return generating process, some approximate covariance structure could be imposed on the data. In an historical (or *ex post*) context, this is true. However, for practical applications of the models the *ex ante* performance is crucial. The relevant question then is whether the inclusion of additional indices or factors really adds information to the model, or merely noise. Important is the degree in which the estimated MIM/MFM (that is imposed on the data) corresponds to the underlying structure (that actually generates the data).

#### 2.4.2.B The nature of the indices or factors

With respect to the nature or identity of the variables, incorporated in a MIM or MFM, we make a distinction between general MIMs, extra-market covariance models, general MFMs and residual-market-index factor models. In this section, we critically discuss these models.

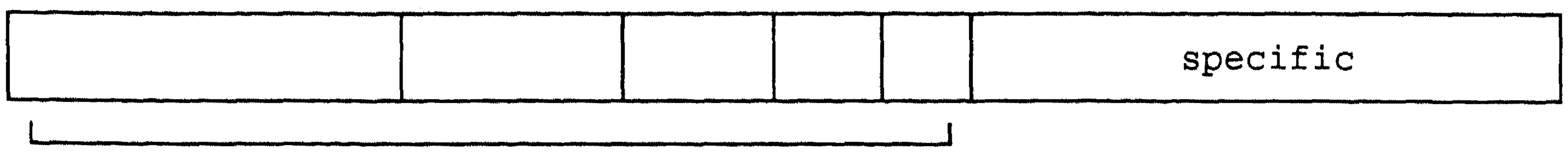
The motivation to incorporate a typology is that ours is different and beyond that of presentations in the literature.[69] Furthermore, a clear typification enables a consistent terminology. Finally, it enables us to show the relationships between the models.

For clarity, we provide schematic representations of the models, depicting return components (or the proportion of variance explained by the components).

---

[69] Elton & Gruber [1991, Ch.6], for example, distinguish between general MIMs (our eq.(2.35)), industry IMs (our extra-market covariance models), mixed models (in which extra-market covariance is predicted using variables as traditional industry classification and fundamental firm variables), and fundamental MIMs (our general MFMs and residual-market-index FMs).

**General MIMs**




- internal indices: · statistically extracted
  · (industry) grouping
- external indices: · prespecified
  · (industry) grouping

General MIMs have the form of eq.(2.35). This class of MIMs can be subdivided by two criteria: whether internal or external indices are used, and whether or not *a priori* information on the identity of the indices is used. These sub-classifications partly overlap. External indices are prespecified indices, like sector indices or industry indices. These indices can already exist or can be constructed using securities from a universe beyond the opportunity set under study (as do Cohen & Pogue [1967], e.g.). Internal indices are either extracted statistically from the data so that no *a priori* information is used (by means of principal components analysis; Elton & Gruber [1973], e.g.), or formed by combining the securities in the opportunity set into group indices (using information on industry affiliation, for example; King [1966]).

The two types of MIMs, developed by Cohen & Pogue [1967], are widely quoted. In their MIM in covariance form, all N securities within the opportunity set are placed in K classes. Within each of these classes, a SIM holds that relates the return on each security to an index in the class to which it belongs. Considering an opportunity set of strictly common stocks, the class indices can be thought of as industry indices; when considering several asset categories, an appropriate index can be defined for each category. The K indices in turn are related to each other by a covariance matrix. This model can readily be transposed to the form of eq.(2.35) by allowing the indices to be correlated and by restricting each security to have only one non-zero index sensitivity (namely for the index of the class to which it belongs).

Their MIM in diagonal form is a hierarchical model in that it incorporates the additional assumption that each class index itself is linearly related to an overall index in a SIM-like fashion.[70] This assumption on the relationships between the class indices diagonalizes their covariance matrix. By substituting the SIM on the index level into the SIM on the security level, this model can also easily be shown to be

[70] In contrast to the exposition of Alexander & Francis [1986, p.84], Cohen & Pogue [1967] do not rely on bivariate normality of the returns on the securities and the class indices in order to specify the linear relationships between their returns. They simply assume linearity.

similar to eq.(2.35).[71] The first index is then the overall index and the subsequent K indices are the transformed class indices, constrained to be uncorrelated with the overall index. The sensitivities of each security for these transformed class indices are (again) assumed to be zero, except for the index of the class to which the security belongs. As for any security the sensitivity to the overall index is constrained to be equal to the product of *(i)* its sensitivity to the class index and *(ii)* this class index' sensitivity to the overall index, the MIM in diagonal form remains in principle a K-index model.[72]

For an exposition of the way in which the resulting formulas for a portfolio's return variance facilitate the mean-variance portfolio selection model, we refer to Cohen & Pogue [1967, pp.189-193]. Judged by the stringency of the underlying assumptions, the MIM in covariance form appears to be more general than the MIM in diagonal form. We note, however, that it may be very difficult (if not impossible) to find a set of class indices that are mutually correlated, whereas at the same time any security is correlated with one and only one of these indices. This, together with the fact that correlated indices can easily be made orthogonal, speaks in favor of the MIM in diagonal form. Considering the power of the simple SIM described in section 2.3, it is then straightforward to adapt the MIM in diagonal form and to consider MIMs in which securities are related to both an overall index and a set of sub-indices. When the constraint on the sensitivity for the overall index is dropped, we arrive at the next class of MIMs.

**Extra-market covariance models**

| | residual | | | | |
|---|---|---|---|---|---|
| index/factor | | | | | specific |

| | | |
|---|---|---|
| - market index | - industries | : internal/external indices |
| · internal | - sectors/groupings | : internal indices |
| · external | - pseudo-industries | or factors |
| - market factor | - factor(s) | : external |

As discussed in section 2.3.3.B, the conditioning of security returns on an existing stock market index (or market factor) does not generally yield a diagonal residual covariance matrix. As 'the market' cannot account for all of the covariability between security returns, there exist "extra-market components of covariance" (Rosenberg [1974, p.263]). In this view, the residual variance of security returns (i.e. after conditioning on 'the' market return) can be decomposed into extra-market covariance and specific variance. Next, the extra-market covariance can

[71] Cf. Alexander & Francis [1986, p.84].

[72] This in contrast to Elton & Gruber [1991, p.137] who designate the model as a general K+1-index model.

be attributed to indices or to factor influences. The extra-market covariance models can thus be regarded as 'extended market index models'.

In extra-market covariance models of the **hierarchical type**, the security returns are first conditioned on a market index (by performing a regression against the index) or on a market factor (extracted from the sample using factor analysis or principal components analysis). Next, the residuals from this step are related to indices or factors. In this hierarchical procedure, orthogonality is imposed on the market-related part of the return and the extra-market return component. The same result is obtained when the additional (i.e. extra-market) indices or factors are orthogonal(-ized) to the market return and then the security returns are related to the market and these indices in one step. In extra-market covariance models of the **general type**, the security returns are related to the market and the additional indices in one step without imposing this orthogonality.

Constraining the first index in eq.(2.35) to be a market index, the extra-market covariance models take the form:

$$(2.41) \qquad \underline{r}_{it} = a_i + b_{im}\underline{r}_{mt} + \sum_{j=2}^{k} b_{ij}\underline{r}_{Ijt} + \underline{e}_{it} \qquad \forall\ i \in N$$

where $\underline{e}_{it}$ is again a random zero-mean disturbance term, satisfying the diagonality assumption, and orthogonal to $\underline{r}_{mt}$ and the indices $\{\underline{r}_{Ij}\}_j$. These extra-market indices may or may not be mutually correlated. For extra-market covariance models of the hierarchical type, we furthermore have:

$$(2.42) \qquad \mathrm{Cov}(\underline{r}_{mt}, \underline{r}_{Ijt}) = 0 \qquad j = 2, \ldots, k$$

Note that when the market index is an internal index, internal inconsistency is implied (cf. section 2.3.2). When using in addition internal extra-market indices, eq.(2.42) cannot hold.

Aside from the classification hierarchical and general, extra-market covariance models can be distinguished by the nature of the extra-market indices or factors. Extra-market covariance can be attributed to industry influences by either using external indices based on traditional industry groupings or using internal indices.[73] Alternatively, securities can be grouped into pseudo-industries according to their actual return behavior.[74] These pseudo-industries

[73] I.e., forming stock groupings according to traditional industries. Examples are King [1966], Meyers [1973b], Fertuck [1975] and Livingston [1977].

[74] King [1966] uses factor analysis and cluster analysis, Meyers [1973b] and Elton & Gruber [1973] use principal components analysis, and Fertuck [1975] uses cluster analysis.

can partly overlap standard industry classifications.

According to their actual return behavior, securities can also be classified into categories that are broader than pseudo-industries. The favorite technique to extract a relatively small number of comovement groups from the security data is cluster analysis. The resulting comovement groups are called 'homogeneous groups' (Farrell [1974,1975] and Martin & Klemkosky [1976]), 'sector indices' (Farrell [1983]) or 'comovement classes' (Arnott [1980]). In addition to internal indices, indices provided by investment services can be used.[75]

Yet another alternative to account for extra-market covariance is to consider macro-economic variables (factors). In this context, considerable attention is paid in the literature to the influence of interest rates on stock returns. Interest rate effects can be incorporated into the model in an indirect way by supplementing the stock market index with one or several bond market indices. A bond market index thus serves as a proxy for one or more combined interest rate factors; the correlation between interest rate changes and the returns on the bond index will be negative. An early example is Bildersee [1973] who relates (common and preferred) stock returns to both a stock market index and a bond market index. This type of return generating process was formalized by Stone [1974, p.717], who used the framework for quantifying the concept of interest rate risk and for deriving a "generalization of the standard asset pricing equation".[76] Flannery & James [1984] use a similar index model.

In the direct approach of incorporating interest rate effects, one or more variables that measure the change in (some) interest rates are added to the stock market index. In this category we have Sweeney & Warga [1986]. Of course, the direct and indirect approaches can be combined into one model, by considering both changes in for example a (short term) interest rate and the return on a (long term) bond index. A specific example is Lynge & Zumwalt [1980], who supplement a stock market index with both a long term and a short term bond index. As the

---

[75] As Aber [1976, p.620] notes, investors and portfolio managers will not restrict themselves to consider stock price behavior in terms of traditional industry classifications. They rather may (more or less arbitrarily) classify stocks into other groupings, use indices provided by investment services, or consider macroeconomic variables.

[76] Korkie [1974] extended Stone's return generating process to a 3-IM where the third index indirectly captures differential real effects of inflation on firm performance. Gultekin & Rogalski [1979, pp.630-631] in turn criticize Stone's asset pricing model. While acknowledging that Stone's derivation is technically correct, they show that he misinterpreted the pricing equation. Consequently, Stone's pricing model turns out to be fully consistent with the standard (Sharpe [1964], Lintner [1965a,b] or Black [1972]) asset pricing model. Gultekin & Rogalski furthermore argue that the term 'interest rate risk' is inconsistent with the one-period context of the model.

remaining maturity of the short term bond index is approximately equal to their observation interval, its holding period return approximately equals the short term interest rate. So, instead of a change in an interest rate, the interest rate itself appears in their model.

Considering the various forms in which an extra-market model can manifest itself, we can specify two extremes. At one end, we have the 'weakest' form, incorporating a market factor supplemented with pseudo-industries/sectors; these models can hardly be distinguished from general MIMs. The 'strongest' form incorporates a market index plus hypothesized factor(s).

The inclusion of a general market index m in a MIM, as indicated by the studies discussed above, may not be interpreted as an indication that $\underline{r}_m$ causes $\underline{r}_i$. As a market index is a convex combination of the incorporated securities, its return is influenced by events or factor movements, just like the individual securities. The index then just represents a certain packaging of factors. Appropriately weighting the incorporated securities with the index weights $\{m_i\}_{i \in m}$ and aggregating yields the MFM (2.35) for the index:

$$(2.43) \qquad \underline{r}_{mt} = a_m + \textstyle\sum_j b_{mj} \underline{\delta}_{jt} + \underline{e}_{mt}$$

where $\{\underline{\delta}_j\}_{j=1,\ldots,k}$ now denote the factors. We see that the market index is a proxy for a specific linear combination of factors that influence stock returns. Incorporating this expression in the MM eq.(2.2) —an '*ex post* factor model'— yields for security i:

$$(2.44) \qquad \underline{r}_{it} = [\alpha_i + \beta_i a_m] + \textstyle\sum_j [\beta_i b_{mj}] \cdot \underline{\delta}_{jt} + [\underline{\varepsilon}_{it} + \beta_i \underline{e}_{mt}]$$

Apparently, the finding of factors in excess of the 'market factor' m indicates that the factor sensitivities of the individual securities differ from the average (i.e. market) sensitivities in the sense that the coefficients $b_{ij}$ in (2.35) differ from the restricted coefficients $\beta_i b_{mj}$ in (2.44). This argument includes the possibility that some factors are 'nihilated' in the index m. In that case, individual securities exhibit non-zero factor sensitivities $b_{ij}$, but the weighted average is zero: $\sum_{i \in m} m_i b_{ij} = 0$ for one or more factors j.[77]

This discussion already brought us to the field of another class of models, to be considered next.

---

[77] By the same token, isolating a 'comovement class' of securities can be interpreted as finding securities that each are dependent of a factor that generates (a part of) their return.

**General MFMs**

| | | | | idiosyncratic |
|---|---|---|---|---|

- statistically extracted factors
- explicit (external) factors
- mimicking portfolios

In general MFMs, a close relationship with the underlying stochastic process generating the security returns is pursued. There is a growing interest in these models, stimulated by the development of Ross' [1976, 1977b] Arbitrage Pricing Theory (APT) that is based on a general multi-factor return generating process (in strict form). We can specify a (partly overlapping) sub-classification of general MFMs using two criteria: whether or not *a priori* information on the identity of the factors is used, and whether the factors are incorporated directly or indirectly by forming mimicking portfolios.

Using no *a priori* factor information implies that the factors are extracted from the data (the return covariance matrix) as statistical artefacts. An early example is King [1966]; in the context of the APT, we have the seminal studies by Gehr [1978] and Roll & Ross [1980], followed by numerous others. The use of factor analysis is a natural choice, since the strict MFM in eqs.(2.35-39) exactly represents the linear hypothesis underlying factor analysis.[78]

An alternative route is to hypothesize on the identity of the return generating factors and to condition security returns on this set of (external) factors. The seminal example of this approach is Chen, Roll & Ross [1986]. An obvious advantage of using *a priori* specified factors is that the MFM gains economic-intuitive content and that economic meaning can be attached to the factor sensitivities. In addition, some statistical (indeterminacy) problems associated with factor analysis are avoided.[79] On the other side, it is difficult —or

---

[78] Cf. Mardia, Kent & Bibby [1979, Ch.9]. Since the pricing in the (pure) APT is asymptotic and allows for some degree of residual correlation, the use of (computationally more simple) principal components analysis is justified to provide asymptotically consistent estimates of an approximate (i.e. non-strict) factor structure in cross-sectionally large samples. Cf. Chamberlain & Rothschild [1983], Grinblatt & Titman [1985] and the extensions of Connor & Korajczyk [1986,1988]. Since it is not clear how large the sample must be to invoke the limit results, the argument is vague. Huang & Jo [1992], however, provide empirical evidence that the use of asymptotic principal components analysis involves virtually no loss of information for even less than 100 securities. Thus the difference between principal components-based MIMs and factor analysis-based MFMs fades.

[79] For these problems, see Elffers, Bethlehem & Gill [1978] and Raveh [1985].

rather impossible– to control the diagonality property: the set of plausible economic variables may be exhausted before all residual return correlation is accounted for.

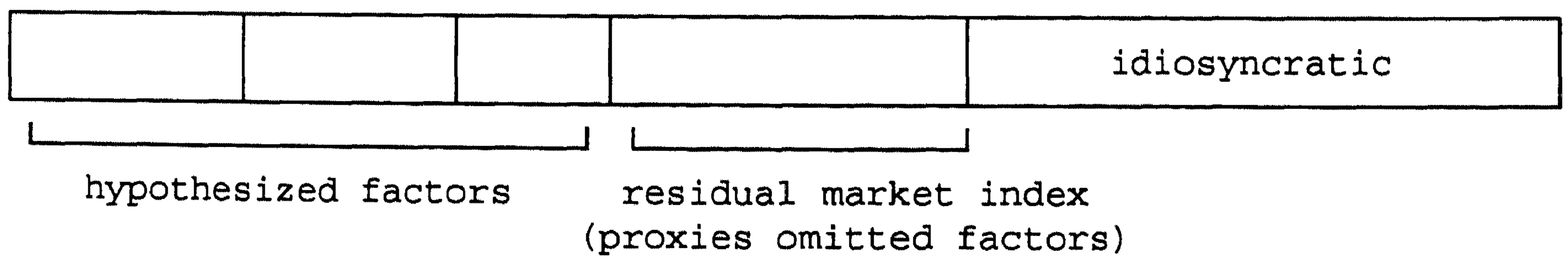
When external factors are specified or statistical factors are extracted, a next step may be taken by proxying these factors by internal portfolios. These 'mimicking portfolios', 'factor portfolios' or 'basis portfolios' are composed from the securities in the opportunity set in order to represent the factors in an adequate way. The basic technique dates back to Black & Scholes [1974, pp.9ff]. In the context of the APT, factor mimicking portfolios are widely applied.[80]

We start with the case in which factor analytical techniques are used to estimate the factor model. Chen [1983, pp.1411-1412] developed a procedure to form factor portfolios with unrestricted sensitivity to the factor being mimicked and zero sensitivities to the other factors. In his optimization, Chen minimized the deviation of the factor portfolio's composition with respect to an equally weighted portfolio. This is only a crude way to control the idiosyncratic risk. More sophisticated is the approach of Lehmann & Modest [1988, pp.224ff], who developed a minimum idiosyncratic risk procedure. They choose portfolio weights to minimize idiosyncratic risks, subject to restrictions of full investment and zero sensitivities to all other factors, but again without any restriction on the portfolio's sensitivity to the particular factor being mimicked. A recent application of this technique is Kryzanowski, Lalancette & To [1994].

In the case where the factor model is estimated by applying principal components analysis, the factors already are expressed as linear combinations of the securities. These internal indices are then mimicking portfolios; see for example Connor & Korajczyk [1986] (and our footnote 78 above).

Finally, for the case in which external (economic) factors are used, the goal is to maximize the correlation with the true economic variables. In this context, Connor & Korajczyk [1991] developed a new technique, that explicitly recognizes measurement errors with respect to the economic variables.

**Residual-market-index factor models**




80) In the context of Equilibrium-APTs, Huberman, Kandel & Stambaugh [1987] provide a theoretical discussion of mimicking portfolios.

In the last category, we have residual-market-index factor models. These are extended general MFMs. Their design is the reverse of hierarchical extra-market covariance models.

Starting point is a general MFM. Next, a market index is regressed against the set of incorporated factors. The residual of this regression represents the component of the index that is uncorrelated with the factors. This residual market index component, generally termed 'residual market factor', is then considered as an additional factor in the securities' factor models. The importance of the market's residual return as a factor was indicated by Wei [1988]. From a theoretical perspective, the inclusion of the residual-market-index factor yields exact pricing in an Equilibrium-APT. In empirical work, the market residual is important in its role as a proxy for unobserved factors. This was independently recognized by McElroy & Burmeister [1988], who considered this variable in their empirical and theoretical studies (Burmeister & McElroy [1988,1991]). When added to economic factors in a MFM, the market index residual may especially be relevant for capturing psychological moods that excert a general influence on the securities in the market.

## 2.5 MIMs AND MFMs AS EMPIRICAL CONCEPTS

This section reviews the empirical specification and performance of MIMs and MFMs. We discuss the models according to the typology of the previous section.

### 2.5.1 MIMs as empirical concepts

#### General MIMs: extracted (internal) indices

Elton & Gruber [1973] perform *ex ante* analyses to compare MIMs with SIMs. In the direct approach (i.e. forecasting the return correlation matrix), the market index-based SIM outperforms the internal index-(first principal component) based SIM and the full historical correlation matrix, and both SIMs outperform principal components-based MIMs. This indicates that correlations are better estimated on the basis of broad market influences rather than influences which are specific to the opportunity set and that, from a forecasting point of view, the assumption of zero extra-market correlations is better than using their historical values. It seems that extra internal indices, although explaining a larger part of historical in-sample correlations, introduce noise rather than information in the forecasts. The SIMs in turn are outperformed by the overall mean correlation model; this is consistent with the findings of Elton, Gruber & Urich [1978].

Partitioning the correlation matrix and averaging correlations within and between either traditional industries or 3-group pseudo-industries, Elton & Gruber [1973] find that differences between correlations of traditional industry groupings convey information. Using more disaggregated pseudo-groupings again produced more noise than information. The indirect approach (using *ex ante* efficient frontiers) confirmed these results.

### General MIMs: industry indices

One of the earliest studies in the field is by Cohen & Pogue [1967], who compared the performance of the SIM and industry-based MIMs to the performance of the full covariance model. They use 10 industry indices, based on traditional industry groupings.

Compared to the conditioning of the stock returns on an overall index, the conditioning on the ten industry indices does not yield a major reduction of historical residual covariability. As the industry indices are highly correlated, this is not very surprising. On average, the SIM better represented the true historical correlation matrix. Although the MIMs approximated more closely the relatively small number of intra-industry stock correlations, the SIM accounted better for the relatively large number of inter-industry correlations. So in an historical context, the MIMs introduced more noise into the implied relationships between stocks in different industries than the SIM.[81]

In the indirect approach (also *ex post*), Cohen & Pogue found that the SIM frontier dominates the MIMs frontiers, especially in the low risk-return region. Also, for equivalent levels of mean returns, the composition of the SIM portfolios much more resembled the composition of the full covariance portfolios than that of the MIMs. Using full covariance information to compute the actual means and variances of the portfolios' returns, it appeared that the SIM understates portfolio variances to a greater extent than the MIM, indicating superfluous (spurious) diversification. Still, the SIM frontier dominates the MIM frontier.

---

[81] Cohen & Pogue [1967, p.178] hypothesize that this is due to the high correlations between the industry indices. We can clarify this matter by recognizing that there are only 11 (sic!) annual return observations in the base period. Ruling out perfect correlations between individual stock returns, the correlation matrix has rank 11. This implies that the correlation matrix can be reproduced exactly by only 11 principal components, leaving only 1 degree of freedom for Cohen & Pogue's MIM. So, using a 11-index model with orthogonal indices, extracted from the data, the correlation matrix could have been reproduced **exactly**. This would be the case if we converted Cohen & Pogue's MIM in diagonal form into a general MIM with one overall (market) index and 10 industry indices.

Wallingford [1967], surprised by Cohen & Pogue's [1967] results, conducted a small scale comparative study. Using an internal index-based SIM and two internal sub-indices in his MIMs, it appeared that the MIMs dominated the SIM throughout almost the entire range of expected returns (as expected, the MIMs were in turn dominated by the full covariance model). Wallingford hypothesized that the conflict with Cohen & Pogue's [1967] findings could be traced to a reduced inter-index correlation, which would improve the performance of the MIMs. In our opinion, a more important factor is that Wallingford's use of internal indices increased the correlations between the individual stocks and these indices. In addition, Wallingford used only two indices in his MIMs. Consequently, the number of correlations between stocks belonging to different index groups is smaller than the number of correlations between stocks belonging to the same index group. The security-index correlations are thus more important than the inter-index correlations. Cohen & Pogue [1967] instead, used ten indices in their MIMs; as a result, there are far more inter-group than intra-group correlations.

Alexander [1978] systematically investigated possible sources of discrepancy between Cohen & Pogue's [1967] and Wallingford's [1967] findings. His results indicate that, in addition to sampling differences, the selection of the indices forms a major element in the relative superiority of SIMs over MIMs. The SIM tends to dominate the MIM when broad market and industry based indices are used (supporting Cohen & Pogue's [1967] findings and Wallingford's [1967, p.103] hypothesis); the use of internal indices enhances the performance of a MIM relative to a SIM.

Alexander [1978] employs the broad market-based index, used in his SIM, as the first index in his MIM. It is then just a small step to extra-market covariance models, considered next.

**Extra-market covariance models**

King [1966] starts from the hypothesis that security returns are generated by a MFM that incorporates a general market effect, an industry effect and a security specific effect. After (factor analytically) extracting the first factor, he finds that the residual covariance matrix shows a partitioning in agreement with traditional industry groupings. As multiple factor analysis confirmed industry effects, he suggests to extend the SIM to a MIM that incorporates industry indices.

By extending the number of industries incorporated in the sample, Meyers [1973b] finds less consistency between traditional industry groupings and pseudo-industries (obtained from cluster analysis and principal components analysis on the extra-market correlation matrix). Fertuck [1975] extends this *ex post* result to an *ex ante* context by observing that the strength of an industry effect crucially depends on the homogenity of the industry and on the level of aggregation in defining an industry. This is consistent with Elton & Gruber's [1973] findings, mentioned before. Livingston [1977] in turn points at the sample sensitivity and technique sensitivity of the results (especially relating to the use of factor analysis and principal components analysis for identifying market and extra-market effects), but confirms Meyer's [1973b] and Fertuck's [1975] findings that industry effects show great variability.

Farrell [1974] draws attention to significant correlations among industry groupings, that can be interpreted as evidence of extra-market comovements that go beyond industry classification. In addition to the three factors considered by King [1966] (i.e. a general market factor, an industry factor and a specific factor), he argues that there exists a broader-than-industry classification according to which industries can be amalgamated into homogeneous groupings or "market sectors" (Farrell [1983, p.45]). These stock groupings are homogeneous in the sense that the correlations between stock returns within each group are high and positive, whereas there exist no (positive) correlations between the groups. Farrell decomposes the extra-market risk components of returns to classify the securities according to their actual behavior instead of to their industry, thereby extending the concept of pseudo-industries to homogeneous groups. Performing cluster analysis on the residual correlation matrix, he suggests a classification of stocks according to *(1)* growth, *(2)* cyclical, *(3)* stable and *(4)* oil stocks.

While the S&P 425 explained on average 31% of the return variance, the four stock groupings (indices) in a four-index model contributed on average an additional 14% to the explanation of stock returns. Farrell [1974, pp.200-201 fn.19] adds 10% for industry effects, concluding that systematic factors (general market, market sector or grouping, and industry) may account for 55% of the return variance. In our opinion, however, it is incorrect to consider a market sector effect as a systematic factor **in addition** to the effects of a general market factor, an industry factor and a specific factor, as Farrell [1974, pp.187,202] does. Both the homogeneous groupings and the industry groupings provide an explanation for extra-market return (co-) variance, although on a different level of aggregation. This means that there exists an overlap between the sector effects and the industry effects; only for the oil stocks both effects coincide. As a result, the 55% systematic component of return variance is an overstatement.

Updating his study, Farrell [1975, 1983] found that the same four homogeneous groups emerged and that the grouping effect seemed to maintain its same relative power in terms of proportion explained return variance. Furthermore, Farrell [1983, pp.53-55] finds that the superiority of the four-index model in an *ex post* context carries over to an *ex ante* context and that the extension of the SIM to this MIM indeed captures more information than noise.

Like Farrell, Arnott [1980] decomposes the extra-market risk components of returns to classify the securities according to their actual behavior instead of to their industry. By stopping the clustering procedure in an earlier stage, however, Arnott finds five 'comovement classes' that are more homogeneous (i.e. contained less weakly related stocks) than the groupings of Farrell and explained a far greater fraction of the extra-market return variability: *(1)* quality growth, *(2)* utilities, *(3)* oil and related, *(4)* basis industries and *(5)* consumer cyclical. Again these groups can be treated as indices and serve as inputs in a MIM in addition to a broad market index.

In our opinion, the contribution of Arnott [1980, pp.58-59] is that he explicitly links the comovement class indices (or 'major extra-market factors' in his terminology), extracted from the actual stock return behavior, to economic variables and fundamental characteristics of the stocks. So is the class of utilities dominated by interest rate sensitivity and is the cyclical class strongly influenced by the economic outlook. This provides an economic flavour to the statistical model.

With Rosenberg & Marathe [1976], we want to stress the fact that indices can be regarded as proxies for true underlying return generating economic factors. By means of a MIM, we actually try to approximate the factor structure of the securities, i.e. to estimate the model with the return generating factors and the sensitivities of the returns for these factors. In isolating a comovement class of securities, one has found a group of securities of which each return is (partly) generated by the same factor. The common dependence on this factor creates the strong inter-relationships between these securities.

**Towards MFMs**

Aber [1976] considers various models, among which a mixed MIM/MFM, and investigates their ability to eliminate residual return covariability. In each of these models, the S&P Index is supplemented with other indices. In the industry-MIM, 9 traditional industry indices are added. In the group-MIM, 6 stock groups (high quality, glamour, cyclical, defensive, conglomerate and growth) are added. In the mixed MIM/MFM finally, 4 macroeconomic variables (monthly Treasury Bill rate, 20-year corporate Aaa bond rate, an index of Industrial Production and the Consumer Price Index) are added to the S&P Index and the 6 stock groups. As Aber's objective is to examine the effect of using nonindustry-based indices, he does not provide the rationale for the stock groupings and the macro-economic variables. We note, however, that some of the groupings are familiar from Farrell's [1974] study and that Smith [1969] used some of the factors in a univariate context.

Aber's results indicate the potential superiority of MIMs to the SIM in producing uncorrelated return residuals, at least in an historical context. Furthermore, they show that nonindustry-based MIMs may outperform industry-based MIMs. Especially the performance of the mixed MIM/MFM warrants interest in the use of economic variables. This approach is discussed below.

### 2.5.2 MFMs as empirical concepts

In discussing empirical aspects of MFMs, we should in fact start with King's [1966] study and explore various applications of factor analytical techniques. However, extracted factors are statistical artefacts. To reduce the level of abstraction somewhat, one can consider factor (mimicking) portfolios or other security groupings that are meant to represent the factors. This approach leads to MIMs, discussed above. In this section, instead, we focus on the role of economic variables as factors.

Many studies explored the link between security returns and specific economic variables. There is, for example, a long time interest in inflation and the seemingly anomalous relationship between inflation and stock returns.[82] Also money supply, and economic activity and business cycles are studied extensively in the inflation context.[83] Other variables that are considered in relation to stock returns are interest rates.[84]

#### Predictive models, pricing models, and risk models

Models that relate security returns to economic variables can be distinguished in predictive models, pricing models and risk models.

In predictive models, security returns are related to economic variables known in advance. More specifically, the **levels** of these variables are considered (like the level of interest rates, the term spread, the dividend yield or the level of expected inflation) and used to detect persistent return components over time and to analyze the time-varying **level** of expected returns or expected risk premia. In this way, expected returns are predicted. Some examples of this approach are

[82] Some classical theoretical and empirical studies in this field are Nelson [1976] and Fama & Schwert [1977].

[83] Money supply is studied by Pearce & Roley [1983], among many others. Real activity, as source for a spurious relationship between inflation and stock returns, is discussed by Fama [1981] and Geske & Roll [1983] (who counter the Modigliani & Cohn [1979] irrationality-hypothesis). Hasbrouck [1984] provides an overview of the arguments. Some recent studies on the relationship between business cycles and stock returns include Bulmash & Trivoli [1991] and Krainer [1992].

[84] Interest rates will be discussed in more detail in chapters three and four. More exotic variables include the 'presidential factor', related to influences of changes in (economic) government policy and/or psychological factors (cf. Herrick [1988]).

Keim & Stambaugh [1986], Campbell [1987], Fama & French [1989], Chen [1991] and Ferson & Harvey [1991].

In pricing models, security returns are related to contemporaneous changes in economic variables. More specifically, **innovations** (unanticipated changes or surprises) in these variables are considered, and the relationship between the securities' sensitivities to these changes and their expected returns is analyzed. In other words: the relationship between expected return and risk is explored. The APT is a marked example.

In risk models, finally, the **contemporaneous** relationship between security returns and economic innovations ('economic news') is analyzed, without a restriction that the corresponding risks (in the form of the securities' sensitivities) are priced. As a result, pricing models are nested in risk models. It is the latter models that we are interested in in this study.

### Chen, Roll & Ross [1986] and related studies

Still, the APT provided a strong stimulus for examining the relationship between economic variables and security returns. The first tests of this model relied on factor analysis (Gehr [1978] and Roll & Ross [1980]), but it soon appeared that the application of this technique carried many statistical problems.[85] In addition, the unknown identity of the key factors is unsatisfying and the opinions on the number of pervasive factors differ.[86]

Chen, Roll & Ross [1986] set the tone by considering explicit economic variables. They start from the traditional discounted cash flow valuation formula and infer that security prices (or returns) will be affected by changes in the expected cash flows and in the risk adjusted discount rate. Real and nominal forces that act upon the expected cash flows are economic activity and inflation. Related to the discount rate are the term structure of interest rates and the risk premium. From these considerations, they *a priori* specify the following set of economic factors:

---

85) Cf. Dhrymes, Friend & Gultekin [1984] and our footnote 79 above.

86) Roll & Ross [1980] found four to five pervasive factors. Using all available CRSP data, Trzcinka [1986] finds only one dominating return generating factor, but considering the long sample period, this may be the result from non-stationarities in an underlying multiple factor structure. Conway & Reinganum's [1988] results suggest the presence of one dominant factor and one minor factor; the application of standard maximum likelihood factor analysis would overstate the number of factors. Brown [1989] shows that the detection of only one return generating factor can be explained when there are in fact multiple 'equally important' **priced** factors. Connor & Korajczyk [1993] finally, find evidence for one to six pervasive factors. So the debute on the number of factors is far from settled.

- the monthly growth rate of industrial production. Since industrial production is a flow during some month, this variable measures the change in industrial production lagged by at least a partial month. They therefore lead this variable one month;
- unanticipated inflation;
- the change in the expected rate of inflation;
- the change in the default risk premium, measured as the difference between the return on a low grade corporate bond index and the return on an index of long term government bonds;
- the change in the term structure spread, measured as the difference between the return on the index of long term government bonds and the Treasury Bill rate.

Note that the last two variables are measured in terms of returns. As bond returns are directly (negatively) related to changes in yields, these variables could equally well be expressed in terms of differences in yield changes. (The change in the term structure spread thus indicates changes in the slope of the term structure of interest rates.) In addition, Chen, Roll & Ross [1986, pp.386-390] intend to incorporate innovations in the variables as relevant factors, but they argue that some of the variables are noisy enough to be incorporated without pre-whitening.[87] On this and many other points, their statistical design is heavily criticized by Poon & Taylor [1991].

Hamao [1988] replicates the Chen, Roll & Ross [1986] study for the Japanese stock market, and many other researchers have gratefully used the Chen, Roll & Ross set of economic factors as a starting point for their own analyses (the way in which the variables are measured sometimes differs, though). Chan, Chen & Hsieh [1985], for example, use the same set of factors supplemented with a business cycle indicator (a measure of net entries to the business populations, partly based on the change in the number of business telephone lines). Burmeister & Wall [1986] and McElroy & Burmeister [1988] in turn, disregard the change in expected inflation, but supplement the set of factors with a market index residual.[88] Furthermore, they proxy the industrial production variable by the unanticipated growth rate of real final sales. Connor & Korajczyk [1991] use the same set of factors as the latter studies (also omitting changes in expected inflation), but proxy the industrial production variable by unanticipated changes in unemployment. Chang [1991] not only adds the changes in industrial production to this set

---

87) Pre-whitening is the process of removing predictable components from a variable, mostly by means of time series (ARIMA-) modelling. When this is accomplished, the residual component is white noise and can be interpreted as an innovation, or surprise, or 'information'.

88) Burmeister & Wall [1986, p.9] use the residual market index return to compare proportion return variance explained by the economic factors and the market index. This motivation to include a residual market index factor is different from the argument used by McElroy & Burmeister [1988] and discussed at the end of section 2.4.2.B.

again, but also considers a foreign exchange rate variable.[89)]

A quite different approach is followed by Kim & Wu [1987], who use principal components analysis to determine the three most significant combinations of a number of 14 economic variables.

**Explanatory power of factor models; risk models**

All of these studies employ the factors in a pricing context.[90)] As their study is intended to be a test of the APT, Chen, Roll & Ross [1986] do not provide much information on the estimation results of the underlying factor model. Only the first version of their paper (Chen, Roll & Ross [1983, Table 3]) incorporates more or less explicit information about the relevance of these factors in a time series context. They (factor analytically) extracted factors from the return covariance matrix and formed five factor mimicking portfolios. The returns on these portfolios were then each related to the economic factors. It appears that the economic factors explained on average a proportion of about ($R^2$=) 5% of the variance of the returns on each of these portfolios. It is discomforting then, that a general stock market index explains a significant proportion of the time series fluctuations of the security returns but appears insignificant in pricing. The industrial production variable, in contrast, is not significantly related to the stock market index but appears to be the most significant variable in the pricing relationship.[91)] This leads Chen, Roll & Ross [1986, p.399] even to remark that "[t]his suggests that the "explanatory power" of the market indices may have less to do with economics and more to do with the statistical observation that large, positively weighted portfolios of random variables are correlated." This ignores the fact that a market index represents the amalgam of factors that influence individual stock returns.

Other pricing studies more explicitly mention the performance of factor models in a time series context. Burmeister & Wall [1986, p.11]

---

89) Bodurtha, Cho & Senbet [1989] explicitly transpose an economic factor model to an international context. In the same context, Beckers, Grinold, Rudd & Stefek [1992] investigate international pricing on the basis of *a priori* specified (inter-)national equity attributes.

90) It may be noted that Kim & Wu [1987] and Chen & Jordan [1993] compare the pricing abilities of extracted factors versus economic variables and find only small differences.

91) This point is also put forward by Poon & Taylor [1991, p.632]. At the same time, however, they state that "[i]t does not matter if the exposure of the stock returns to particular macroeconomic factors is statistically insignificant or erratic. What matters is that cross-sectional returns are linearly related to measures of exposure to that macroeconomic factor". However, as Chen, Roll & Ross [1986] use Fama & McBeth's [1973] two-stage procedure, the use of insignificant first-stage time series sensitivities in the second stage cross-sectional regressions is questionable, in our opinion.

and McElroy & Burmeister [1988, p.38] report that their four economic factors explain about 25% to 30% of the variance of the S&P 500.[92]

In addition, there are studies that consider factor models only in a risk context (i.e. without imposing cross-sectional constraints).

Berry, Burmeister & McElroy [1988a] use the factor set of Burmeister & Wall [1986] and McElroy & Burmeister [1988], and confirm that the four economic variables explain about 25% of the S&P 500 return variance. Including the residual-market-index factor, they report proportions explained variance ranging from 50% to 84% for seven economic sectors and from 15% to 75% for 82 industries.[93]

Chang [1991] explicitly investigates the intertemporal stationarity of factor sensitivities and the factor model's conditional forecasting ability. He finds that the market model performs better than the economic variable model; after including the residual-market-index factor in the model, its forecasting ability equals that of the market model. Kim & Wu [1987, pp.95-96] find that the conditional forecasting ability of the simple market model and of the economic factor model are quite similar. The gain of using economic variables may then not be measured in terms of statistical performance, but defended by increased insight of risk structures.

## 2.6 SUMMARY AND CONCLUSIONS

This chapter presented a critical overview of various return generating processes. We discussed the models in both a theoretical and an empirical context and explored their role in portfolio theory.

---

92) Connor & Korajczyk [1991, p.12] also report substantial correlation between market index returns and economic variables. As they proxy the latter variables by 'rotated factors' (i.e. linear combinations of five statistically extracted factors exhibiting maximum correlation with each of the macro-economic variables), these correlations overstate the return correlations with the actual underlying economic variables.

For the Japanese stock market, Elton & Gruber [1988, p.41] use canonical correlation analysis to link an extensive set of economic variables to four factor-analytically extracted factors and report correlations of 30% and 64%, depending on the time period. By the same argument, these numbers overstate the degree of correlation between returns and economic variables.

93) Although we have some comments on his methodology and conclusions, we mention that Asprem [1989, p.606] reports $R^2$'s ranging from .24 (Italy) to .71 (the Netherlands). He estimated economic factor models for ten European countries using 14 years of quarterly data. For three European countries, Wasserfallen [1989] uses only 9 years of quarterly data and finds that the univariate relationships between stock returns and macro-economic factors is very weak.

From a theoretical point of view, we defined the models and discussed the underlying assumptions. In particular, we touched upon the singularity property of the market model, upon the diagonality condition of single index models and strict multi-index / multi-factor models, and upon the internal consistency problem of the latter models. Although these models can be **imposed** on the data, we investigated the conditions under which these linear models are **implied** by the data. It appeared that a joint symmetric elliptical distribution of security returns and index or factor returns – which also covers a joint symmetric stable (Paretian) distribution – implies linearity of the models.

From an empirical point of view, we emphasized the problems encountered in estimating the market model. This special attention is warranted since many empirical problems readily carry over to the other models we discussed. Whether estimating the market model, index models or factor models, one is confronted with issues of stability and stationarity. The tendencies of estimated sensitivities to change over time (as a result of estimation errors or true underlying causes) and the effects of portfolio grouping (either beneficial or detrimental) should clearly be recognized in a much wider frame.

Index models and factor models play an important role in portfolio theory as they allow to simplify the securities' return interactions. In this respect, we touched upon the dangers of oversimplifying the covariance structure of security returns, a typical situation where a beautiful hypothesis is slayed by an ugly fact. We provided a categorization of a wide range of proposed models and commented on their theoretical characteristics and empirical relevance. The categorization provided ample opportunities to compare the models.

In the particular interest of our study, we paid special attention to multi-factor models, which relate security returns to tangible economic variables. In presenting a portfolio model, Elton & Gruber [1992, p.8], referred to the case "when the indexes in the return-generating process are economic variables, as suggested by the recent literature", on which their footnote 5 (p.15) simply states: "See, for example, ???" This nicely indicates the large choice in *a priori* variables that can be considered and implies the need for an overview.

We made a distinction between predictive models, pricing models and risk models. For our purpose, we are interested in risk models. In risk models, a contemporaneous relationship between security returns and innovations ('news' or 'surprises') regarding economic variables (factors) is specified. Many researchers rely on the factor sets of previous studies, most notably Chen, Roll & Ross [1986]. But this is a pricing study, and the set of relevant variables is not necessarily restricted to priced factors as we argued before. Outside a pricing context, factor models have received only limited attention. As a result, the empirical properties of factor models and economic variables in a time series context are not as well documented as their pricing abilities. What we did observe, however, was a discrepancy between variables that perform well in a pricing context and variables that are able to explain significant proportions of return variability. Especially the latter aspect (in an *ex ante* context) is crucial for factor models as risk models. Most promising are variables that refer to financial markets, like interest rates, risk premia and exchange rates. Because they are linked to securities that are traded on liquid and competitive markets, they reflect economic news in an immediate fashion. In chapter four, we elaborate further on this issue.

In any case, the quest for relevant economic factors continues. In the context of risk models, this is still a large field for future research. The experience gained by estimating the simple model with which we started off this chapter, should be of help in efforts to model the relationship between security returns and economic variables, directed at a more complete understanding of investment risk.