Chapter 3: FACTOR MODELS AS STATISTICAL AND ECONOMIC CONCEPTS

3.1 INTRODUCTION

In general, the adequate representation of risk is a predominant problem in investment analysis. In this light, multi-factor models (MFMs) form a field of increasing theoretical challenge and practical application. Theoretical aspects of MFMs are the subject of this chapter. We discuss the meaning and importance of these models by highlighting some of their statistical and economical aspects that are relevant, especially in the field of portfolio analysis. A discussion of empirical aspects and (potential) applications of MFMs is postponed until chapters four and five.

Generating and conditioning in factor models

In the previous chapter, the central focus was on models that linked security returns to one or more (stochastic) variables. Although the starting point of that chapter was fairly general, these models were, almost without exception, simplified in some specific way (for example, linearized in the parameters) and proposed **ex cathedra**. In this chapter, we first return to the general setting.

Formally, we start again from the general assumption that a security's return in period t is generated by some function $\Phi_t(\cdot)$ of k factors, summarized in the vector $\underline{\boldsymbol{\delta}}_t = [\underline{\delta}_{jt}]_j$:

$$(2.1) \quad \underline{r}_t = \Phi_t(\underline{\boldsymbol{\delta}}_t) + \underline{\xi}_t$$

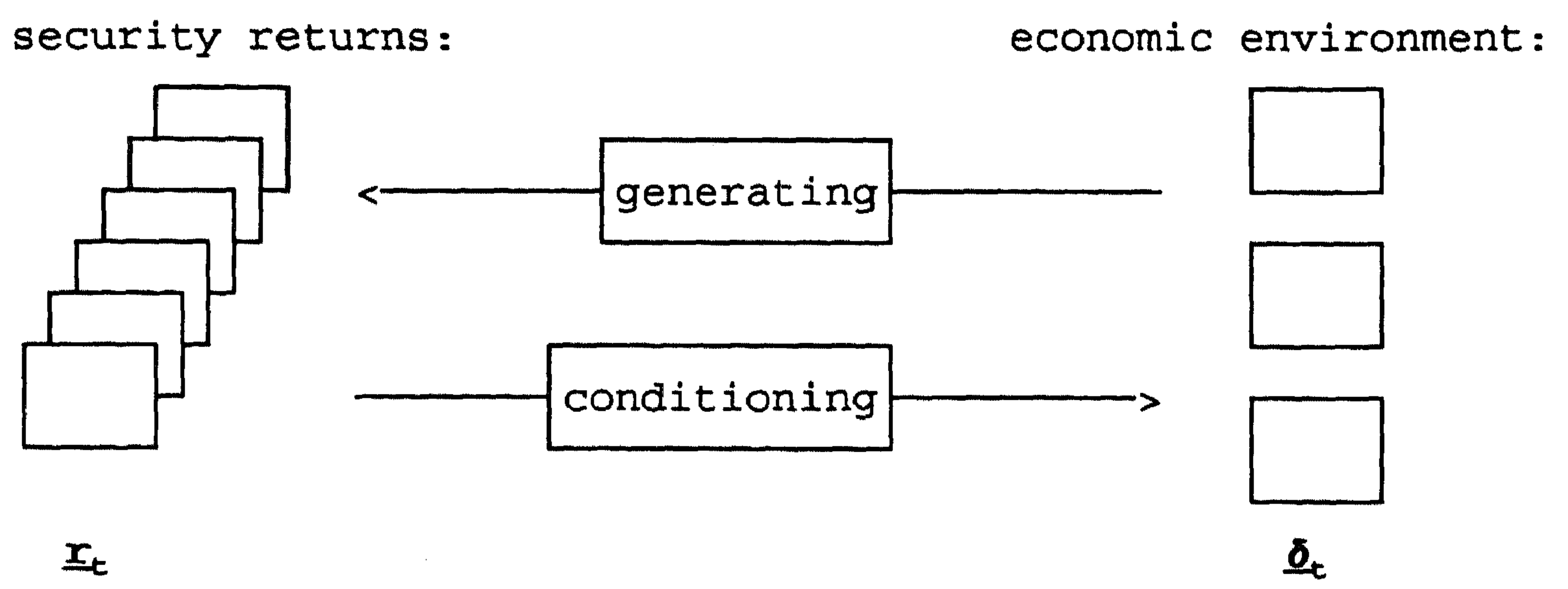
In this MFM, the random error term $\underline{\xi}_{it}$ denotes the unmodelled part of the return and accounts for the influence of various other factors. It has a zero mean and satisfies:

$$(3.1) \quad E(\underline{\xi}_t | \underline{\boldsymbol{\delta}}_t) = 0$$

Unless indicated otherwise, we will tacitly assume intertemporal stationarity of the functional relationship, so $\Phi_t(\cdot) = \Phi(\cdot)$.

Conceptually, a MFM entails a representation of the linkage between security returns and the relevant economic environment. The 'relevant economic environment' is then the generic term for a set of common underlying stochastic state variables or factors. This linkage has two interpretations. In one view, we start from the economic environment. A MFM then tells us how changes in (investors' perceptions of) this environment generate changes in security prices and, hence, returns. The complementary view starts from the security returns; a MFM then represents a vehicle to condition these returns on changes in the environment. This is schematically depicted in Figure 3.1. In section 3.2, the process of conditioning security prices (and returns) on state variables is discussed in more detail. The linearization of the models and the interpretation of the resulting sensitivity coefficients will have special attention.

Figure 3.1: The linkage between security returns and the relevant economic environment.



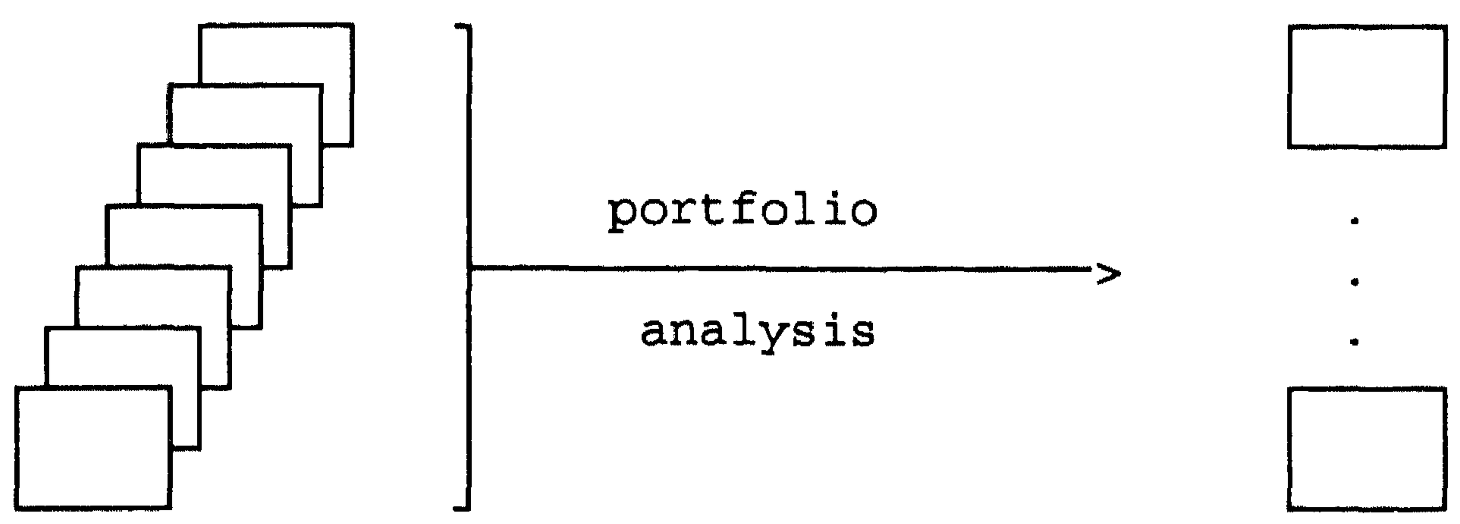
Factor models and the reduction of complexity

Section 3.3 further develops MFMs as statistical concepts and shows how conditioning can reduce two fundamental problems of the investment decision: the information problem and the combination problem.

These two complex problems stem from the interrelatedness of security returns. As an investment portfolio is a convex combination of individual securities, the distribution of any portfolio's return is a (more or less complex) blend of the distributions of the returns on the individual securities that are included in the portfolio. As these returns are statistically dependent, their **joint** distribution is relevant for portfolio analysis. Hence, when considering an opportunity set of N securities, all possible interactions between the N securities must explicitly be considered in the security analysis. This introduces an information problem: how to estimate the necessary inputs for the portfolio problem on the level of the individual securities? Once this problem has been overcome, we are faced with the combination problem: how can the relevant characteristics of the securities be processed and aggregated into portfolio characteristics? The problems of assessing the relevant probability characteristics of both individual securities and any portfolio is depicted in Figure 3.2.

Figure 3.2: Two central problems of portfolio selection: gathering the necessary input information on the security level and aggregating the relevant information to the portfolio level.

information problem: **combination problem:**



security returns: $\underline{r}_{it}$ portfolio returns: $\underline{r}_{pt}$

For the sake of tractability, the joint distribution of the securities needs to be simplified. By using **exhaustive conditioning**[1] and a portfolio context, the attention can shift from the joint distribution of security returns to the joint distribution of the factors. When the returns on a number of securities are conditioned on a far less number of common factors, this reduces the dimensionality of the ways in which the returns interact within a portfolio. In this way, MFMs serve the role of simplifying distribution assumptions.

'Risk' in a factor model

By conditioning security returns on a set of common factors, the model (2.1) allows the decomposition of these returns into a 'systematic' part $\Phi(\cdot)$ and an 'unsystematic' part $\underline{\varepsilon}_{it}$. This implies that the distribution of a security's return is generated by the distributions of the common factors under consideration as well as by the distributions of other factors. The conditioning of the security returns materializes in the form of sensitivity coefficients or 'betas'. Section 3.4 explores to what extent and in what circumstances these sensitivities can be regarded as risk measures.

In a very general way, the risk accompanying a portfolio's return is related to the shape of its return distribution. Under exhaustive conditioning, the distribution of the unsystematic (i.e. idiosyncratic) portfolio return component degenerates during the process of (naively) diversifying the portfolio. So the return distribution of a well-

[1] Exhaustive conditioning implies that the conditioning factors account for all of the interrelationships that exist between the securities. The residual return components are then truly idiosyncratic: they are independent of the factors and independent over the securities.

diversified portfolio is (by definition) completely generated by the common factors. Hence, as in a portfolio context the sensitivity coefficients link the return distributions and the factor distributions, these sensitivities can be considered as risk measures. At the same time, this interpretation of betas as risk measures is consistent with the tenet of portfolio theory that a security's marginal contribution to a portfolio is relevant. Under the assumptions made, a portfolio's return distribution is completely determined by its sensitivities and the factor distributions. As the security's sensitivities reflect its marginal contribution to the portfolio's sensitivities, the former sensitivities can be interpreted as risk measures.

When the conditioning is not exhaustive, the joint distribution of the unsystematic return components determines the degree in which portfolio diversification can increase the relative importance of the systematic return components. At least to an approximate degree, the arguments of the exhaustive case can be carried over. More important in this more general case, however, is that still a part of a portfolio's return distribution can be manipulated via its sensitivities. For example, expectations of a portfolio's return can be formed or forecasts can be made, conditional on some values for the factors under consideration. This implies that a portfolio can be formed on the basis of specific expectations for future materializations of the factors. When this constellation of anticipated factor values (or 'factor scenario') does not come true, the unanticipated factor movements together with the corresponding sensitivities are responsible for generating part of the portfolio return. Again, the sensitivities can be interpreted as risk measures.

Deriving factor models of security returns

The risk concept that emerges from conditioning investment returns on multiple variables has two important characteristics. First, it is a **conditional** risk concept. The uncertain future portfolio returns depend on both the uncertain future realizations of the factors and the portfolio's factor sensitivities. Using these factor sensitivities as inputs for the process of risk management, portfolio choice is characterized by conditional optimality. Second, as a factor model implies a decomposition of risk, investment risk has become a **multi-dimensional** concept, in which the sensitivities are attributes of the securities. In comparison with the (E,σ^2)-framework, this has great intuitive appeal. However, in order to enhance the interpretation of the risk concept, it is required that the factors are no anonymous statistical constructs but appealing economic variables. So, if we would be able to relate the thusfar anonymous general factors to identifiable economic variables, then the transparancy of the risk concept is enlarged. The use of economic scenarios (see chapter five, section 5.4.1) then fits in the portfolio selection process. By elaborating the multi-farious risk concept in **economic-intuitive** terms, we expect to join in with investment decision-making in practice.

The economic content of the risk concept can only materialize when the return generating function $\Phi(\cdot)$ is concretisized, especially with respect to the identity of its arguments. Unfortunately, there exists no agreement on the appropriate number and the identity of the relevant conditioning variables. In section 3.5, we will explore a methodology for deriving multi-factor models of security returns. The basic idea is to develop an adequate price function for securities in which state variables act as arguments. With the help of a closed form version of this price function, a relationship between the securities' returns and changes in the state variables can be modelled.

The theoretical basis is the present value or discounted cash flow framework, indeed a classic tool in finance. In present value models, it is assumed that the current price (present value) of a security is equal to the appropriately discounted value of all future expected cash flows accruing to the investor. This present value function replaces the anonymous price function. It can now be assumed that future cash flows are influenced by a number of economic variables or factors, characterizing the economic environment in which these cash flows are generated. This implies that the present value function can be recasted in terms of current expectations with respect to future realizations of these economic variables. The latter expectations can be termed 'information variables' and —in the form of state variables— they now act as arguments of the price function. A revision of the expectations towards the factors (induced by unanticipated factor movements, e.g.) will result in changes in security prices, generating a return. Hence, the return generating process can be cast in the form of a multi-factor model, where changes in factor anticipations (i.e. changes in information variables) are linked to security returns.

The advantage of this present value approach lies in our opinion in the fact that the resulting expressions of the elasticities or sensitivities for the factors can be examined, providing insight in the nature of the relations between returns and factor movements and in their expected intertemporal stationarity. As 'Dividend Discount Models', present value models are often used to derive the required or expected return on securities. Instead of then relying on other models to derive risk measures of these securities, we show how the risks, accompanying this expected return, can be identified and measured.

We realize that the quality of inferences from a model crucially depend on the quality of the model (like the validity of its assumptions). We therefore also the address the empirical validity of present value models.

Section 3.6 summarizes this chapter.

3.2 CONDITIONING SECURITY PRICES/RETURNS ON STATE VARIABLES

In deriving a multi-factor representation of security returns, we start from the paradigm that at any time t, a security's price P can be functionally related to some set $\{\theta_j\}_{j\in k}$ of k state variables. At time t, the values of the factors can be summarized in the vector $\boldsymbol{\theta}_t = [\theta_{jt}]_j$. With time, these state variables describe the states of the world. A security's **price function** at time t (i.e. at the beginning of period t) can then be denoted as:

$$(3.2) \qquad P_t = P(\boldsymbol{\theta}_t;t)$$

where the time t is explicitly incorporated as an argument. Note that the price function is not synonymous with the pricing function, which specifies the equilibrium price of the security. Eq. (3.2) gives the security price conditional on time and the values of the state variables at that time.

This conditioning can take several forms. We have **complete conditioning** when the total set of state variables is considered. When only some subset of state variables is considered, then the conditioning is **partial** (or incomplete). In the latter case, an error term that accounts for the effects of the omitted factors must be added to the RHS of eq. (3.2). For the time being, we assume that the conditioning is complete.[2]

Having linked the security's price to a set of state variables and time, we can proceed to derive the relationship between the security's return and a change in the value of the factors over some period of time. Under discrete compounding, the return on the security over the period t (i.e. over the interval $[t,t+\Delta t]$) is:

$$(3.3) \qquad r_t = P(\boldsymbol{\theta}_{t+\Delta t};t+\Delta t)/P(\boldsymbol{\theta}_t;t) - 1 = \Delta P_t/P_t$$

where Δ is the forward difference operator. The return will have three components:

[2] Under complete conditioning, the state variables describe all possible states of the world. Hence, all changes in security prices can be allocated to changes in the state variables. Since there is no error term, this implies that we have exhaustive conditioning (cf. footnote 1 above). Under partial conditioning, the securities' error terms may be interrelated or not. Only in the latter case we have exhaustive conditioning. So complete conditioning implies exhaustive conditioning, but the reverse is not necessarily true.

1. the instantaneous return as a result of a change in the values of the state variables ($\boldsymbol{\theta}_t \rightarrow \boldsymbol{\theta}_t+\Delta\boldsymbol{\theta}_t$),
2. the return as a result of a change in time ($t \rightarrow t+\Delta t$), and
3. the cross effects of 1. and 2.

Assuming that the price function $P(\cdot)$ is continuous and analytic on some closed interval of the arguments, it can be approximated to an arbitrary degree of accuracy on the interval by some polynomial.[3)]

In the rest of this section, a general multi-factor representation is developed as starting point for (approximately) linear MFMs. Then the relationship between conditioning, Taylor series approximations and least squares approximations is explored.

3.2.1 A general multi-factor representation

The most straightforward way to relate the change in the security's price to a change in the state variables and time is to apply the multivariate version of Taylor's theorem on eq. (3.2). This yields:

$$(3.4) \quad \Delta P(\boldsymbol{\theta}_t;t) = \sum_{i=1}^{\infty} \frac{1}{i!} d^i P(\boldsymbol{\theta}_t;t)$$

where the i-th differential is defined as:

$$(3.5) \quad d^i P(\boldsymbol{\theta}_t;t) \equiv \left[\textstyle\sum_j \Delta\theta_{jt} \frac{\partial}{\partial\theta_j} + \Delta t \frac{\partial}{\partial t} \right]^i P(\boldsymbol{\theta}_t;t)$$

In interpreting these differentials, the symbolic power between the square brackets is first to be expanded by the multinomial theorem and then the powers of $\partial/\partial\theta_j$ and $\partial/\partial t$ multiplied by $P(\boldsymbol{\theta}_t;t)$ are to be replaced by the corresponding i-th partial derivatives $\partial^i P(\cdot)/\partial\theta_1{}^i$, $\partial^i P(\cdot)/(\partial\theta_1{}^{i-1}\partial\theta_2)$ &c., evaluated at the spanning point $(\boldsymbol{\theta}_t,t)$.[4)]

---

3) This is the Weierstrass approximation theorem; cf. Young & Gregory [1972, p.308]. A function is analytic or regular when the function is defined and differentiable at all points of the domain.

4) For this Taylor series representation of the price function to be notationaly valid, we assumed that the interdependence between the arguments of $P(\cdot)$ is at most of first order. We therefore allow the state variables to be linearly related.
For a proof of the expansion with a remainder term in the case of (two) independent variables, we refer to Lang [1973, p.124] or Courant & John [1974, pp.68-70].
In the general case of dependent variables, the terms $\Delta\theta_j = d\theta_j$ in eq. (3.5) do no longer represent arbitrary increments of θ_j but differentials of θ_j. As a result, additional terms with $d^i\theta_j$
(footnote continued on next page)

Finally dividing both sides of (3.4) by $P(\boldsymbol{\theta}_t;t)$ yields the factor model representation of the security return:

$$(3.6)\quad \Delta P_t/P_t = \sum_j \frac{\partial P(\cdot)/P(\cdot)}{\partial\theta_j}\Delta\theta_{jt} + \tfrac{1}{2}\sum_i\sum_j \frac{\partial^2 P(\cdot)/P(\cdot)}{\partial\theta_i\partial\theta_j}\Delta\theta_{it}\Delta\theta_{jt}$$

$$+ \sum_j \frac{\partial^2 P(\cdot)/P(\cdot)}{\partial\theta_j\partial t}\Delta\theta_{jt}\Delta t$$

$$+ \frac{\partial P(\cdot)/P(\cdot)}{\partial t}\Delta t + \tfrac{1}{2}\frac{\partial^2 P(\cdot)/P(\cdot)}{\partial t^2}(\Delta t)^2$$

$$+ \text{ terms of higher order}$$

with $P(\cdot) = P(\boldsymbol{\theta}_t;t)$.[5] The first line shows the instantaneous return as a result of a change in the factors and the third line represents the return as a result of a change in time. The second line represents the first order cross effects of the state variables and time. The terms preceding ($\Delta\theta_{jt}$, Δt) and ($\Delta\theta_{it}\Delta\theta_{jt}$, $\Delta\theta_{jt}\Delta t$, Δt^2) are the first and second order semi-elasticities of the security price with respect to the factors and time, respectively. These semi-elasticities are the sensitivities of the security's return for first and second (cross) order changes in time and the state variables. Note that the factor model (3.6) is linear in the parameters (the sensitivities), but not necessarily linear in the factors.

(i.e. differentials) appear next to the terms $(d\theta_j)^i$ in the second and higher order differentials $d^iP(\boldsymbol{\theta}_t;t)$, $i\geq 2$. The Taylor series expansion then becomes very complex and unworkable. However, for i=1, eq. (3.5) is valid whether or not the arguments are independent (Yamane [1968, pp.129,137-138]). For i=2, eq. (3.5) is also valid when the arguments are linearly dependent (Yamane [1968, pp.129-130]). Whereas for the sake of generality we allow the state variables $\boldsymbol{\theta}$ to be mutually dependent, we restrict the dependence to a first order relation only for the sake of **notational** simplicity. As a result, the differentials $d^i\theta_j$ equal zero for $i\geq 2$ and the Taylor series takes the presented form.

5) Alternatively, one may expand $\ln[P(\boldsymbol{\theta};t+\Delta t)]$ in order to get the continuously compounded return $\Delta\ln[P(\cdot)]$ as a function of $\Delta\theta_j$ and Δt. Without loss of generality, still another possibility is to express all variables in terms of natural logarithms, since $y = g(x_1,\dots,x_k)$ implies $\ln(y) = \ln\{g[\exp(\ln x_1),\dots,\exp(\ln x_k)]\} = f(\ln x_1,\dots,\ln x_k)$, provided that $y, x_j>0\ \forall j$. We can then write $\Delta\ln[P(\cdot)]$ as a function of $\Delta\ln[\theta_j]$.

3.2.2 Linear and approximately linear multi-factor representations

Discrete time formulation

Under the assumption that terms beyond a certain order are of negligible magnitude, the Taylor series can be truncated. For example, as a first order approximation, the multi-factor representation of the security return eq. (3.4) can be simplified to the total differential:[6]

$$(3.7) \quad \Delta P_t/P_t \approx \frac{\partial P(\cdot)/P(\cdot)}{\partial t} \Delta t + \sum_j \frac{\partial P(\cdot)/P(\cdot)}{\partial \theta_j} \Delta\theta_{jt}$$

$$= a(\boldsymbol{\theta}_t;t)\ \Delta t + \sum_j b_j(\boldsymbol{\theta}_t;t)\ \Delta\theta_{jt}$$

where the drift term $a(\cdot)$ is the security return conditional on no change in the factors and $b_j(\cdot)$ is the sensitivity coefficient of the return for changes in factor j. Note that $a(\cdot)$ and $b(\cdot)$ are functions of the state variables and time. Provided that the changes in θ_{jt} are small and are considered over a small interval, we expect the first order approximation to be reasonable.

The ultimate simplification, finally, is to assume that the first order semi-elasticities are (locally) linear in $\boldsymbol{\theta}_t$ and t. The drift term and the sensitivity coefficients of the return for the factors are then constants (for a specific range of θ_{jt} and t):

$$(3.8) \quad a(\boldsymbol{\theta}_t;t) = \frac{\partial P(\cdot)/P(\cdot)}{\partial t} = a$$

and

$$(3.9) \quad b_j(\boldsymbol{\theta}_t;t) = \frac{\partial P(\cdot)/P(\cdot)}{\partial \theta_j} = b_j, \qquad \forall\ j \in k$$

Under these assumptions, the return generating process eq. (3.7) reduces to:

$$(3.10) \quad r_t = a\ \Delta t + \sum_j b_j\ \Delta\theta_{jt} + \epsilon_t$$

As we used an equality sign, we added an error term ϵ_t. In case of complete conditioning, the error term accounts for the higher order effects of the incorporated factors and time. In case of partial conditioning, this error term also accounts for the influence of omitted

[6] Any (algebraic) dependence structure among the state variables and time is allowed; see footnote 4.

factors (i.e. relevant factors not included in the set $\{\theta_j\}_{j\in k}$). From now on, we will assume the less restrictive partial conditioning.

So far, we have considered the security return over some time interval $[t,t+\Delta t]$ as a result of a projected change in the state variables and time. We can now expand (3.10) by first adding a stochastic flavor to the mathematical results. For this, we attach probability distributions to the changes in the state variables over $[t,t+\Delta t]$ and consider eq. (3.4) over all these possible changes. We assume again that second and higher order effects are negligible and that the first order semi-elasticities are constant in the sense that they do not depend on the particular values that are drawn from the distributions of the state variables. The latter we can accomplish by choosing the expected values of the factors as spanning points. Eq. (3.10) is now valid for any possible (small) changes in the factors and time. Second, we attach the time dimension Δt to r_t and $\Delta\theta_{jt}$, where Δt is independent of t. This implies that we can consider security returns and factor changes over any time intervals of length Δt. Therefore, we can drop the term Δt on the RHS of (3.10).

As a result of these operations, eq. (3.10) becomes:

$$(3.11)\qquad \underline{r}_t = a + \textstyle\sum_j b_j\ \underline{\Delta\theta}_{jt} + \underline{\varepsilon}_t$$

where in chapter two $\underline{\Delta\theta}_{jt}$ is denoted by $\underline{\delta}_{jt}$.
The (unconditional) expected return over the period t is:

$$(3.12)\qquad E(\underline{r}_t) = a + \textstyle\sum_j b_j\ E(\underline{\Delta\theta}_{jt})$$

where we have assumed, without loss of generality, that the error term satisfies:[7)]

$$(3.13)\qquad E(\underline{\varepsilon}_t) = 0$$

Furthermore, we will assume:

$$(3.14)\qquad E(\underline{\varepsilon}_t\cdot\underline{\Delta\theta}_{jt}) = 0 \qquad\qquad \forall\ j\in k$$

Combining (3.12) and (3.13) gives:

$$(3.15)\qquad \underline{r}_t = E(\underline{r}_t) + \textstyle\sum_j b_j[\underline{\Delta\theta}_{jt} - E(\underline{\Delta\theta}_{jt})] + \underline{\varepsilon}_t\ ,$$

which is the general linear MFM in unconditional expectations. According to (3.15), the error term contains the (linear as well as non-linear)

---

7) However, the consequence of eq.(3.13) is that the constant a in (3.12) is no longer defined by eq.(3.8). Likewise, the consequence of eq.(3.14) is that the coefficients b_j in (3.12) are no longer defined by eq.(3.9). See section 3.2.4 and Appendix 3.B.

effects of omitted factors insofar these are uncorrelated with the selected factors.

If the conditioning is exhaustive, the error term represents a factor that is idiosyncratic for each security. For each pair of different securities i and h we then have:

(3.16) $E(\underline{\varepsilon}_{it} \cdot \underline{\varepsilon}_{ht}) = 0 \qquad \forall\ i \neq h$

The error process of the security return, $\underline{r}_t - E(\underline{r}_t)$, is then decomposed in a systematic or factor-related part and an idiosyncratic part.

If the conditioning is non-exhaustive, the error term captures (in part) the effect of omitted common factors and eq.(3.16) will generally not hold. Still, the error process of the security return is decomposed but now in a systematic and an unsystematic part, where 'systematic' is defined with respect to the set of selected factors.

The expected return over period t, **conditional** on the factors at the beginning of that period, is:

(3.17) $E_t(\underline{r}_t) = a + \sum_j b_j\ E_t(\underline{\Delta\theta}_{jt}) + E_t(\underline{\varepsilon}_t)$

where we use the notation $E_t(\cdot) \equiv E(\cdot\,|\,\theta_t)$.
When we are willing to assume that:

(3.18) $E_t(\underline{\varepsilon}_t) = E(\underline{\varepsilon}_t) = 0$

then we get from (3.11):

(3.19) $\underline{r}_t = E_t(\underline{r}_t) + \sum_j b_j[\underline{\Delta\theta}_{jt} - E_t(\underline{\Delta\theta}_{jt})] + \underline{\varepsilon}_t$

Eqs.(3.19) and (3.18) specify the general linear MFM in conditional expectations.

In contrast to (3.13), eq.(3.18) is not at all a trivial assumption and implies that the MFM is truly linear in the selected state variables. In that case, the error term does not contain higher order effects of the changes in the selected factors and can only account for the effects of possibly omitted common factors that are unrelated to (i.e. independent of) the included factors. This can be clarified by adopting the more appropriate framework of continuous time analysis.

Continuous time formulation

Although the discrete time exposition as presented above looks convincing, it rests on the calculus of deterministic variables. However, as the variables under consideration are stochastic, the use of stochastic calculus is appropriate. Doing so, we enter the continuous time framework, which in many cases is preferred over a discrete time setting because of its (relative) analytical elegance and simplicity.

Continuous time multi-factor models have frequently been used. For example, Cox, Ingersoll & Ross [1985b] consider multiple interest rates, and Breeden [1979] and Cox, Ingersoll & Ross [1985a] use multiple unspecified state variables. Treynor & Black [1976] and Turnbull [1977] assume that a firm's future cash flows are generated by economic variables and let the cash flow factor model propagate into a factor model of firm value.

In discrete time, the validity of the general linear MFM in conditional expectations (eq.(3.19)) crucially depends on the intrinsic linearity of the relationship between security prices (or returns) and (changes in) factors. In continuous time, when $\Delta t \rightarrow dt$, all relationships are (locally) linear. So even if $P(\cdot)$ is a non-linear function of $\boldsymbol{\theta}$, in the limit of continuous time (i.e. to order dt), their instantaneous contemporaneous changes will be perfectly correlated.[8] Furthermore, under stationarity, normality is implied by the central limit theorem, so no distributional assumptions are required.

The derivation of a multi-factor model in continuous time is contained in Appendix 3.A. The resulting model clearly shows how the conditional expected and the unanticipated components of changes in the state variables propagate into the respective components of the security return. Furthermore, it is in this case even more apparent that, although the price function $P(\cdot)$ may be non-linear, (local) linearity of the model is preserved by letting the sensitivities change continuously over time. Forcing the sensitivities to be constant over time introduces an error term in the factor model that contains the non-linear effects of the changes in the state variables. Furthermore, it implies that the sensitivities will depend on the (range in the) specific values that the state variables take on. Considering only a sub-set of the state variables (partial conditioning) adds the (linear and non-linear) effects of omitted factors to the error term. The effects of non-stationarities and omitted variables are further discussed in section 3.2.4.

Both the discrete time and the continuous time factor model appear to represent relationships whose parameters (given a sample of data) can be estimated using some statistical or econometric technique. In order to shed light on this issue, section 3.2.3 first explores the factor model that 'best' fits the actual relationships. Section 3.2.4 next discusses the differences between this model and the first order approximations as presented above.

[8] Cf. Merton [1982, p.32].

3.2.3 Conditioning and least squares

As a prelude to the discussion of some statistical aspects of MFMs, we consider in this section the conditioning of security returns on factors according to the least squares principle. Especially when we do not know the identity of all factors that possibly influence the security returns, what is a 'best' factor model and under what conditions is this model linear in the parameters?

We start from the general assumption that a security's return is generated by some unspecified true set of factors through some unspecified true function. We next consider the return as a function $\Phi(\cdot)$ of a (sub-) set of k factors $\{\delta_j\}_{j\in k}$ which we are interested in:

$$(2.1) \quad \underline{r}_t = \Phi(\underline{\boldsymbol{\delta}}_t) + \underline{\xi}_t$$

where $\underline{\boldsymbol{\delta}}_t = [\underline{\delta}_{jt}]_j$ is the vector with factor values and $\underline{\xi}_t$ is an additive (zero-mean) error term. This error term may account for the influence of various factors other than $\{\delta_j\}_{j\in k}$ (these factors can be of a common or more idiosyncratic nature). However, for this moment, we do not yet make any assumptions regarding the relationship between the error term and the k-vector $\underline{\boldsymbol{\delta}}_t$.

Now we want to find the function that 'fits best' in eq.(2.1). Of course, this depends on the criterion employed. The mean square error is a natural choice, so we are interested in the function $\Phi(\cdot)$ that forms the least squares approximation to (2.1). We assume that the relevant first and second moments of the variates exist and choose:

$$(3.20) \qquad \Phi(\cdot) = \Phi(\underline{\delta}_{1t}, \ldots, \underline{\delta}_{kt}) \quad | \quad E\{\underline{\xi}_t^2\} \text{ minimum}$$

This implies that we are looking for the function $\Phi(\cdot)$ that minimizes the mean square error:

$$(3.21) \qquad E\{[\underline{r}_t - \Phi(\cdot)]^2\}$$

According to the law of iterated expectations, we can write:

$$(3.22) \qquad E\{[\underline{r}_t - \Phi(\cdot)]^2\} = E\{\ E[\ (\underline{r}_t - \Phi(\cdot))^2 | \underline{\boldsymbol{\delta}}_t\ ]\ \}$$

Within the conditional expectation, $\boldsymbol{\delta}_t$ and hence $\Phi(\cdot)$ is treated as a constant. As for a constant c, the expression $E\{(\underline{r}_t - c)^2\}$ is minimal for $c = E\{\underline{r}_t\}$, we find that $\Phi(\cdot) = E\{\underline{r}_t | \boldsymbol{\delta}_t\}$ minimizes $E\{[\underline{r}_t - \Phi(\cdot)]^2 | \boldsymbol{\delta}_t\}$. Hence,

$$(3.23) \qquad \Phi(\cdot) = E\{\underline{r}_t | \underline{\boldsymbol{\delta}}_t\}$$

minimizes (3.21).[9] So, the expression

$$(3.24) \quad \underline{r}_t = E\{\underline{r}_t | \boldsymbol{\underline{\delta}}_t\} + \underline{\xi}_t$$

forms the least squares representation of (2.1). This in turn implies that:

$$(3.25) \quad E\{\underline{\xi}_t | \boldsymbol{\underline{\delta}}_t\} = 0$$

so that the error term is uncorrelated with (any transformation or function of) the factors.

When employing a mean square error criterion, relating the return to factors is optimal when the conditional return expectation is considered. Equivalently, from a least squares viewpoint, conditioning the return on factors through the function $\Phi(\cdot)$ has the maximum payoff when $\Phi(\cdot)$ equals the conditional expectation of the return, given $\boldsymbol{\underline{\delta}}_t$.

The rationale for conditioning security returns on state variables or factors, then, is that conditioning reduces return variability. This follows directly from the well-known expression:[10]

$$(3.26) \quad E\{Var(\underline{r}_t | \boldsymbol{\underline{\delta}}_{jt})\} = Var(\underline{r}_t) - Var(E\{\underline{r}_t | \boldsymbol{\underline{\delta}}_{jt}\}) \leq Var(\underline{r}_t)$$

According to eq.(3.25), the least squares criterion ensures that all information regarding the factors is used in the conditioning.

In practice, we would like to replace eq.(2.1) by a return generating process that is linear in the (selected) factors. This can now be defended on three grounds.

First, when the joint distribution of $\underline{r}_{it}$ and $\boldsymbol{\underline{\delta}}_t$ belongs to the elliptical (spherical) family, the conditional expectation eq.(3.23) is actually linear. Hence,

$$(3.27) \quad \underline{r}_t = a + b_1 \underline{\delta}_{1t} + \ldots + b_k \underline{\delta}_{kt} + \underline{\xi}_t$$

and the error term $\underline{\xi}_t$ satisfies eq.(3.25).[11] For a further discussion

---

9) Note that $E\{\cdot | \delta_{jt}\}$ is non-stochastic. The random variable $E\{\cdot | \underline{\delta}_{jt}\}$ is to be interpreted as the expectation, conditional to the σ-field relative to which $\underline{\delta}_{jt}$ is defined. See chapter two, footnote 9, and Spanos [1986, p.125] on this point.

10) See for example Mood, Graybill & Boes [1974, p.159], Maddala [1977, p.27] or Spanos [1986, p.126].

11) Only when $\underline{r}_{it}$ and $\boldsymbol{\underline{\delta}}_t$ are distributed jointly normal, however, the error term is homoskedastic: $E\{\underline{\xi}_t^2 | \boldsymbol{\underline{\delta}}_t\} = E\{\underline{\xi}_t^2\} = 0$. Also for symmetric stable (Paretian) distributions is the conditional expectation linear (provided the characteristic exponent is greater than one) but, except for the normal case, the second moment does not exist. This implies that we cannot use the mean square error criterion in theoretical expositions.

of this case, we refer to chapter two, section 2.2.2.C, where the linearity of the market model was discussed.

Second, we could simply **assume** that the conditional expectation (3.23) is a linear function.[12] A linear model can then be imposed on the actual (non-linear) return generating process by forcing a least squares regression:

$$(3.28) \qquad \underline{r}_t = a + b_1\underline{\delta}_{1t} + \ldots + b_k\underline{\delta}_{kt} + \underline{\varepsilon}_t$$

from which the parameters a and $\{b_j\}_j$ follow and where, by construction, the error term satisfies:

$$(3.29) \qquad E\{\underline{\varepsilon}_t \cdot \underline{\delta}_{jt}\} = 0 \qquad j=1,..,k$$

In this case, however,

$$(3.30) \qquad E\{\underline{\varepsilon}_t | \underline{\boldsymbol{\delta}}_t\} \neq 0$$

so that the error term is uncorrelated with, but not independent of, the factors.

Third, a linear model may be interpreted either as a linear least squares approximation to the unknown return generating function $\Phi(\cdot)$ (White [1980]), or as the best linear prediction of the return $\underline{r}_j$ given the factors $\{\underline{\delta}_{jt}\}_j$ (Van Praag [1981]). In this case, the eqs.(3.28-30) also apply. The next section discusses this issue in more detail.

3.2.4 Taylor series approximation versus least squares approximation

This section confronts the linear least squares approximation and the first order Taylor series approximation, as discussed in sections 3.2.2 and 3.2.3. Starting point is again the general assumption (2.1) that a security's return is generated by some function $\Phi(\cdot)$ of k factors $\underline{\boldsymbol{\delta}}$, where the additive zero mean and constant (finite) variance error term $\underline{\varepsilon}$ satisfies (3.1) and accounts for the influence of various other factors.

For reasons of simplicity, linear (or linearized) models are attractive. First, one does not have to search for an adequate form of the systematic return generating function $\Phi(\cdot)$. Second, statistical theory mainly considers linear models and there is a host of conventional econometric techniques that can be used for estimating parameters or making inferences from sample data. Third, as the portfolio operator is linear in discretely compounded returns, the sensitivities (return generating parameters) of individual securities can then easily be aggregated into portfolio sensitivities.[13]

---

[12] Cf. Fama [1976, p.78] in the context of the market model.

[13] See also Appendix 4.B to chapter four, and section 5.2 of chapter five on this point.

Two linear approximations

We can specify two linearized versions of eq.(2.1). For brevity, we will omit the time subscripts in the following. On one hand, we have a truncated first order Taylor series approximation around the expected values of the factors, as discussed in section 3.2.2:

$$(3.11') \quad \underline{r} = a' + \textstyle\sum_j b'_j \underline{\delta}_j + \underline{e}'$$

where the term $\underline{e}'$ reflects the approximation error as well as the true stochastic error. The sensitivities are the gradients (cf eq.(3.9)):

$$(3.31) \quad b'_j = \frac{\partial\Phi\{E(\underline{\boldsymbol{\delta}})\}}{\partial\underline{\delta}_j} \qquad j=1,..,k$$

where $\Phi\{E(\underline{\boldsymbol{\delta}})\}$ indicates that the partial derivatives are evaluated at the expected values of the factors.[14)]

On the other hand, we have the linear least squares approximation

$$(3.32) \quad \underline{r} = a + \textstyle\sum_j b_j \underline{\delta}_j + \underline{\varepsilon}$$

where the zero-mean error $\underline{\varepsilon}$ includes both the approximation error and the true stochastic error $\underline{\xi}$. In this case, the parameters a and $\{b_j\}_{j=1,..,k}$ are chosen to minimize $E(\underline{\varepsilon}^2)$. By standard statistical arguments, it follows that these parameters are the multiple regression coefficients and that

$$(3.33) \quad E(\underline{\varepsilon}\cdot\underline{\delta}_j) = 0 \qquad j=1,..,k$$

If the identities of the k factors are known, a linearized version of eq. (2.1) can be estimated by means of ordinary linear least squares (OLS) regression.

OLS slopes and gradients

It is then tempting to consider the estimated coefficients in eq.(3.32) as estimates of the parameters in the Taylor series expansion eq.(3.11').[15)] However, White [1980, p.152] notes that "OLS estimates do

---

14) Although the expected value is an obvious choice, the spanning point is more or less arbitrary, as remarked before.

15) Actually, this is done by Cramer [1971, pp.79-83], among others. Also Kennedy [1992, p.102] makes an ambiguous remark in this direction: "A nonlinear function can be restated, via a Taylor series expansion, as a polynomial. Estimating a linear function is in effect omitting the higher-order terms of this polynomial." For the implied 'omitted variables bias', we refer to the end of this section.

not necessarily provide reliable information about the local properties (derivatives, elasticities) of unknown functions." So it is very dangerous to equate least squares coefficients with Taylor series expansion coefficients. Hence, tests of hypotheses (concerning sign or magnitude of coefficients) based on Taylor approximation properties may be seriously misleading. The severe limitations of the Taylor series approximation interpretation for OLS has two grounds. First, the choice of the spanning point of the expansion is arbitrary; there is no guarantee that the expected value of the arguments is the 'best' spanning point. Second, OLS provides a global approximation of $\Phi(\cdot)$ whereas a truncated Taylor series only provides a local approximation.

The Taylor approximation goes **along** the curve of a function, while the OLS approximation goes **through** the function.[16] So there may be some point where the OLS slope estimate coincides with the tangency line at the spanning point. But as White [1980, p.150-151] remarks, locating this point requires knowledge of the unknown function; furthermore, in the multivariate case or for higher order approximations no such point need exist. White [1980, pp.152-154] derives conditions under which the OLS estimator provides information (asymptotically) about local properties of the unknown function (evaluated at the means of the regressors), and these conditions are very restrictive.[17] It follows that in the univariate case the extent of inconsistency between the first derivative and the OLS slope estimate depends on:

- the variance of the independent variable;
- the amount of skewness (as measured by the third central moment of the distribution) of the independent variable;
- the amount of either concavity or convexity of the unknown function over the range of values that the independent variable can take, or, when skewness is zero, the differences in either concavity or convexity.

The relationship between OLS coefficients and gradients

In order to explore the relationship between the first order Taylor series approximation (3.11′) and the linear least squares approximation (3.32), we relate the linearized least squares MFM (which is imposed on the return) to the non-linear general MFM (2.1) (which drives the return). For a detailed analysis, we refer to Appendix 3.B. With respect to White's [1980] analysis, our contribution is twofold. First, we adopt

[16] Note that the OLS approximation passes through the means $(E(\underline{\delta}_j), E\Phi(\cdot))$, whereas the Taylor approximation is tangent to the underlying function in the point $(E(\underline{\delta}_j), \Phi[E(\underline{\delta}_j)])$.

[17] For example, the unknown function is restricted to a subset of either the concave or the convex functions. Furthermore, the mere existence of a bound on the partial derivatives, given the OLS estimates, depends on characteristics of the inverse of the regressors' covariance matrix.

a different set of assumptions and second, assuming normal factors, we express least squares coefficients in terms of gradients.

We can summarize the results from Appendix 3.B as follows. When the return generating process is non-linear, we will make an approximation error by using a linearized MFM. This approximation error very likely generates discrepancies between the first order Taylor series coefficients b'_j in (3.31) and the linear least squares coefficients b_j in (3.32). The Taylor series coefficients equal the gradients evaluated at the spanning point, $\Phi'[E(\cdot)]$. Assuming joint normality for the factors $\{\delta_j\}_{j=1,..,k}$, we derive that the least squares coefficients represent the expected values of the gradients of the unknown function:

$$(3.34) \quad b_j = E\left[\frac{\partial \Phi(\cdot)}{\partial \underline{\delta}_j} \right] \qquad j=1,..,k$$

The discrepancies between the two sets of coefficients are due to non-linearities related to the third and higher order (partial) derivatives of $\Phi(\cdot)$, and (because of the assumed symmetry of the distribution) not to its second order derivatives. This implies that not concavity or convexity per se, but **differences** in concavity and convexity over the ranges of $\{\underline{\delta}_j\}_j$ are relevant. We therefore allow **mixed** convexity / concavity of $\Phi(\cdot)$. When the function $\Phi(\cdot)$ is characterized by either increasing convexity or decreasing concavity in the direction of $\underline{\delta}_j$, the corresponding least squares coefficient b_j exceeds the Taylor series coefficient b'_j. In case of decreasing convexity or increasing concavity we have the reverse. We show that the discrepancy between the two coefficients depends on the variance of the corresponding factor and the rate of change in convexity or concavity of $\Phi(\cdot)$, and derive bounds on the differences between the coefficients.

It is difficult to say exactly how the interpretation of the OLS estimates as the expected value of the gradients carries over to the situation where the factors are distributed only approximately normal, or where they just follow any symmetrical distribution.[18] However, the

[18] However, we note an unexpected and remarkable parallel between this interpretation of the regression coefficient under normality on the one hand and some robust regression estimators that are designed to cope with **deviations** from normality (especially outliers) on the other. Theil [1950] and Sen [1968], for example, developed estimators for the slope coefficient which are the median of the set of slopes $\Delta y/\Delta x$ joining pairs of points (x,y). The slopes $\Delta y/\Delta x$ can be imagined as gradients, so these estimators would resemble the median of gradients. Another example are Hinich & Talwar [1975], who propose the (trimmed) mean or median of slope coefficients from a large number of non-overlapping sub-sample OLS regressions. As these sub-samples are very small, the coefficient from each of these regressions may be considered as a local linear approximation to the underlying function (i.e. gradient), evaluated in the sub-sample mean. Hence, the (trimmed) mean or median of these coefficients can

conclusion concerning the relationship between the global least squares coefficients and the local Taylor coefficients generalizes to this less restrictive setting. Also for this case we derive bounds on their differences.

Usefulness of OLS approximation

In contrast to the very limited ability of least squares to provide information about partial derivatives or elasticities of unknown underlying functions, the least squares approximation is very useful from another viewpoint. White [1980, p.155] considers the minimization of the mean square error (MSE) of the approximation and proves that linear least squares estimates provide consistent (i.e. asymptotically correct) estimates of a well-defined weighted linear least squares approximation to the true but unknown function.[19] The distribution function of the independent variables hereby serves as weighting function. We have in our case that the least squares parameters a and $\{b_j\}_j$ are estimators for the parameters a^* and $\{b^*_j\}_j$ that satisfy:[20]

$$(3.35) \quad \min_{a^*, \{b^*_j\}_j} \quad E\{ [\Phi(\underline{\boldsymbol{\delta}}) - a^* - \textstyle\sum_j b^*_j \underline{\delta}_j + \underline{\varepsilon}]^2 \}$$

Given sample estimates for a^* and $\{b^*_j\}_j$, this implies that $a + \sum_j b_j \underline{\delta}_j$ is the best linear approximation (or prediction) of $\underline{r}$ by $\underline{\boldsymbol{\delta}}$. So the general usefulness of the least squares approximation is its optimality as a predictor, which holds "regardless of omitted variables, aggregation error, errors in variables, simultaneous equation error, non-additivity of the disturbance, or other forms of functional misspecification" (White [1980, p.163]).

White [1981] extends this result by showing that this interpretation is also valid for non-linear least squares estimates. Another important extension is by Van Praag [1980, p.5; 1981, p.142], who shows that linear least squares coefficients are the best linear predictors whether or not there exists an underlying functional model at all. In this 'model free' regression context we can even interprete the linear least squares estimates $\{b_j\}_j$ without assuming any underlying function $\Phi(\cdot)$ at all. However, at any time, the interpretation of least squares coefficients as gradients of an assumed underlying function must be mistrusted. This is especially relevant for section 3.5, where expansions of present value models will be used as linearized factor models, and for chapter four, where we will estimate factor models.

---

then be considered as the mean (or median) of the gradients.

19) Spanos [1986, p.459] also gives a clear exposition.

20) Cf. eq.(3.B.7) in Appendix 3.B. Note that the MSE can be decomposed in the MSE due to misspecification, $E\{[\Phi(\underline{\boldsymbol{\delta}})-a-\sum_j b_j \underline{\delta}_j]^2\}$, and the (true) error variance, $Var(\underline{\varepsilon})$.

Non-stationarity and omitted variables

The obvious source for non-stationarities in a factor model is a change of the function $\Phi(\cdot)$ over time. Indeed, the intertemporal stationarity assumption in section 3.1 that $\Phi_t(\cdot) = \Phi(\cdot)$ may be too heroic. Not only the functional form of $\Phi_t(\cdot)$ may change over time, also the set of arguments may change. For example, in one time period one set of state variables is (considered to be) relevant while in another period attention shifts to another set of state variables. Especially when the state variables are interrelated, a formerly relevant state variable may still pick up some influence of the now relevant factors. As a result, its estimated sensitivity is likely to change over time.[21]

Aside from this non-stationarity of the (non-linear) relationships, it is clear from the discussion above that there also exists non-stationarity of **linearized** relationships. This can best be seen in the continuous time factor model, where the sensitivities are explicitly time-subscripted. These sensitivities are gradients and change over time as the position on the curve of $\Phi(\cdot)$ changes. Hence, the response coefficients are only correct for small deviations of the factors. This also applies for the discrete-conditional case. The discrete time unconditional formulation of a factor model resembles a linear least squares regression equation. When applying OLS regression, however, we stressed that the estimated slope coefficients do provide only very limited information on the gradients in some spanning point like the sample mean. Not only when the (mean) levels of the factors change, but also when their variability changes over time, this implies that the sensitivities for these factors may change.

Since a non-linear function can be expanded in a Taylor series, one can take the view that estimating a linear function boils down to omitting the higher-order terms in this expansion (cf. Kennedy [1992, p.102]). This would introduce an 'omitted variables bias' in the estimated coefficients, but we have dealt with this case above. However, when there is only partial conditioning (so only a sub-set of relevant state variables is included in the factor model), we are most explicitly confronted with the omitted variables bias. Spanos [1986, pp.420-421] criticizes the textbook discussion[22] of the omitted variables bias problem. The conventional argument incorrectly combines two statistical models with different parametrizations. There is a difference in parametrization of the models, since the coefficients in the 'complete' model and the 'incomplete' model depend on different sample information. The bias does then arise when one tries to estimate some coefficients in the true 'complete' model by estimating these coefficients in the

[21] This is a general problem when proxy variables are used for the true underlying factors (cf. section 4.1). When the relationship between the proxy and the true factor changes over time, the sensitivity for the proxy variable changes over time.

[22] See for example Maddala [1977, pp.155-156], Johnston [1984, pp.260-261] or Kennedy [1992, pp.91-92].

'incomplete' model. However, there arises no omitted variables problem when estimating coefficients in the incomplete model and considering these as just the coefficients for the incomplete model. Of course, insofar an omitted factor is correlated with some incorporated factor(s), part of the influence of this omitted factor is accounted for by the coefficients of the incorporated factors. When the relationship with the omitted factor changes over time, also the coefficients of the incorporated factors will change over time. But this is just an example of general non-stationarity and not due to the omitted variables bias.

3.3 FACTOR MODELS AS SIMPLIFYING DISTRIBUTION ASSUMPTIONS

In this section, we consider security returns as interrelated random variables. We argue that a factor model representation of these returns can be used to derive an essential input for the portfolio selection problem, viz. information concerning the joint probability distribution of future portfolio returns. Section 3.3.1 indicates the complexity of assessing distributions of portfolio returns. In section 3.3.2, we show how conditioning can reduce this complexity, both on the level of a particular portfolio and on the level of individual securities that can be combined into a multiplicity of portfolios. In addition, we show how conditioning is related to sensitivity analysis and probabilistic modelling approaches and we provide some more examples of the use of conditioning. Finally, in section 3.3.3, the general ideas are concretisized to the problem of obtaining probabilistic information on the portfolio level. This is accentuated to multi-moment portfolio analysis.

3.3.1 Distributions of securities and portfolios

In the most general terms, the investment problem entails the confrontation of an investor's preference structure and the relevant attributes of the investment. The monetary attributes of an investment relate to the probability distribution of its return over a specified future horizon. In portfolio theory, it is generally assumed that risky investment alternatives can be fully characterized by the (subjective) distribution of their outcomes: the single dimensional returns. Consequently, two investment alternatives with the same probability distribution[23)] will be perfect substitutes. Starting from the statistical interdependency of security returns, the general diversification theorems of Samuelson [1967], Brumelle [1974] and

23) More accurate: with the same position in the joint return distribution.

MacMinn [1984] imply that an investor will in general combine several individual securities into a portfolio.[24] Investment decisions can thus be viewed as choices among alternative probability distributions of portfolio returns (the optimal choice being determined by the specific preference functional of the investor).

Portfolio selection models then require the assessment of the probability distributions of portfolio returns. This is complex for two reasons. First, the direct assessment of a probability distribution of future returns is difficult in its own right. Second, an investor does not want to appraise the distribution of some given portfolio; instead he is interested what probability distributions can emerge from blending individual securities into a portfolio. As a portfolio is a convex combination of individual securities, the distribution of a portfolio's return is a convolution of the distributions of the returns on the individual securities that are included in the portfolio. Even if the (marginal) distributions of all the individual securities in the opportunity set are known, we have a problem in deriving portfolio return distributions. Because of the statistical interdependency of security returns (the mechanism that drives diversification), their joint distribution is not simply the product of their marginal distributions. This makes the computation of the convolution of their distributions theoretically very complicated, if not practically impossible. This renders the direct assessment of their joint distribution infeasible.

Focussing on portfolios[25], the central question is thus how the distributions of portfolio returns can be assessed, or, equivalently, how the distribution of a (linear) combination of returns can be derived from the joint distribution of these returns. A means by which the joint probability distribution of the returns could be simplified and that has the additional feature that the probability distribution of a linear combination of the variables can be adequately represented, would thus

---

[24] Samuelson [1967] and Brumelle [1974] show that non-negative interdependence (which is more general than linear interdependence or correlation that underlies Markowitz' [1952] diversification) of security returns yields an incentive to diversify. MacMinn [1984] extends these results by also considering positive interdependence.

[25] In effect, we do not need diversification theorems in order to defend the focus on portfolios. We only have to observe practice, where we see that investors combine securities into portfolios. This observation was also the starting point of Markowitz' [1952, p.78] analysis, who was able to **explain** this incentive to diversify: "Diversification is both observed and sensible; a rule of behavior which does not imply the superiority of diversification must be rejected both as a hypothesis and as a maxim." The directive to diversify can then be interpreted as a conditional-normative complement to this positive component of his model. In this sense, the contribution of Markowitz' analysis is **not** the much emphasized advice **that** investors should diversify (because they already do). The contribution is that Markowitz shows **how** investors can diversify their portfolios adequately.

be welcome.

3.3.2 Conditioning and the reduction of complexity

The reduction of complexity of the beforementioned assessment problem is discussed on two levels: (A) for one particular portfolio, and (B) for a (feasible) set of multiple portfolios that can be formed from the individual securities in the opportunity set. In addition, the link with sensitivity analysis and probabilistic modelling approaches is discussed (C) and some examples on the use of conditioning are mentioned (D).

3.3.2.A The portfolio level

In order to evaluate a portfolio, an investor must assess its (subjective) return distribution. The direct assessment of a portfolio's distribution may be problematic, particularly because return distributions depend on the (economic) context in which the returns are generated. In order to simplify (subjective) probability assessments and to improve their consistency, Ravinder, Kleinmuntz & Dyer [1988] developed a general framework that relies on **decomposition**. Decomposition is a technique for reducing the complexity of a judgment problem by dividing it into a set of smaller (and presumably easier) judgments. Specifically, Ravinder, Kleinmuntz & Dyer [1988, p.187] assume that there exist "background events" that describe the relevant context of the "target event". Their decomposition approach then captures the relationship between these events in the form of conditional probabilities of the target event. These conditional distributions can be aggregated by combining them with the (marginal) distributions of the background events.[26)]

Transposed to the portfolio problem, the portfolio return $\underline{r}_p$ (the target event) is conditioned on economic events (the background events). We can take factors $\{\underline{\delta}_j\}_{j=1,..,k}$ as representative for the relevant economic events and express the relationship between portfolio returns and factors in period t as:

$$\underline{r}_{pt} = \Phi_t(\underline{\delta}_{jt}) \qquad (3.36)$$

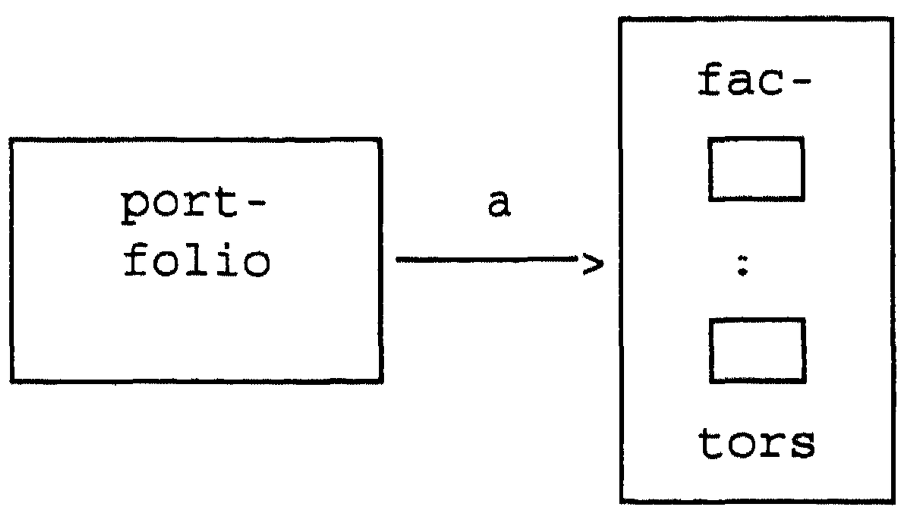
[26)] Ravinder, Kleinmuntz & Dyer [1988] employ a discrete probability framework. The emphasis of their paper is on the measurement error that can be attributed to the use of decomposition when compared to direct assessment. We here only refer to the conceptual aspects of their approach. The general idea of decomposition for the resolution or reduction of complexity can be traced to Simon [1962]. Decomposition is mainly applied as a method of numerical calculation or as an instrument of economic analysis. An early example is Baumol & Fabian [1964]. For a recent and extensive treatment, we refer to Goedhart [1994].

The function $\Phi(\cdot)$ then represents the process of conditioning and forms the link between the distributions of the factors and the portfolio return. The underlying decomposition hypothesis is that there exist a number of factors that generate the returns. When more than one factor is introduced, the dimensionality of the portfolio return increases. Reduction of complexity is achieved when both the link between returns and factors is clear (for generating conditional return distributions) and the investor has information concerning the (marginal) distributions of the factors. The procedure is depicted in Figure 3.3.A.

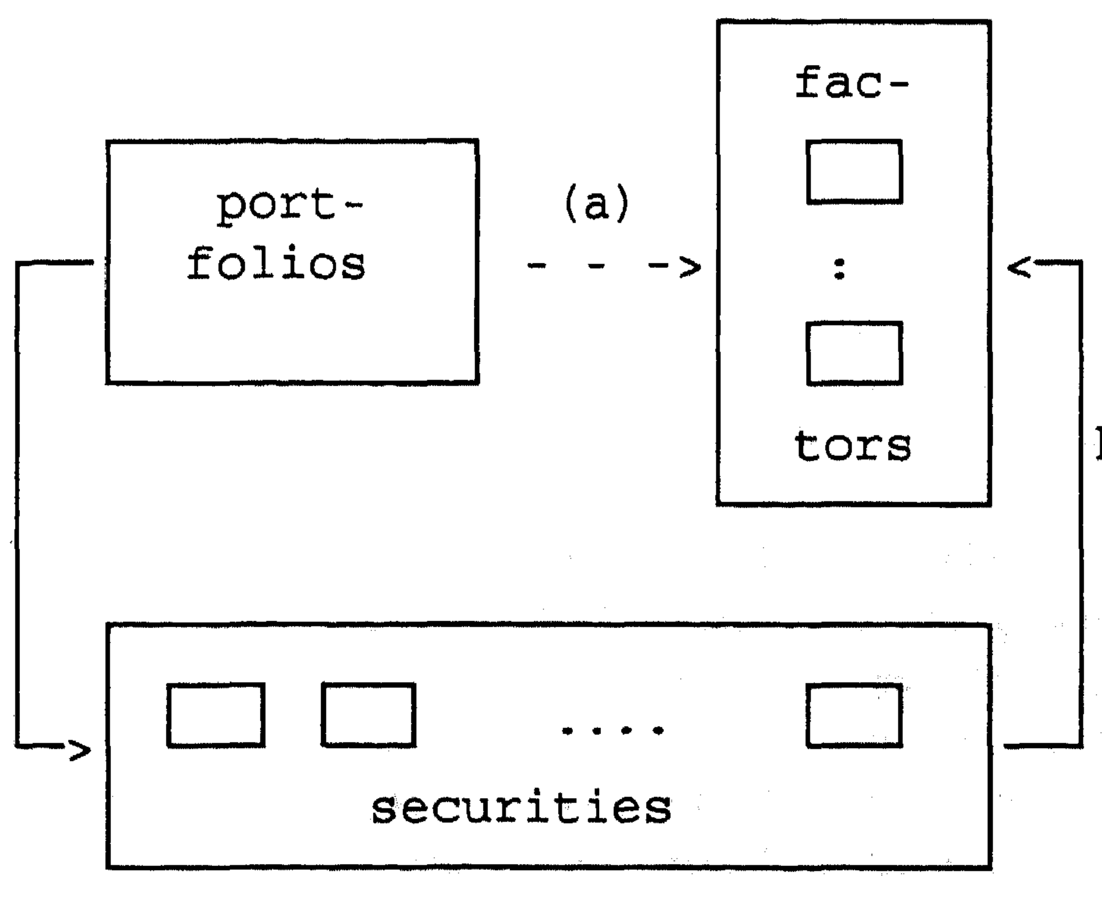
The reduction of complexity is certainly achieved when not considering the return on one particular portfolio, but instead the returns on a (feasible) **set** of portfolios. Although in the case of k background events the dimensionality of a portfolio return increases from one to k, the dimensionality of the returns on a set of P >> k portfolios remains k. This is elaborated further in the next section.

Figure 3.3: Conditioning reduces complexity. The arrows indicate conditioning. For generating, the direction of the arrows should be reversed.

Panel A: Decomposition and the assessment of a portfolio's return distribution (a)



Panel B: Hierarchical structures and the assessment of the securities' joint return distribution (b)



3.3.2.B **The securities' level**

Conditioning of a portfolio's return on factors can reduce the complexity of the assessment of its distribution. As mentioned above, however, another complexity is archetypical for the portfolio selection problem. An investor does not deal with one portfolio, as a given combination of securities, but is confronted with an unlimited number of feasible portfolios that can be constructed from the securities in the opportunity set. In order to choose between these portfolios, an investor must assess their (subjective) return distributions. So an investor is not so much interested in the distribution of one particular portfolio, but more in the joint distributions of the individual securities that generate the whole set of feasible portfolios. The complexity we encounter here is the interdependence of security returns that obstructs the assessment of their joint distribution. For the sake of tractability, the joint distribution of the securities needs to be simplified. One way to reduce the dimensionality of the securities' joint return distribution is conditioning.

To simplify the assessment of joint distributions of some interdependent random variables $\{\underline{x}\}$, Raiffa [1968, pp.260-261] discusses **hierarchical structures**.[27] The procedure rests on the assumptions that:

- the variates $\{\underline{x}\}$ are interdependent because of their common dependence on some (smaller number of) underlying variables $\{\underline{y}\}$;
- the variates $\{\underline{x}\}$ are mutually independent, conditional on the underlying variates $\{\underline{y}\}$ taking any specific values $\{y\}$.

When the underlying variates $\{\underline{y}\}$ are independent, we can assess $\{\underline{x}\}$ for each of the assumed values of $\{\underline{y}\}$. When the underlying variates $\{\underline{y}\}$ in turn are interdependent, we can add a level to the hierarchy and consider (an again smaller number of) variates $\{\underline{z}\}$ which render $\{\underline{y}\}$ conditionally independent.[28]

Transposed to the investment problem[29], we can assume that:

---

27) These hierarchies must not be confounded with the hierarchies that are used in indicating the relative importance of various selection **criteria** in a decision problem. For an exposition of the latter hierarchy approach in portfolio selection, see for example Saaty, Rogers & Pell [1980].

28) "And so it goes. A little science and a bag of tricks" (Raiffa [1968, p.261]). Science, or better: economic intuition, enters in finding meaningful underlying variables on which the original variables can be conditioned. This underlies the decomposition of portfolio returns as discussed above. See also sections 3.5 and 4.1.

29) Raiffa [1968, p.261] gives a short example in which security returns are related to indices that represent homogeneous classes, which in turn are related to underlying economic factors. Comparable, but more extensive, is Rudd & Rosenberg [1980] who first relate stock returns to sector indices and then specify a factor model for each of these indices.

- security returns $\{\underline{r}_i\}_{i\in N}$ are interdependent because of their common dependence on some underlying factors $\{\underline{\delta}_j\}_{j\in k}$ (with k<<N);
- security returns $\{\underline{r}_i\}_i$ are mutually independent, conditional on the underlying common factors $\{\underline{\delta}_j\}_j$ taking any specific values.

When we in addition assume that:

- the conditioning of the security returns is taken with respect to the distributions of all possible values that the underlying common factors can take on; and
- the process of conditioning is linear,

then we can write:

$$\underline{r}_{it} = a_i + \textstyle\sum_j b_{ij}\underline{\delta}_{jt} + \underline{\xi}_{it} \tag{3.37}$$

with:

$$E(\underline{\xi}_{it}) = E(\underline{\xi}_{it}|\underline{\delta}_{jt}) = 0 \qquad \forall\ i,j \tag{3.38}$$

and where the common factors $\{\underline{\delta}_{jt}\}_j$ and the errors $\{\underline{\xi}_{it}\}_i$ are independent. The error term reflects the fact that, given some values for the common factors, we cannot know the security return with certainty. More specific: the disturbance term incorporates the influence of idiosyncratic factors. The conditioning is exhaustive in the sense that all factors, common to more than one security, are accounted for.

In eq.(3.37), we recognize the linear, strict multi-factor model from section 2.4. Note, however, that (3.37) is a more restrictive representation than eq.(2.35), because now both the errors and the factors are not 'just' uncorrelated, but independent.[30]

The hierarchical structure implies conditioning, by means of which the returns are linked with a set of (exogeneous) factors. From the complementary point of view, the returns are generated by the factors, and the variability of these factors gives rise to the variability of the returns. The link between the return and each of the factors is explicitized by means of the sensitivity coefficients $\{b_{ij}\}$. Hence, these coefficients determine the extent to which the characteristics of the factors' distributions are passed through to the distributions of the security returns. This is depicted in Figure 3.4.A.[31]

---

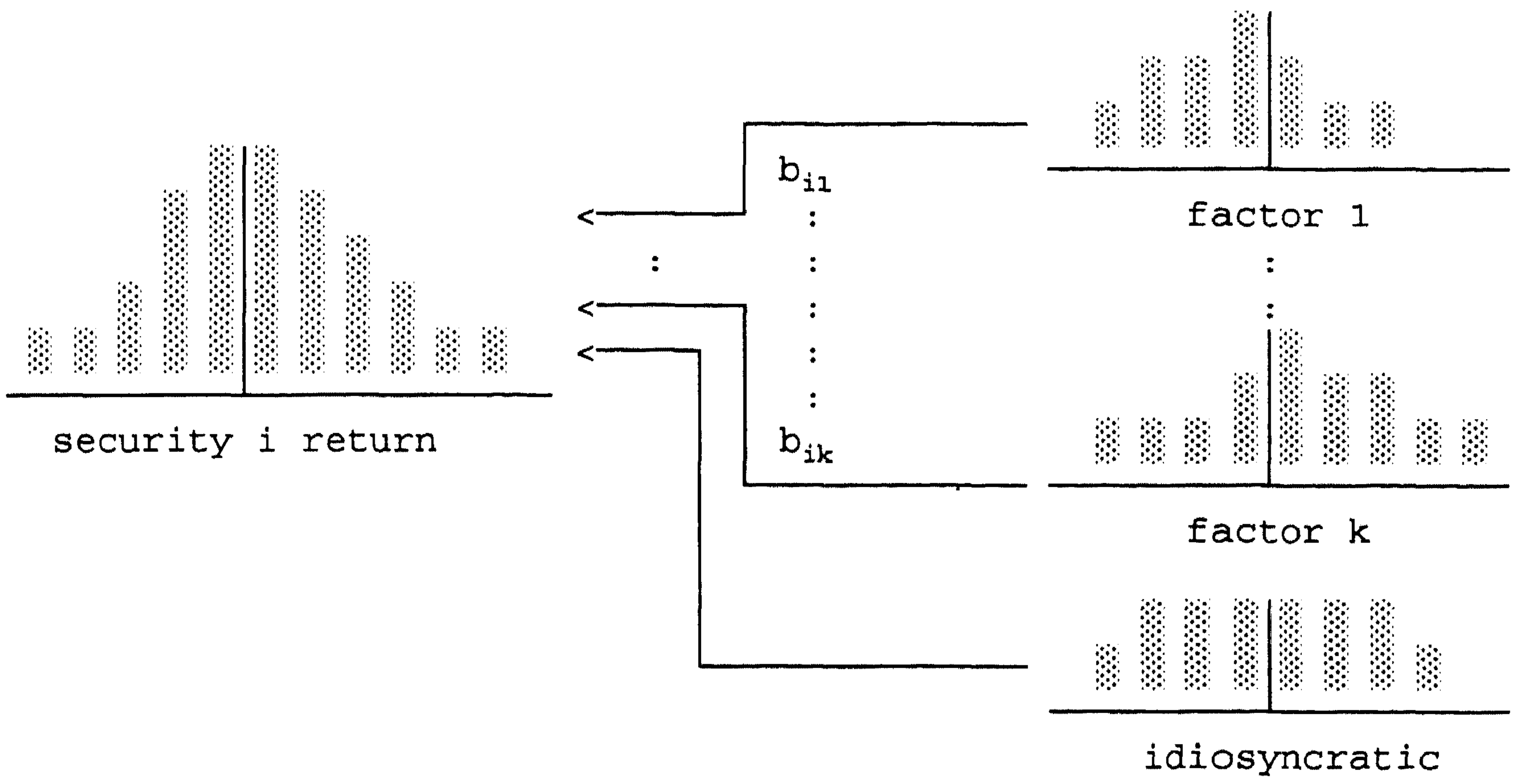
[30] A set of factors can always be orthogonalized. The factors can be made independent, in contrast, **only** when their joint probability density function $f(\delta_1,..,\delta_k)$ can be factorized into a product of functions of the individual factors alone (i.e. their marginal density functions): $f(\delta_1)\cdot\ldots\cdot f(\delta_k)$. This directly follows from the definition of statistical independence.

Also, in a regression context, we have the orthogonality $E(\underline{\xi}_{it}\cdot\underline{\delta}_{jt}) = 0$ by construction, but there is no mechanism to guarantee the stronger semi-independence $E(\underline{\xi}_{it}|\underline{\delta}_{jt}) = 0$.

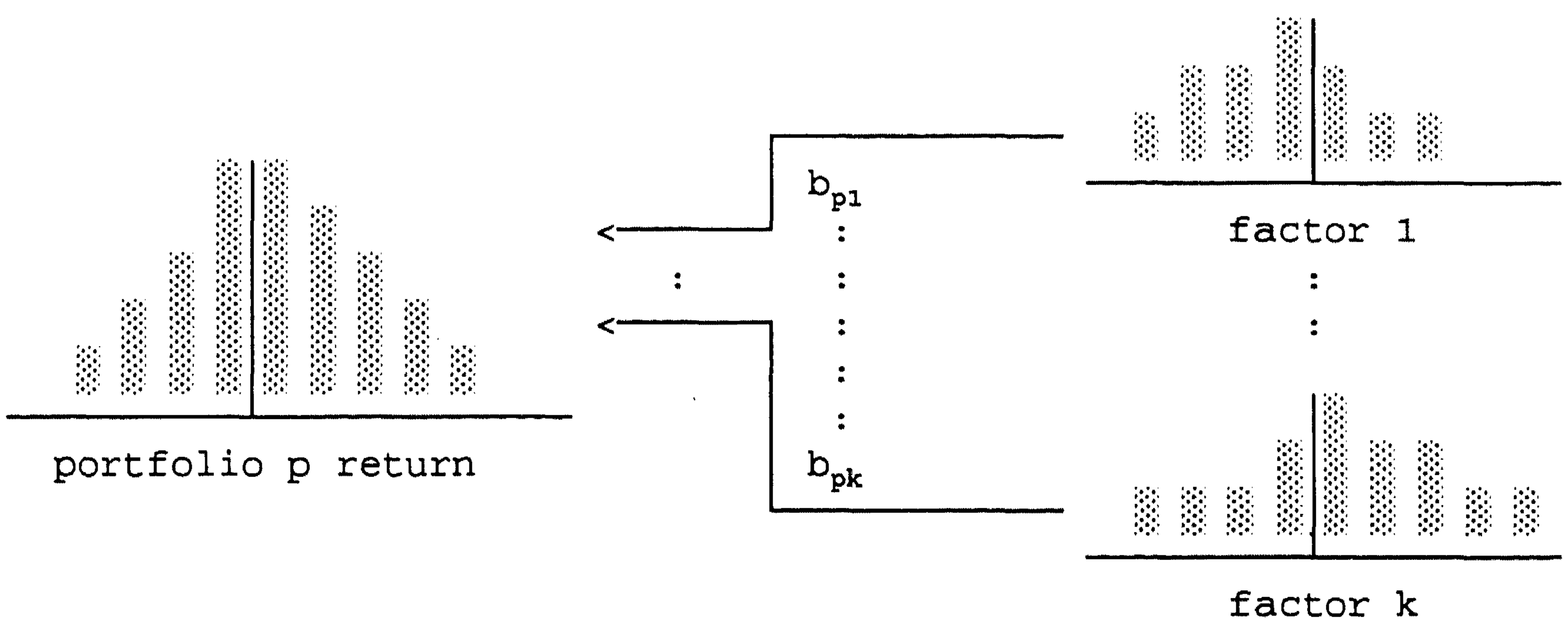
[31] Sharpe [1985, pp.184,189,190] presents comparable figures for the mean-variance framework.

Figure 3.4: Conditioning returns on factors.

Panel A: Distributions of individual security returns and factors.



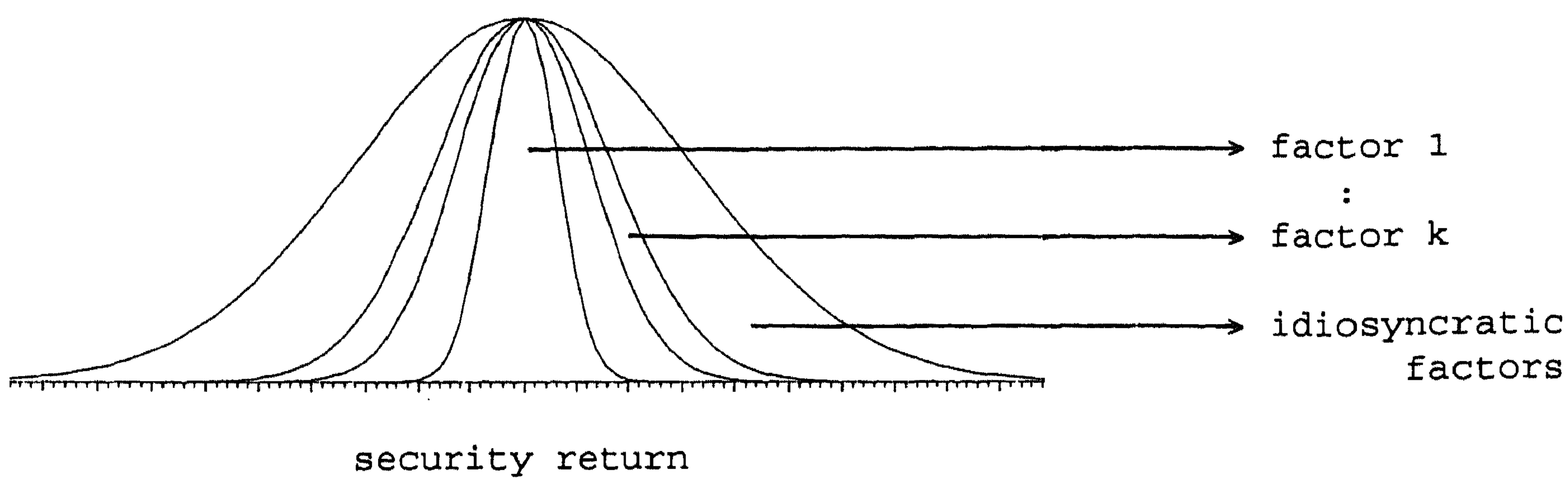
Panel B: Distributions of (well diversified) portfolio returns and factors.



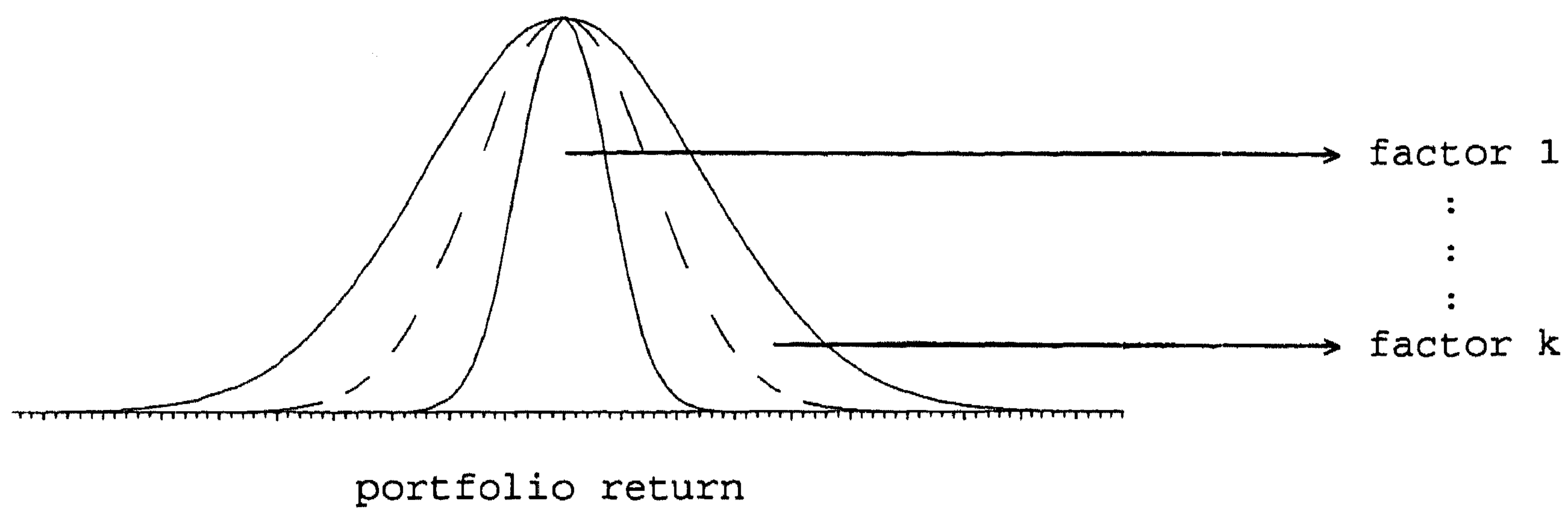
Because of the independence assumptions that are here imposed on the MFM, the return on a security can be decomposed into two independent parts. The first part is the systematic or factor-related return component, which in turn may be the sum of independent parts, each of which is generated by a different factor (this is the decomposition hypothesis on the portfolio level, as discussed above). The second part is the truly idiosyncratic return component. This decomposition hypothesis on the security level imposes a strong dichotomy between the common and idiosyncratic parts of a security's return. When combining securities into a portfolio, their factor sensitivities $\{b_{ij}\}_{i,j}$ are averaged in a linear fashion into the portfolio sensitivities $\{b_{pj}\}_j$.

Figure 3.5: Conditioning returns on factors: an alternative representation.

Panel A: The distributions of the factors constitute a security's return distribution.



Panel B: The distributions of the factors constitute a (well-diversified) portfolio's return distribution.



(For this reason we assumed the conditioning to be linear.) When the portfolio is large and not extremely weighted towards a limited number of securities, the law of large numbers can be invoked and the distribution of the idiosyncratic return components degenerates. Hence, the portfolio becomes 'well diversified' with respect to the factors. So, when aggregating the security return distributions into a portfolio return distribution, we need only be concerned with the (independent) distributions of the limited number of factors and the portfolio's sensitivities for these factors. Given the factor distributions, the factor sensitivities are in this sense the attributes that characterize the distributions of portfolio returns. This is depicted in Figures 3.3.B and 3.4.B. Likewise, in a portfolio context, the factor sensitivities are the attributes that characterize the joint distribution of security returns.

Figure 3.5 contains an alternative representation of the conditioning.[32] It is explicitly shown how the various factors constitute the return distributions of both individual securities and well diversified portfolios.

Summarizing so far, we conclude that a MFM can be used as simplifying distribution assumption in the sense that portfolio return distributions can be adequately approximated from information about the factor representations of the individual securities. The precise nature of this assumption is that a k-factor model involves a restriction on the joint distribution of the N security returns. On the security level, we have a decomposition hypothesis in the sense of a dichotomy between factor-related and idiosyncratic return parts. In this way, the dimensionality of the interrelationships between individual security returns is **reduced** from N to k dimensions, thus reducing the complexity of probability assessments of portfolios. On the portfolio level, we have (for k>1) a decomposition hypothesis in the sense that a one-dimensional portfolio return can be **extended** towards the k dimensions of the factors. Considering the possible returns on a (feasible) set of a multitude of P portfolios that can be formed from the individual securities, the dimensionality of the portfolio returns of P >> k portfolios remains k.

3.3.2.C Sensitivity analysis and probabilistic modelling

To put the conditioning that underlies the decomposition approach in a broader perspective, we briefly discuss the link with sensitivity analysis and probabilistic modelling. We then provide examples from the literature in which conditioning is suggested in an implicit or explicit way to gather probabilistic information.[33]

Sensitivity analysis

The decomposition approach, as described above, is in a sense related to sensitivity analysis. In sensitivity analysis, a link is established between one or more output variables and several underlying input variables. Using this model, one can analyze how changes in the values

---

[32] Harrington [1987, p.124] presents a figure comparable to 3.5.A for the representation of systematic and unsystematic risk in the market model.

[33] These examples either present analytical results, or rely on simulation analysis. For the sake of completeness, we note that not only simulation analysis can be used to shape the process of conditioning, but also that conditioning can be used to enhance inferences from simulation analysis. In that respect, we mention the role of conditioning as a technique for reducing the variance of estimators obtained from simulation studies (see for example Ross [1985, pp.479-482]). Cf. our section 3.2.3, especialy eq.(3.26).

of the input variables (or errors in these values) affect the output variable(s). This 'what if' approach to the analysis of uncertainty has a long tradition in the field of capital budgeting. As the values of input variables are uncertain and subject to estimation error, one is interested in the effect of relaxing assumptions on the profitability of an investment project.[34] Especially the uncertainty about the project's cash flows is recognized.

Of course, sensitivity analysis copes with uncertainty only in a very stylized way. Generally, in addition to the most likely value (or 'base case' value), either an 'optimistic' and a 'pessimistic' value for an input variable is chosen, or various values in some interval are considered. Given this range in each input variable, the effect on the output variable can be analyzed in isolation, or a composite estimation error can be found by aggregating the errors in the inputs. In both cases, it is more or less implicitly assumed that the values of the input variables are uniformly distributed. This is most clear when the output variable is plotted against each of the input variables.[35]

In contrast to conventional 'deterministic' sensitivity analysis, the 'stochastic sensitivity analysis', as proposed by Eschenbach & Gimpel [1990], starts from particular distributions of the input variables. For each input variable, the output variable is plotted against the input variable's cumulative distribution function. When the output variable monotonically increases (or decreases) with the input variable, the resulting curve represents the output variable's distribution function, conditional on the other variables' base case values. By combining the graphs of all input variables, a 'stochastic spider plot' is obtained. This technique reveals the impact of individual input variables on the output variable and indicates the input variables' relative importance. Also, because probability information is included, it much resembles the probabilistic approach as discussed below. However, although the output variable's uncertainty is connected to uncertainty in the input variables and although the information concerning different input variables can be presented in a single graph, the analysis remains partial (i.e. 'ceteris paribus' the values of the other input variables).

Instead of analyzing the effect of uncertainty about an input variable in isolation, a composite estimation error can be analyzed. Yoon [1990] aggregates individual estimation errors according to the 'law of propagation of errors'[36]; when the ranges of the input

---

[34] An early reference is Solomon [1966]. See also the discussion of Joy & Bradley [1973,1978] and Smith [1978b], and the closely related analyses of Whisler [1976] and Hsiao & Smith [1978].

[35] For some guidelines for constructing well-designed graphs, see for example Eschenbach & McKeague [1989].

[36] Given some variables $x_1,..,x_n$, measured with (uncorrelated) random errors, and some function $\underline{y} = f(\underline{x}_1,..,\underline{x}_n)$, the errors in the x_i propagate into y by:

variables around their expected values are assumed to be proportional to their standard deviations, some statistical flavor is added to sensitivity analysis. However, the information is limited to a range of possible values of the output variable, and not its distribution.

Probabilistic modelling

Still, it is a small step from here to the 'probabilistic approach' of Hillier [1963, 1965]. Probabilistic approaches in general aim to provide the probability distribution of the internal rate of return, the present value of a project or some other investment decision criterion. Hillier [1963, 1965] provides an analytical framework which rests on the assumption that cash flows emanate from different sources. This allows him to decompose the cash flows into uncorrelated (independent) and perfectly correlated parts, so that their covariances can easily be handled. Because of the normality assumption (or at least because of the assumption that only means and variances matter), the probabilistic information is quite limited. Also, as non-linearities would destroy the (joint) normality of the input variables, the application of Hillier's technique is limited to linear relationships between input and output variables.[37)]

The decomposition approach can in fact directly be recognized in the more sophisticated 'probabilistic approach' of Hertz [1964, 1968]. Summarized (Hertz [1968, p.99]), the analysis starts with the identification of key factors that make up the cash flows. Next, (subjective) probability distributions are specified for these factors. These "uncertainty profiles" are then combined into a "risk profile": the probability distribution of the particular investment criterion. The combination of the factors' distributions into the criterion's distribution is achieved by Monte Carlo simulation (via repeated sampling), taking into account the interdependencies that may exist between the factors. The output of the whole process is a probability

---

$$\mathrm{Var}(\underline{y}) \sim \textstyle\sum_i (\partial f/\partial x_i)^2 \mathrm{Var}(\underline{x}_i)$$

This readily follows from a first order Taylor series expansion of $f(\cdot)$. Compare with eq. (3.A.8) from Appendix 3.A.

37) Hillier's analytical approach is extended for practical implementation, for example, by Wagle [1967], who assumes that cash flows components are generated by different factors; by Bussey & Stevens [1972] and Giaccotto [1990], who provide general models for cross- and auto-correlations between cash flows; and by Zinn, Lesso & Motazed [1977] who also consider semi-variance (but do not decompose the cash flows).

distribution of the relevant investment criterion.[38)]

In this place, outweighing the accuracy of an analytical approach against the flexibility and simplicity of a simulation approach is not relevant. Important is the conceptual point that decomposition by conditioning can simplify the assessment of probabilistic information. Applications are general, but the principle remains the same. In particular, Hertz' [1964, p.102] Exhibit II closely resembles our Figure 3.5.B.[39)]

3.3.2.D Other examples of conditioning

In addition to those mentioned above, there are several other studies following the lines of Hillier [1963] and Hertz [1964]. Most of these studies, however, are confined to the mean-variance framework.

Cohen & Elton [1967], for example, simulate joint returns in order to calculate the variance-covariance matrix as an input for a (mixed real and financial investment) mean-variance portfolio selection model. In their approach, they first condition future cash flows on factors. Then they specify separate distributions for each underlying factor and the interrelationships among the factors. Next they simulate the (period by period) the values of the factors and infer the values of the corresponding cash flows. Discounting these cash flows finally yields present values, for which the covariance matrix can be estimated.

Fried [1970], following an analytical route, developed a formal method for obtaining *a priori* distributions for mean-variance portfolio models. For this, he specifies 'economic structures' in which a future return is decomposed in two independent parts: a deterministic part that is generated by fixed and known independent variables (for example, analysts' forecasts) and a random part. In this way, he decomposes the **expected** security returns, which are allowed to change over time. The risk is obtained from the covariance matrix of the forecast errors.

In essence, chapter two contains many examples of conditioning security returns. The first hierarchical structure in portfolio theory, designed to simplify the representation of return distributions, is of course the single index model or single factor model (Markowitz [1959, pp.96-101], Sharpe [1963]), discussed in section 2.3. In the familiar mean-variance

---

[38)] For the evaluation of this distribution, information about the decision maker's preferences is needed. Hertz [1968, pp.103-104] suggests a mean-variance efficiency criterion, implying a quite limited use of the full probability information.

[39)] There are several criticisms raised against the probabilistic approaches of Hillier [1963] and Hertz [1964]. These criticisms, however, do not concern the underlying decomposition framework, but aim at the use of a single certainty equivalent factor to adjust for the risk of a cash flow stream (Keeley & Westerfield [1972]), or at the use of a stochastic internal rate of return (Ekern [1979]).

applications of this model and its multivariate extensions (discussed in section 2.4), the independence assumptions can be replaced by the less restrictive linear independence assumptions.

A small number of studies, however, indicate the applicability of conditioning for obtaining **full** probability information. In presenting a Bayesian approach to assess return probability distributions, Winkler [1972, p.144; 1973, pp.399-401] suggests conditioning of the distribution's parameters on different sets of economic conditions. Or, "[a]nother possible way to arrive at a simplified, tractable model is to assume a functional relationship among [security returns and economic] variables and to deal with the parameters of this functional relationship" (Winkler [1973, p.402; 1972, p.145]). This alternative explicitly refers to the use of factor models.

Another example is Grauer [1981], who investigates the so-called 'growth optimal' investment policy (i.e. maximize expected logarithmic utility, or maximize the geometric mean return). For the operational feasibility of the involved specific optimization, he uses the single index model to generate (discrete) probability distributions. He suggests the use of the model for generating the probability inputs to expected utility formulated investment problems in general.

When considering the probability distribution of a portfolio that combines many securities and options on these securities, the complexity of the assessment problem becomes even greater. Bookstaber & Clarke [1981, 1983] analyze the problem of deriving return distributions for optioned stock portfolios. As each option position can be translated into 'share equivalents' through the use of option hedge ratios, the effect of an option position on a portfolio can easily be determined. However, the hedge ratio, and hence the share equivalent, of each option will change whenever the price of the underlying stock changes. So this procedure is only valid in continuous time, where the option positions can be revised continuously over infinitesimally intervals. Over some finite horizon, the various stocks in the portfolio can take on any of a number of values and there will be an infinite number of combinations of returns on the individual stocks that yield one specific return on the stock portfolio. This implies that, for a given return on the underlying stock portfolio, an unlimited number of returns on the optioned portfolio are possible. But at the same time, there are an unlimited number of returns possible on the underlying stock portfolio itself. So, for computing the return distribution of an optioned portfolio, one must consider the share equivalent of each option for every possible value of the underlying stock, conditional on the returns on every other stock in the portfolio. As the analytical solution to this problem is very complex, Bookstaber & Clarke [1983] consider the expected value of the optioned portfolio conditional on the returns on the stock portfolio as an approximation. For deriving the returns on the stock portfolio in their simulations, they use in turn the single index model.

In the context of the investment decision, conditioning in the form of MFMs can reduce two essential problems: the information problem and the combination problem (how to assess the relevant probability characteristics on the level of individual securities and on the portfolio level; see Figure 3.2). Above, we have mainly considered full probability information. However, aside from stochastic dominance rules and growth optimal investment policies, there are no portfolio selection criteria in which the full probability distribution of portfolio returns is used in an explicit way. Instead, in most cases **summary measures** of the distributions are used. This can be motivated either by the argument that only some summary measures are relevant for an exact representation of distributions, or because one is satisfied with only an approximate representation of the full probability distribution. In both cases, we are left with the problem how to account for this probability information in a tractable way. For this reason, we elaborate in section 3.3.3 on the role of MFMs as simplifying distribution assumptions and illustrate their use in the context of multi-moment portfolio analysis. In this context, the factor approach will prove to be indispensable.

3.3.3 (Multi-) factor models and multi-moment portfolio analysis

Factor models, moments and distributions

Instead of considering the full probability distribution of the returns, we could alternatively focus on summarizing characteristics of the distribution, either as an exact representation or as an approximation. When the security returns come from a distribution family that is known and stable, the parameters of the distribution can be used. Alternatively, when they exist, the statistical moments can be used to characterize distributions.[40)]

This opens two possible routes to the use of summary measures (moments). The first is the most general and considers only an **indirect** use of the statistical moments in order to obtain full probability distribution information. The procedure is to represent a true but unknown underlying probability distribution in terms of an alternative (approximating) distribution by using its moments in an asymptotic series expansion. In this context, Kendall & Stuart [1969, pp.156-167] discuss the Gram-Charlier, the Edgeworth and the Cornish & Fisher series expansions, which use the standard normal distribution as alternative distribution. Next, only a limited number of terms can be considered in

---

40) The distinction between moments and parameters is that the parameters of a distribution appear explicitly in the mathematical specification of the distribution. For the normal distribution, the first two moments also happen to be its parameters. So, referring to (E,σ^2)-analysis as 'two-parameter analysis' implicitly assumes normality of the returns. The more general term is 'two-moment analysis'.

the expansions in order to obtain a satisfactory approximation to the true distribution function. This implies that only a few (four, e.g.) moments are used for approximating the true distribution. Jennings [1971, pp.800-802], for example, employs a truncated (two terms, three moments) Cornish & Fisher series expansion in order to approximate the probability of obtaining a below-target return (cf. chapter one, section 1.2.2). A more recent example is Jarrow & Rudd [1982], who use a generalized Edgeworth series expansion (on the basis of four moments), which allows for arbitrary approximating distributions and the possibility of non-existing moments.

Apart from this general argument, one could raise utility-theoretic arguments to consider statistical moments in a **direct** way, as representing the relevant probabilistic information. In effect, the choice for the moments is in a natural way suggested by the Taylor series expansion of the investor's expected utility function (see chapter one, section 1.1.3). The first moment then refers to the locus of the distribution ('expected return') and the higher moments refer to its shape ('risk').

When a set of moments of a portfolio's return distribution is given, their direct and indirect use, as set out above, is more or less straightforward. In a portfolio context, however, we encounter *(i)* the combination problem: how to determine the portfolio moments from the moment information of the individual securities, and *(ii)* the information problem: how to assess the securities' necessary moment information. To tackle both problems, the use of factor models is crucial. In this section, we accentuate the general principle to multi-moment portfolio analysis, remembering that the underlying idea carries over to a much wider field of application.

The use of moments

Already at the dawning of (E,σ^2)-analysis, concerns arose about neglecting higher moments of the probability distributions.[41] Aside from theoretic and empirical investigations into the role of the

[41] The first sentence of Brozen [1950, p.326], who discussed Markowitz' [1950] contribution to the Chicago Meeting of the Econometric Society, was: "Markowitz disregards the influence of the third moment"! Even earlier, Hicks [1939, p.125 fn.1] was concerned about the role of skewness aside from expectation and dispersion. The idea of approximating the distribution of investment returns by a limited number of moments was already raised by Hicks [1934, p.195]: "Each [frequency] curve could be rigidly defined by taking a sufficiently large number of moments, and an approximation to the situation obtained by taking a small number". The case where only the expected value and standard deviation were considered was apparently treated by Hicks in detail.

third moment[42], however, it was not until some 20 years later that Jean [1971] gave impetus to a general extension of the two-parameter portfolio analysis by also considering the third and higher distribution moments.[43]

Before stepping to the n-moment portfolio analysis and the potential role of MFMs therein, some remarks about the use of moments are in place. For brevity, we will indicate central moments as moments.

Firstly, do moments of the distribution exist and up to what order do they exist? Within the stable (Paretian) family, members with characteristic exponent smaller than two do not possess a (finite) variance and members with characteristic exponent smaller than one do not even possess a mean. In these cases, only the parameters of the distributions can be used. The point can be weakened, however, by observing that "if one is willing to assume [that] the values of the random variables are finite with probability 1 and the number of states-of-the-world is finite, all central moments must also be finite since they represent finite sums of finite numbers" (Rubinstein [1973a, p.62 fn.3]). So, finite-moment alternatives can be formulated for the stable (Paretian) hypothesis.

Secondly, we have the 'moment problem': does the entire set of moments (assuming that they all exist) determine a distribution uniquely? The answer is in general no: there exist distributions that are not determined by their moments.[44] One familiar example is the lognormal distribution, but in that case the indeterminacy can be removed by log-transforming the variables.[45] Kendall & Stuart [1969, p.109] provide other examples, but note at the same time that "[f]or all ordinary purposes (...) a knowledge of the moments, when they all exist, is equivalent to a knowledge of the distribution function: equivalent, that is, in the sense that it should be possible **theoretically** to exhibit all the properties of the distribution in terms of the moments" (Kendall & Stuart [1969, p.87]).

Instead of focussing on all moments, n-moment portfolio analysis considers only the lower n moments of a distribution. This can be motivated (or defended) by appropriate utility-theoretic considerations, or by assuming that higher order moments are exact functions of the n lower moments. In this context the 'principle of moments' is relevant,

---

42) See for example Friedman & Savage [1948], Pratt [1964], Arditti [1967], Levy [1969] and Arditti [1969, 1971], Samuelson [1970] and Tsiang [1972].

43) The normative analysis for an individual investor was commented on and extended by Arditti & Levy [1972, 1975], Jean [1972, 1973], Ingersoll [1975], Schweser [1978b] and Granito & Walsh [1978]. The three-moment normative case was extended to a positive pricing context by Kraus & Litzenberger [1976, 1983] and Barone-Adesi [1985]. See Schweser [1978a] for an early comparison.

44) For the conditions, we refer to Kendall & Stuart [1969, pp.109-112] or Spanos [1986, p.74].

45) Cf. Feldstein [1969] on this point; see also Elton & Gruber [1974].

which says that distributions which have a finite number of the lower moments in common will be (least squares) approximations to each other. Even when only a limited set of three or four moments is considered, the approximations often turn out to be very good in practice.[46) Furthermore, as the order of the moment increases, it becomes more and more difficult to find distributions whose moments up to some order are equal, whereas their higher moments are different.

Thirdly, we have the interpretation problem of the moments. Assuming all moments finite, the first two moments can be interpreted as measures of location and dispersion of the distribution. The third moment or its normalized equivalent, the skewness (defined as the third moment over the cube of the standard deviation), is often interpreted as a measure of asymmetry. However, a zero third moment is a necessary condition for symmetry, not a sufficient. So although some distribution possesses a zero third moment, it may be very skewed. As Mood, Graybill & Boes [1974, pp.75-76] remark: "knowledge of the third moment gives almost no clue as to the shape of the distribution".[47)
The fourth moment (or, in its normalized form, the kurtosis) is often regarded as a measure for the degree of peakedness of a distribution. Positive (or lepto-) kurtosis is then interpreted as to indicate that the distribution is more peaked around its center than the (mesokurtic) normal distribution, while negative (or platy-) kurtosis then indicates that the distribution is more flat. However, as shown by Kaplansky [1945], any combination of lepto- and platykurtosis and peakedness of a distribution is possible. Neither does lepto- and platykurtosis give information about the fatness of the distribution at the tails.[48)
To the best of my knowledge, any intuitive explanations of the fifth and higher moments are not available (but if they were, we would expect them to break down just as in the case of the third and fourth moment).

Multi-moment portfolio selection

In an n-moment portfolio selection problem, the particular investor chooses on the basis of the first n moments of the portfolio return

---

46) See Kendall & Stuart [1969, pp.87-88]. However, for various counter examples, see Mood, Graybill & Boes [1974, p.77] and the references cited in Barnes, Zinn & Eldred [1978, p.227].

47) See also Tsiang [1972, p.359, fn.6] on this point. Furthermore, the argument can be generalized: zero odd-order moments are a necessary and not a sufficient condition for symmetry, so we can have an asymmetrical distribution with as many zero odd-order moments as desired (Kendall & Stuart [1969, p.85]).

48) The third distribution presented by Kaplansky [1945] has excess kurtosis, but it can easily be checked that its tails are thinner than the tails of the normal distribution already beyond less than one standard deviation from the mean.

distribution.[49] The problem is conceptually identical to the two-parameter portfolio selection problem. The investor constructs the feasible set of portfolios using his subjective joint probability distribution for the returns on the risky securities. Then a risk free security is introduced and tangents, which originate from the risk free rate on the expected return axis, are drawn to the bounding surface of the risky-securities-only efficient set. The tangency point between this new efficient surface and the investor's indifference surface indicates his optimal portfolio. Thus a 'micro separation theorem' is invoked, according to which the investor's optimal combination of risky securities is given by the tangency point of the efficient surface with the risky-securities-only efficient surface.

The question is how the investor's optimum combination of risky securities can be obtained. Aside from the observation that the investor should adjust his holdings in each risky security until the derived optimality conditions are met[50], no operational directions or clues are given and the investor's optimal risky securities portfolio is considered as given. Consequently, the problem of obtaining and processing the required input data for the portfolio selection problem is sidestepped. Jean [1971, pp.508-509] only provides expressions for the moments of a two-security portfolio, but under the separation theorem this is of course sufficient. Things change when we want to determine the composition of the investor's optimal risky securities portfolio.

The complexity of multi-moment portfolio analysis

In the following, we will consider central moments, i.e. moments around the mean.[51] Given N securities with returns $\{\underline{r}_i\}_{i\in N}$, the n-th moment of a portfolio p with investment weights $\{x_i\}_{i\in N}$ is:

$$(3.39)\quad M_n(\underline{r}_p) = M_{np} = E\{[\underline{r}_p - E(\underline{r}_p)]^n\} = E\left\{ \left[\sum_{i=1}^{N} x_i\,[\underline{r}_i - E(\underline{r}_i)\right]^n \right\}$$

$$= \sum_{i_1=1}^{N} \cdots\cdots \sum_{i_n=1}^{N} x_{i_1}\cdot\ldots\cdot x_{i_n}\cdot E\{[\underline{r}_{i_1} - E(\underline{r}_{i_1})]\cdot\ldots\cdot[\underline{r}_{i_n} - E(\underline{r}_{i_n})]\}$$

49) It is then assumed that the n-th moment exists, implying that the n-1 lower order moments also exist. In this place, we do not worry about an adequate representation of an investor's preferences in terms of moments; see Brockett & Kahane [1992] on this point.

50) For example Jean [1971, pp.513-514], Jean [1973, p.490] and Ingersoll [1975, p.793].

51) When starting from raw moments (i.e. moments around the origin), the central moments can easily be derived from the raw moments by the 'parallel axis theorem' (a binomial expansion; cf. Kendall & Stuart [1969, pp.55-56]).

So for N securities in the portfolio, the expression for the n-th portfolio moment consists of N^n terms! In addition to the n-th moments of the individual security returns, eq. (3.39) contains $N(N^{n-1}-1)$ intermediate terms, which represent the comoments.[52] They are measures of the interdependence or covariability between the N securities.[53]

Although not all of these intermediate terms are different[54], the expressions for a portfolio moment increase in complexity when higher order moments are considered. This means that in the computation of higher order portfolio moments, more and more detailed information about the covariability between the security returns is required. In order to enhance the tractability (and hence applicability) of multi-moment portfolio selection models, procedures must be found to cope efficiently with the multivariate interactions. One (and in fact, the only) solution to this problem is to resort to conditioning as in factor models.[55]

---

[52] Also termed cross moments, mixed moments, joint moments or product-moments.

[53] As a second order covariability measure, the covariance measures linear correlation. Higher order covariability measures such as coskewness and cokurtosis indicate curvilinear associations.

[54] We can see this when we expand the first line of eq. (3.39) by repeatingly applying the binomial theorem. The resulting multinomial expansion takes the form:

$$M_{np} = \sum_{j_1 \ldots j_N}^{n} C_{j_{1\ldots N}} \cdot x_1^{j_1} \cdot x_2^{j_2} \cdot \ldots \cdot x_N^{j_N}$$

$$\times\, E\{\ [\underline{r}_1 - E(\underline{r}_1)]^{j_1} \cdot [\underline{r}_2 - E(\underline{r}_2)]^{j_2} \cdot \ldots \cdot [\underline{r}_N - E(\underline{r}_N)]^{j_N}\ \}$$

where the summation is to be taken over all N summation indices such that their sum is n (i.e., $j_1 + \ldots + j_N = n$), and where the multinomial coefficient $C_{j_{1\ldots N}}$ is defined as:

$$C_{j_{1\ldots N}} \equiv \frac{n!}{(j_1!)(j_2!)\cdot \ldots \cdot (j_N!)}$$

(For an equivalent expression, see Mood, Graybill & Boes [1974, p.531].)

[55] Instead of the direct computation of the moments, we could think to construct the moment generating function of the portfolio return, by means of which we could calculate these moments indirectly and thus simplify their representation. When it exists, the moment generating function determines the portfolio return distribution uniquely. The problem is how to derive the moment generating function of the portfolio return from the moment generating function of the individual securities. In addition, there are serious deficiencies attached to the use of the moment generating function. A random variable can have no moments at all but still have a moment generating function; conversely, a random variable can have all or
(footnote continued on next page)

By extending mean-variance analysis to include skewness, both the information problem and the computation problem are magnified disproportionally. In this context, Elton & Gruber [1991, p.231] sigh that "practical portfolio analysis in three moments must await development of a set of analytical techniques for estimating and solving problems involving skewness measures." To the best of my knowledge, these techniques haven't been developed since. Actually, these techniques do not have to be developed since they already exist, albeit somewhat disguised. After all, the role and application of index (or factor) models in the mean-variance framework can easily be extended towards multi-moment portfolio analysis. As there exist no studies in which the use of factor (or index) models as a solution to the problem is explicitly advocated[56], we illustrate how the concept of 'derived covariances' in the single index (or factor) model can be extended to 'derived comoments', simplifying the covariability structure of security returns.

Reducing complexity by conditioning in factor models

We start off with conditioning security returns on one or more external indices or factors. From section 2.4.1 we have for security i:

$$(3.40) \quad \underline{r}_i = a_i + \sum_{j=1}^{k} b_{ij}\underline{\delta}_j + \underline{e}_i \qquad i \in N$$

together with $E(\underline{e}_i) = 0$ and (without loss of generality) $E(\underline{\delta}_j) = 0$.

In mean-variance analysis, the error terms $\underline{e}_i$ are assumed to be mutually uncorrelated and uncorrelated with the common factors $\{\underline{\delta}\}_j$. This implies that the model incorporates all factors that account for **common variance** among the security returns. The stronger assumption needed for multi-moment portfolio analysis is that now the conditioning

some moments but no moment generating function. A random variable can even have all or some moments and a moment generating function, but the latter does not generate moments. For more details and examples, we refer to Kotlarski [1975].

For this reason, the characteristic function is used in theoretical analyses. The characteristic function exists for every distribution and determines it uniquely. The main problem again is how to derive the characteristic function of the portfolio return from the characteristic function of the (dependent) individual securities.

56) An exception that apparently remained unnoticed, however, is Stone [1973]. He uses the single index model to decompose skewness in a common and a specific component. Furthermore, he presents a linear approximation to the cubic programming problem in which "[s]kewness, a previously omitted variable, has been explicitly included (...) in a computationally feasible fashion" (p.632).

is exhaustive in a broader sense: all factors that account for **common variability** among the security returns are accounted for. Consequently, the model is 'strict' in the sense that residual returns are truly security specific or idiosyncratic. Hence, the error terms $\underline{e}_i$ are assumed to be mutually independent[57] and independent of the common factors. For large, approximately equally weighted, portfolios ($x_i \sim 1/N$) the law of large numbers can be invoked, so the distribution of $\underline{e}_p$ degenerates as N increases. This process of naive diversification eliminates (at least in theory) the idiosyncratic part of a portfolio's return distribution. The complementary common part of a portfolio's return distribution can be manipulated analogously to the Markowitz diversification in mean-variance analysis, i.e. by considering the comoments between the securities. A multi-factor model then allows for simple specifications of these comoments.

For a portfolio p that is perfectly diversified in the sense that the distribution of $\underline{e}_p$ has degenerated, we have from eq.(3.40):

$$(3.41) \quad \underline{r}_p = a_p + \sum_j b_{pj} \underline{\delta}_j$$

where:

$$(3.42) \quad b_{pj} \equiv \sum_{i=1}^{N} x_i \cdot b_{ij} \qquad j=1,\ldots,k$$

The n-th moment of the portfolio return is then given by:

$$(3.43) \quad M_n(\underline{r}_p) = E\{ \left[\sum_{j=1}^{k} b_{pj} \underline{\delta}_j \right]^n \}$$

$$= \sum_{j_1=1}^{k} \cdots\cdot \sum_{j_n=1}^{k} b_{pj_1} \cdot\ldots\cdot b_{pj_n} \cdot E\{\underline{\delta}_{j_1} \cdot\ldots\cdot \underline{\delta}_{j_n}\}$$

which, for k<N, is a much simpler expression than eq.(3.39). With the assumptions made above, the dimensionality of the security return covariability structure is reduced from $N(N^{n-1}-1)$ relationships between the security returns to *(i)* the $k(k^{n-1}-1)$ relationships between the factors and *(ii)* the Nxk relations between the securities and the factors, as measured by the response coefficients $\{b_{ij}\}_{i,j}$.

Under the additional assumption that the common factors themselves are mutually independent, the expression for the n-th moment

---

57) Note that we point at the independence of the whole set of error terms. This **complete independence** is stronger than the pairwise independence of the error terms. See Kendall & Stuart [1969, p.24] on this point.

of the portfolio return can further be simplified to:[58]

$$(3.44)\quad M_n(\underline{r}_p) = \sum_j b_{pj}^{\,n} \cdot M_n(\underline{\delta}_j)$$

Of course, the assumption that it is possible to construct perfectly diversified portfolios will only hold in limit cases. Furthermore, considering approximately equally weighted portfolios (in order to invoke the law of large numbers) removes most of the flexibility in portfolio composition, implying that the portfolio selection problem would already be solved. So, considering well balanced and hence well diversified portfolios, the distribution of the idiosyncratic return component is not degenerate for finite N. The same applies to unbalanced portfolios. More realistically, instead, it can therefore be assumed that the contribution of the idiosyncratic return component to the n-th moment is negligible:

$$(3.45)\quad M_n(\underline{r}_p) = \sum_j b_{pj}^{\,n} \cdot M_n(\underline{\delta}_j) + M_n(\underline{e}_p)$$
$$\approx \sum_j b_{pj}^{\,n} \cdot M_n(\underline{\delta}_j)$$

The use of approximate moments implies that the corresponding underlying (joint) distribution is only represented in an approximate way. However, Davies & Ronning [1974, p.143] proved that "if [the investor's] estimate of the joint distribution function of the random returns is not too far away from the actual distribution, then his optimal choice based on this estimate will be not too far away from the actual optimal choice". So, if the factors can account for the major part of the multivariate interactions between the security returns, we can expect the representation to be adequate.

At the beginning of section 3.3.3, we distinguished between the direct and the indirect use of moments for obtaining probabilistic information. Multi-moment portfolio analysis entails a direct use of the moments, and the importance of factor models will be clear from the exposition above. The scheme underlying the indirect use can now be clarified in terms of Figure 3.4 in section 3.3.2. Starting point is the joint distribution of security returns, which is next summarized by their moments and comoments. This representation is then simplified by relating security returns to factors, as a result of which the securities' joint distribution can be summarized in an approximate way by the factor (co—)moments and the securities' factor sensitivities. From this information, the moments of any portfolio's return distribution can readily be computed. Finally, using a series expansion like Edgeworth's, the portfolio's full return distribution can be constructed in an approximate way, thus enabling the inference of probabilistic

[58] Cf. Sharpe's [1967, p.502] approximation for the single index model in the (E,σ^2)-context.

information (like the probability of below-target returns &c.).

The foregoing might suggest that conditioning by factor models can only be employed when the statistical moments of the distributions exist. We therefore stress that the applicability of conditioning is of a very general nature. For example, when security returns follow stable (Paretian) distributions with characteristic exponent smaller than 2, variance does not exist and covariance is not a well-defined statistical concept. Still, **covariabilities** do exist and factor models can be used to account efficiently for these interactions. In this way, factor models provide a natural extension of the role of the single index model in a stable (Paretian) market, as indicated by Fama [1965b] (cf. section 2.3).

3.4 SENSITIVITY COEFFICIENTS (BETA & BETAS) AS RISK MEASURES

When applied to portfolio selection, the question rises to what extent the sensitivity coefficients in the models, as discussed before and in chapter two, are relevant security characteristics. In particular, we are interested in the role of these coefficients as measures for investment risk.

Blume [1971, pp.1-4] considers the beta coefficient in the single index model and offers two interpretations of this coefficient as a measure for 'non-diversifiable risk' in a mean-variance context. The **portfolio approach** rests on the effects of naive diversification on the variance of the specific return component in the single index model. According to the law of large numbers, the return variance of a large, approximately equally weighted ('well diversified') portfolio is proportional to the variance on the index, where the beta coefficient is the proportionality factor (see section 2.3.1). In the **equilibrium approach**, the (market model) beta is the proportionality factor between the risk premium on an individual security and the risk premium on the market portfolio (as follows from the definition of the SML; see section 2.2.2.E). Considering the results of the tests of the CAPM so far, Blume [1970, p.4] remarks that the equilibrium approach "is a weaker and considerably less robust justification than that provided by the portfolio approach."

Babcock [1972] offers a third justification, that does not depend on the market model or single index model. In his **covariance approach**, he investigates under what condition a security can be used to lower the variance of the market portfolio. This condition is that the security's beta relative to the market portfolio is smaller than one.[59]

In this section, we argue that the last two arguments are variations on one and the same theme: they follow from what we will call a portfolio approach. This approach is different from Blume's [1971] portfolio approach. Aside from the portfolio approach there is a second justification for either beta or other sensitivity coefficients as risk measures. This alternative justification directly follows from the conditioning aspect of the models in which these sensitivity coefficients appear and it underlies Blume's [1971] portfolio argument.

We first deal with the portfolio approach. We discuss the role of security characteristics in a portfolio context and analyze the role of beta and betas as relevant security characteristics. We then outline the conditioning approach. We conclude with a resume.

3.4.1 The portfolio approach

For developing the portfolio approach, we first discuss the portfolio optimality principle. We then apply this principle to mean-variance and multi-moment portfolio analysis.

Portfolio optimality principle

We start from a general formulation of an investor's one-period portfolio selection problem. The investor considers the set of portfolio attributes $\{a_{pj}\}_{j\in k}$ relevant, and these attributes enter into a objective function: the preference functional $Z(\cdot)$. The problem for choosing an optimal portfolio p can then be formulated as:

$$(3.46)\qquad \max_{\{x_i\}_{i\in N}} \; Z(a_{p1}, \ldots, a_{pk}) \quad \text{subject to} \quad \textstyle\sum_i x_i = 1$$

where x_i denotes the fraction of initial wealth invested in security i from the opportunity set of N securities. For simplicity, we will not worry about additional constraints.

Forming a Lagrangian, the necessary first order optimality conditions are (together with the full investment condition):

---

[59] Although Babcock [1972, p.701 fn.4] is well aware that the security is already included in the market portfolio, we note that it is odd to consider increasing the security's market proportion. This problem is removed when some other portfolio is chosen as a starting point and, consequently, the beta with respect to this portfolio is considered.

$$(3.47)\quad \frac{\partial Z(\cdot)}{\partial x_i} = \sum_j \frac{\partial Z(\cdot)}{\partial a_{pj}} \cdot \frac{\partial a_{pj}}{\partial x_i} = \lambda \qquad \forall\ i\in p$$

where λ is the Lagrange multiplier.[60]

Rewriting (3.47) as:

$$(3.48)\quad \sum_j \phi_j a_{ij} = \lambda$$

the coefficients

$$a_{ij} \equiv \partial a_{pj}/\partial x_i \qquad \forall\ i\in p,\ j=1,\dots,k$$

represent the marginal contribution of security i to the portfolio attribute a_{pj}, and

$$\phi_j \equiv \partial Z(\cdot)/\partial a_{pj} \qquad \forall\ i\in p,\ j=1,\dots,k$$

is the marginal contribution of this portfolio attribute to the objective function. It is clear then that, in a portfolio context, the characteristics of individual securities are relevant insofar they contribute to the characteristics of the portfolio. This is the main message of 'portfolio theory' (whether 'modern' or not) and it makes 'investment analysis' synonymous to 'portfolio analysis'.

Furthermore, the **marginal** contribution to portfolio attributes is important. This implies that the marginal contributions to the respective portfolio attributes are the relevant security attributes. As, for consistency with $\sum_i x_i = 1$, the portfolio operator is linear, it in turn follows that the relevant security attributes are necessarily linear (or additive) in the portfolio weights. This can be seen by considering securities u and v that are incorporated in the optimal portfolio p with fractions x_u and x_v. Next, we combine these securities into a 'sub'-portfolio w so that $x_w = x_u + x_v$, or equivalently, $x_u = \alpha x_w$ and $x_v = (1-\alpha)x_w$ for some constant α.[61] The investment fractions of the securities u and v in the 'sub'-portfolio w are thus α and $(1-\alpha)$, respectively. As:

$$(3.49)\quad \frac{\partial a_{pj}}{\partial x_w} = \frac{\partial a_{pj}}{\partial x_u} \cdot \frac{\partial x_u}{\partial x_w} + \frac{\partial a_{pj}}{\partial x_v} \cdot \frac{\partial x_v}{\partial x_w}$$

---

60) We assume that the second order conditions are satisfied (for example because $Z(\cdot)$ is concave). When short selling is not allowed, eq.(3.47) must hold for all securities i that have $x_i > 0$ in the optimal portfolio.

61) This, of course, suggests an alternative procedure: to swap security w for securities u and v in the portfolio.

we have:

$$(3.50) \quad a_{wj} = \alpha a_{uj} + (1-\alpha) a_{vj} \qquad j=1,..,k$$

So the relevant attributes of w are weighted averages of the attributes of u and v, where the investment fractions α and $(1-\alpha)$ serve as weights. From combining eqs.(3.49), (3.50) and (3.48), we see that when u and v are combined into (or replaced by) w (or when w is decomposed into u and v), the portfolio remains optimal, since

$$\begin{aligned} \textstyle\sum_j \phi_j a_{wj} &= \textstyle\sum_j \phi_j [\alpha a_{uj} + (1-\alpha) a_{vj}] \\ &= \alpha \textstyle\sum_j \phi_j a_{uj} + (1-\alpha) \textstyle\sum_j \phi_j a_{vj} \\ &= \alpha \lambda + (1-\alpha) \lambda = \lambda \quad , \quad \forall \alpha \end{aligned}$$

When (3.50) is violated, we can only conclude that the initial portfolio was not optimal. The argument can be repeated for any combination of two or more securities in the portfolio. Hence, we conclude that in an optimal portfolio, the relevant attributes of a linear combination of securities are a linear combination of the respective relevant attributes.[62]

Eq.(3.48) says that, in an optimal portfolio, the properly weighted average of a security's marginal contributions to the portfolio attributes is the same for all securities. As a result, all securities in that portfolio plot on a hyperplane in $(\phi_1, \ldots, \phi_k)$-space. This is a generalized version of the **'portfolio optimality principle'** as we know it from mean-variance analysis.

Mean-variance and multi-moment portfolio analysis

Sharpe [1985, pp.156-158] introduced the term 'portfolio optimality principle' in mean-variance portfolio analysis. In that context we have

---

[62] A corollary that we like to mention is that, in an equilibrium context, relevant security attributes are all priced linearly. This follows directly from the linearity of these attributes in the portfolio weights. Suppose that our investor is representative for all investors in the market and that the market portfolio m is in equilibrium and is hence the optimal portfolio. Choosing expected return E_m as the first market portfolio attribute, $a_{m1} \equiv E_m = \sum_{i\in m} x_i E_i$, we have $\partial a_{m1}/\partial x_i = E_i$. Inserting this in eq.(3.48) and rewriting, we get $E_i = \omega_1 + \omega_2 a_{i2} + \ldots + \omega_k a_{ik}$, $\forall$ i$\in$m. The term ω_1 can be interpreted as the 'zero-attribute' return and the terms $\{\omega_j\}_{j=2,..,k}$ are the premia for units of the respective attributes.
It also follows that security attributes will be 'priced' linearly whenever they are computed with respect to a proxy portfolio that is optimal with respect to these attributes. This extends Roll's [1977] critique beyond the mean-variance framework.

$Z(\cdot) = Z(E_p, \sigma^2_p)$, where the subscript p refers to the portfolio.[63] Portfolio optimality then dictates that there exists a linear relationship between the securities' marginal expected returns and marginal variances.[64,65] Also in the mean-variance context, Sharpe [1987, p.158] employs this insight to develop an "embarrassingly simple" algorithm for the iterative (and possibly interactive) revision of a feasible portfolio subject to some simple constraints.

We can illustrate (generalized) portfolio optimality in the context of multi-moment portfolio analysis. Suppose an investor considers the first k (central) moments of portfolio returns important. His objective function thus has the form: $Z(\cdot) = Z\,(E_p, M_{2p}, .., M_{kp})$.

For the expected return $E_p = \sum_{i\in p} x_i E_i$, we have $\partial E_p/\partial x_i = E_i$, so marginal expected return equals expected return.

When the higher portfolio moments are measures for risk, the marginal contributions to the portfolio moments are the securities' relevant risk measures. For $j \geq 2$, we have:

$$(3.51)\quad M_{jp} \equiv E\{(\underline{r}_p - E_p)^j\} = \mathrm{Cov}\{\underline{r}_p, (\underline{r}_p - E_p)^{j-1}\}$$
$$= \textstyle\sum_i x_i \,\mathrm{Cov}\{\underline{r}_i, (\underline{r}_p - E_p)^{j-1}\}$$

so:[66]

$$(3.52)\quad \partial M_{jp}/\partial x_i = j\cdot \mathrm{Cov}\{\underline{r}_i, (\underline{r}_p - E_p)^{j-1}\} \qquad j \geq 2$$

This expression represents the marginal contribution of security i to the portfolio's j-th moment. Alternatively, denoting the j-th root of M_{jp} by m_{jp}, so $(m_{jp})^j = M_{jp}$, we have the semi-elasticities:

$$(3.53)\quad \frac{\partial m_{jp}/m_{jp}}{\partial x_i} = \frac{\mathrm{Cov}\{\underline{r}_i, (\underline{r}_p - E_p)^{j-1}\}}{M_{jp}} = \frac{E\{(\underline{r}_i - E_i)(\underline{r}_p - E_p)^{j-1}\}}{M_{jp}} \qquad j \geq 2$$

which measure the **relative** marginal contributions.

Relative marginal contributions have more appeal for two reasons. As a result of the scaling, their (portfolio) weighted average is equal to one. Furthermore, they represent terms that explicitly appear in the

---

63) Fama & Miller [1972, pp.247-250] use $Z(\cdot) = Z(E_p, \sigma_p)$. As we will show below, the choice for standard deviation over variance is more intuitive for analyzing security risk attributes.

64) In order to infer the linearity of this relationship, Sharpe [1985, pp.158,160] uses the linearity of both the (marginal) expected return and the marginal variance in the portfolio weights. However, as we argued above, the linearity follows automatically from the linearity of the portfolio operator.

65) Rosenberg [1981, p.6] refers to "the Markowitz condition".

66) As M_{jp} is homogeneous of degree j in $\{x_i\}_i$, we have by Euler's theorem: $j\cdot M_{jp} = \sum_i x_i\, \partial M_{jp}/\partial x_i$.

relevant literature. For example, for j=2 we have:

$$(3.54)\quad \frac{\partial\sigma_p/\sigma_p}{\partial x_i} = \frac{\mathrm{Cov}(\underline{r}_i,\underline{r}_p)}{\mathrm{Var}(\underline{r}_p)} = \beta_{ip}$$

which is the regression coefficient in the 'market index model':

$$(3.55)\quad \underline{r}_{it} = \alpha_i + \beta_{ip}(\underline{r}_{pt}-E_p) + \underline{e}_{it} \qquad i\in p$$

where the optimal portfolio p takes the place of the index. When we step over from the investor's portfolio equilibrium to market equilibrium, the optimal risky portfolio p is the market portfolio m. Exchanging the subscripts 'p' for 'm', we find that $\beta_{im}=\beta_i$ is 'the' beta coefficient from the standard CAPM and that eq.(3.55) is the market model eq.(2.2) or (2.21). So, when measured relative to the relevant portfolio, return sensitivities are relevant measures for risk.

For the case that j=3, we have:[67]

$$(3.56)\quad \frac{\partial m_{3p}/m_{3p}}{\partial x_i} = \frac{\mathrm{Cov}\{\underline{r}_i,(\underline{r}_p-E_p)^2\}}{M_{3p}} = \frac{E\{(\underline{r}_i-E_i)(\underline{r}_p-E_p)^2\}}{M_{3p}} = \gamma_{ip}$$

Just like the 'beta' above, this is a standardized co-moment. Exchanging investor equilibrium with market equilibrium, $\gamma_{im}=\gamma_i$ is the 'gamma' coefficient from Kraus & Litzenberger's [1976] three-moment CAPM.

We can now consider a generalized or **'polynomial market model'** (that can be regarded as an extension of Kraus & Litzenberger's [1976, p.1090] quadratic market model):

$$(3.57)\quad \underline{r}_{it} = a_i + b_{i1}(\underline{r}_{mt}-E_m) + b_{i2}(\underline{r}_{mt}-E_m)^2 + \dots + b_{ik}(\underline{r}_{mt}-E_m)^k + \underline{e}_{it} \qquad i\in m$$

where the zero-mean error term $\underline{e}_{it}$ is independent of $\underline{r}_{mt}$. The coefficients $\{b_{ij}\}_j$ measure the sensitivity of security i's return for various powers of the demeaned market return. Note that $a_m = 0$, $\underline{e}_{mt} = 0$, $b_{m1} = 1$ and $b_{mj} = 0$ for j=2,..,k. It readily follows that the relevant security attributes are simple functions of the coefficients in this theoretical regression model:

$$(3.58)\quad \frac{\partial m_{jm}/m_{jm}}{\partial x_i} = \frac{\mathrm{Cov}\{\underline{r}_i,(\underline{r}_m-E_m)^j\}}{M_j(\underline{r}_m)} = b_1 + \sum_{s=2}^{k} b_s \frac{M_{s+j-1}(\underline{r}_m)}{M_j(\underline{r}_m)} \qquad j=2,..,k$$

This implies in turn that the regression coefficients in the polynomial

[67] Provided $M_{3p} \neq 0$. Otherwise $\gamma_{ip} \equiv \mathrm{Cov}\{\underline{r}_i,(\underline{r}_p-E_p)^2\}$.

market model are relevant security attributes. When the second and higher moments are considered as portfolio risk measures, the sensitivities are security risk measures. More general, when the second and higher moments are considered as relevant portfolio attributes, for example because they describe the shape of the portfolio return distribution, then the sensitivities are relevant security attributes.

As an investor combines securities into a portfolio, a security's contribution to portfolio risk is important. We showed that betas in a (polynomial) market (index) model can be regarded as risk measures because they are measures for a security's marginal contribution to portfolio risk. As a result of the portfolio optimality principle, an investor will balance expected return and marginal risk(s) of the securities in his portfolio. Aside from an individual investor's portfolio equilibrium conditions (which mark normative investment theory), we can consider market equilibrium conditions (as studied by positive investment theory). In the latter case, the aggregate risky portfolio is optimal and marginal risks are now measured relative to this portfolio. As the portfolio argument includes both the case of a 'micro equilibrium' and a 'macro equilibrium', it embraces both Babcock's [1972] covariance argument and Blume's [1971] equilibrium argument.

3.4.2 The conditioning approach

In developing the CAPM, Sharpe [1964, pp.438-439] characterizes beta as a measure for the predicted response of a security's return to changes in the market portfolio's return. This is consistent with the portfolio approach, in which the optimal (individual or aggregate) portfolio is the object against which the relevant security attributes are specified. Later on, to develop an intuition for the SML, Sharpe [1964, pp.441-442] states that "only the responsiveness of an asset's rate of return to the level of economic activity is relevant in assessing its risk." It will be clear that this argument no longer rests on the market model[68] but on a factor model. In effect, securities are related to quantities other than the (optimal) portfolio in which they are incorporated. This second approach for justifying beta and betas as risk measures will be termed the conditioning approach. Depending on the quantities to which security returns are related to (or conditioned on), we can distinguish different forms.

68) Sharpe [1964, p.438 fn.23] refers to the single index model and not to the market model.

Conditioning on a market index

When the market portfolio is not an investor's optimal portfolio it can still be sensible to relate security returns to a (general) stock market index as a proxy for this market portfolio. A broad market index is relevant in its own as it represents the aggregate or average of all investors' portfolios (or the portfolio of the average investor), and reflects whether 'the market' is dull, bull or bear. A portfolio beta above or below unity then characterizes an aggressive or a defensive portfolio. Basically, an investor can exploit the aggressivenes or defensiveness of a portfolio in two ways. As the market as a whole is the average of all investors' portfolios, an investor can evaluate the return variability (riskiness) of his portfolio relative to that of the average investor's portfolio and specify some target beta. In addition to this passive strategy, an investor can follow an active strategy and change the beta of his portfolio in anticipation of upward or downward market movements (market timing). So, also (or even in praticular) outside the equilibrium framework of the CAPM, the portfolio's sensitivity to the return on the 'market portfolio' can be important.

It is striking that in the more popular literature (viz. The Institutional Investor), the "beta revolution" (Welles [1971, p.21]) was first explained by interpreting beta as a sensitivity coefficient for changes in the market return, whereas a decade later (in Wallace [1980]) beta is inextricably bound up with the CAPM (and goes down with it).[69] In perspective of the long existent but recently re-emphasized doubts concerning the empirical validity of the CAPM[70], the interpretation of beta as a sensitivity coefficient has gained both significance and popularity. However, Rosenberg [1981, p.12] argued that the construct of an approximated market portfolio remains important despite recognition of the falsehood of the CAPM, since "It is difficult to imagine a market equilibrium in which covariance with the equity market portfolio, or some risk measure that is closely akin to this, would not be rewarded". He concludes (p.16) that "The concept of reward to equity market risk (or beta) is a theoretical insight, that, in my view, is likely to endure." Rudd & Rosenberg [1980, p.599] take another viewpoint by defending the use of betas since "in the absence of a means of identifying the true portfolio, it seems reasonable to employ broad-based sectoral portfolios to define normal investment policy."

---

[69] Dealing with the two interpretations of beta, Lindahl-Stevens [1978] can be regarded as a reaction on Welles [1971], whereas Rosenberg [1981] and Markowitz [1984] comment on the views expressed in Wallace [1980]. Wagner & Quay [1974, p.37] also note that "beta has different meanings for different leanings."

[70] Cf. Fama & French [1992] and the following discussion (for exámple Black [1993], Grinold [1993] and Chan & Lakonishok [1993]).

Conditioning on factors

It is a little step from relating security returns to a market index or sector indices, to relating returns to underlying factors. This is the second form of the conditioning approach. In fact, by imposing the market model on a multi-factor model, Rosenberg [1974] analyzed extra-market components of return covariance, Sharpe [1977] derived a multi-beta CAPM, and Rosenberg [1981] (less formally) defended the role of beta in the light of a market risk premium.

Exhaustive conditioning

In the two forms of the conditioning approach above, the interpretation of beta and betas as sensitivity coefficients does not critically depend on their evaluation in a portfolio context. In the third form, however, the context of a large 'well diversified' portfolio is crucial. As before, security returns are conditioned on one or several external indices or factors, but now the conditioning is exhaustive: all factors that account for common variability of security returns are accounted for. Consequently, the model is 'strict' in the sense that residual returns are security specific. As a result of naive diversification, the return distributions of well diversified portfolios can be completely described in terms of their factor sensitivities and the (joint) distributions of the factors. As these sensitivities determine the extent to which the characteristics of the factor distributions are passed through to portfolio return distributions, the sensitivities can be considered as risk measures. This was depicted in Figure 3.4.B; in section 3.3.3, especially eq.(3.45), this is shown for the case where the distributions can be characterized by their moments.

The crucial condition for this interpretation is whether the set of factors is exhaustive or not. In a portfolio context, we expect the factor risk to constitute the major part of the total risk. In practice, however, the residual portfolio return component may be too large to be neglected.

3.4.3 Resume

In the portfolio approach, securities are evaluated in a portfolio context, relative to the (optimal) portfolio. In the conditioning approach, securities are still evaluated in a portfolio context, but now relative to other quantities (like market indices or factors) whose stochastic nature generates portfolio return distributions. The differences between the two approaches is parallel to the differences between the market model and its variants on one side and the single index/factor model and its generalizations on the other.[71] We showed how the results can be transposed to a general setting.

3.5 PRESENT VALUE MODELS AS A BASIS FOR DERIVING FACTOR MODELS

This section elaborates on a methodology for deriving multi-factor models of security returns, shifting the emphasis from statistical aspects to economical content and transparancy. The theoretical basis is provided by the familiar present value framework that serves to develop an adequate price function for securities in which state variables act as arguments. With the help of a closed form version of this price function, a simple relationship between the securities' returns and changes in the state variables can be modelled (cf. section 3.2). In using the present value framework for deriving the price function, we closely join in with standard immunization theory in which the duration and (hyper-) convexity characterictics of bonds are used to relate bond returns to changes in interest rates.

The advantage of this approach is in our opinion two-fold: an underlying present value model can provide clues to what factors to consider and the resulting expressions of the factor elasticities can be examined, providing insight in the nature of the relations between returns and factor movements as well as in their expected intertemporal stationarity.

Section 3.5.1 provides a general discussion of present value models. In sections 3.5.2 and 3.5.3, the step from present value model model to multi-factor model is discussed for both bonds and stocks. Section 3.5.4 evaluates the empirical validity of present value models.

3.5.1 Present value models for financial securities

3.5.1.A General formulation of present value models

The present value principle dates back to Fisher [1906, Ch.XI-4 and Ch.XIII-8]. In the case of certainty, the market price of a financial asset (representing a claim on a future cash flow stream) in a perfect capital market equals its capitalized value, i.e. the value of the cash flow stream, discounted at the market rate of interest (the risk free rate). This is shown for a constant and varying (i.e. time dependent, non-stochastic) interest rate by Samuelson [1937, pp.477,485], who uses an arbitrage argument. In that case, the rate of return on the asset

[71] The many different uses for beta as suggested in the abundant literature can be traced to these two justifications of beta coefficients as risk measures.

equals the risk free rate.

In general, the value of a (capital or financial) asset is equal to the summed, appropriately discounted value of all future revenues that are generated by that asset. In equilibrium, the current (market) price equals the present value. In this section, we assume that this condition is satisfied. Under uncertainty, there are two approaches: the 'risk adjusted discount rate' (RADR) method and the 'certainty equivalent' (CE) method.[72]

The RADR approach

In the RADR approach, the discount rates are adjusted according to the (cash flow) risks involved[73] Assuming that the generated cash flows $\{\underline{CF}_t\}_t$ are equidistant in time, the current (t=0) market value of the cash flow stream is:

$$(3.59)\quad P_0 = \sum_{t=1}^{T} \frac{E_0(\underline{CF}_t)}{(1+k_t)^t}$$

where T is the horizon and where $E_\tau(.)$ represents the expectation conditional on the (market) information set at $t=\tau$. Almost without exception, it is further assumed that one single RADR applies for all cash flow dates: $k_t = k$. One argument used in defense of specifying a constant discount rate is that one is interested in the present value of the **stream** of cash flows, and not in the present values of each single cash flow.[74] A stronger argument is that expression (3.59) is "nearly empty" because there will exist many sequences of discount rates $\{k_t\}_t$ for which the relationship holds (LeRoy [1989, p.239]). Therefore, (3.59) is specialized by assuming a constant discount rate.[75] This restriction in turn implies that the conditional expected return on the underlying asset equals the constant k, and is hence independent of the conditioning set (Samuelson [1965, 1973]).

---

72) These methods assume that the discount rate is deterministic (but it may be non-constant over time). A third discounting approach that emerged only recently is 'stochastic discounting' (Bühlmann [1992]), where this assumption is dropped. We will not consider this approach here, however.

73) Fisher [1907, pp.207-208]. Fisher [1907, p.213] calls this discount rate a "pseudo or impure rate of interest".

74) Cf. Chen [1967, pp.316-317].

75) Brennan [1971, p.1120], however, goes as far as saying that a constant k "is but an algebraic artefact, and as such should be irrelevant for decision-making purposes", especially for evaluating a firm's investment and dividend policies.

The CE approach

The alternative approach entails discounting the certainty equivalents of the cash flows at the risk free rate. The certainty equivalent cash flow $CE_{\tau}(\underline{CF}_t)$, evaluated at time $\tau<t$, is the smallest amount to be received with certainty for which an investor would exchange the expected cash flow $E_{\tau}(\underline{CF}_t)$. Using the certainty equivalent factor $\alpha_{\tau,t} \equiv CE_{\tau}(\underline{CF}_t)/E_{\tau}(\underline{CF}_t)$, the analogon to (3.59) becomes:[76]

$$(3.60) \quad P_0 = \sum_{t=1}^{T} \frac{CE_0(\underline{CF}_t)}{(1+r)^t} = \sum_{t=1}^{T} \frac{\alpha_{0,t}\, E_0(\underline{CF}_t)}{(1+r)^t}$$

where r denotes the (constant) risk free rate and $0 \le \alpha_{0,t} \le 1$.

Comparing eqs. (3.59) and (3.60) shows that the RADR and CE approaches are consistent if and only if uncertainty is resolved at a constant rate over time, so that $\alpha_{0,t} = \alpha^t$.[77] Thus, discounting at the RADR adjusts distant cash flows downward to compensate for risk in a way that is not easily predictable. From this analysis, Robichek & Myers [1966, p.728] conclude that "[o]n a conceptual level, therefore, the certainty-equivalent framework appears superior to the use of a constant, risk-adjusted discount rate, since it is applicable in a wider variety of situations."[78,79]

3.5.1.B Bonds versus stocks: specifying cash flow patterns

Bonds

From the general specification of present value models for financial

---

[76] In the certainty equivalent factor we can actually recognize the "coefficient of caution", defined by Fisher [1906, p.276] as the ratio of "commercial value" (i.e. the value an investor is willing to pay for a gamble) and the "mathematical value" (i.e. the expected value of the payoff(s) of that gamble).

[77] Cf. Gordon [1962, pp.63-65] and Robichek & Myers [1966]. The choice of $\alpha_{0,t} = \alpha^t$ with $0 \le \alpha \le 1$, indicates intertemporal constant risk aversion according to the RADR framework. More detailed specifications of $\alpha_{0,t}$ would introduce arbitrary choices.

[78] See also the discussion between Chen [1967] and Robichek & Myers [1968]. In general, we could have the counter-intuitive $\alpha_{\tau,t} < 0$ or $\alpha_{\tau,t} > 1$, as illustrated by Bar-Yosef & Mesznik [1977] and Beedles [1978a]. Since dividends and coupons are non-negative cash flows, $E_{\tau}(\underline{CF}_t)$ and $CE_{\tau}(\underline{CF}_t)$ are always positive, so the bounds $0 \le \alpha_{\tau,t} \le 1$ are satisfied in the cases we consider later. But even then, decreasing certainty equivalent factors $\alpha_{\tau,t}$ do not unambiguously reflect increasing risk; see Beedles [1978b] on this point.

[79] The CE approach is quite popular in capital budgeting theory. For an application of the CE method to estimate the 'term structure' of certainty equivalent factors for risky bonds, see for example Silvers [1973] and Carleton & Cooper [1976].

assets, we can derive two extremes. On the one hand, we have a straight[80] default free, fixed coupon bond. Immediately after payment of the last coupon, the current (t=0) value B_0 of this bond equals the present value of the future (deterministic) payoffs:

$$(3.61)\quad B_0 = \sum_{t=1}^{T} \frac{CF_t}{(1+{}_0r_t)^t} = \sum_t \frac{CF_t}{(1+y)^t}$$

where T = the remaining maturity of the bond
CF_t = the fixed coupon of the bond for $0 < t < T$
CF_T = the coupon plus the face value, to be received at time T
y = the yield to maturity of the bond,

and where ${}_0r_t$ denotes the discretely compounded spot rate over the interval [0,t]. The set $\{{}_0r_t\}_t$ denotes the term structure of interest rates. In case of a flat term structure, all interest rates equal the yield to maturity y of any bond.

Stocks

On the other hand, we have stocks. The present value (current price) of a share of common stock is equal to the appropriate discounted value of all expected future cash flows accruing to the stockholder. This raises two questions: what should be discounted and at what rates? As shown by Miller & Modigliani [1961] in a perfect market environment, several approaches are equivalent. For simplicity, following Williams [1938, pp.57-58,80], it is generally assumed that the relevant cash flows to be capitalized are ultimately the expected dividends.[81] For this reason, present value models for stocks are termed 'dividend discount models'. The expected dividend stream can show various patterns over time and the actual cash flows that accrue to the stockholder can differ significantly from their expected values. Indeed, the different nature of the relevant future cash flows is the most outstanding feature separating bonds and common stocks. For this reason, we resign ourselves (for the time being) with the assumptions of intertemporal constant risk aversion and a flat term structure of interest rates; this implies a single, constant risk adjusted discount rate. Instead, we will concentrate on the (expected) dividend stream over time.

---

[80] So there are no options embedded in the bond.

[81] Total earnings **per se** are not capitalized in order to avoid double counting, since the portion not paid out as dividends is reinvested to maintain the level of the future earnings and dividend stream. So "[e]arnings are only a means to an end, and the means should not be mistaken for the end. Therefore we must say that a stock derives its value from its dividends, not from its earnings. In short, a stock is worth only *what you can get out of it*" (Williams [1938, p.57], original italics). See Sloane & Reisman [1968] for a classification system of equity valuation models according to what is discounted.

Under these assumptions, the current (i.e. at t=0) theoretical equilibrium stock price, immediately after paying the current dividend D_0, can be expressed as:

$$(3.62)\quad P_0 = \sum_{t=1}^{\infty} \frac{E_0(\underline{D}_t)}{(1+k_t)^t} = \sum_{t=1}^{\infty} \frac{E_0(\underline{D}_t)}{(1+k)^t} = \sum_{t=1}^{\infty} \frac{CE_0(\underline{D}_t)}{(1+r)^t}$$

where $E_0(\underline{D}_t)$ = the current expectation for the dividend at time t
k = $r + \pi$ = the RADR
r = the risk free interest rate
π = the risk premium
$CE_0(\underline{D}_t)$ = the certainty equivalent for the dividend at time t.
The variable k, again, can be identified as the expected return on the stock.

In order to enhance the tractability of the present value model in theoretical analyses, it must be simplified even further. This can be done by specifying growth patterns of the (expected) dividend stream over time. The most simple closed-form present value formula results if it is assumed that expected dividends grow at a constant rate g:

$$(3.63)\quad P_0 = \sum_{t=1}^{\infty} \frac{D_0(1+g)^t}{(1+k)^t} = \frac{D_0(1+g)}{k-g} = \frac{E_0(\underline{D}_1)}{k-g} \qquad \text{for } k > g,$$

or in the certainty equivalent framework:

$$(3.64)\quad P_0 = \sum_{t=1}^{\infty} \frac{D_0(1+g')^t}{(1+r)^t} = \frac{D_0(1+g')}{r-g'} = \frac{CE_0(\underline{D}_1)}{r-g'} \qquad \text{for } r > g',$$

where g' is the growth rate of the certainty equivalent dividends, $g' \equiv \alpha(1+g)$.[82]

While the derivation of eq. (3.63) is often attributed to Williams [1938, p.88, eq.(17a)] and is even widely identified as the Gordon & Shapiro [1956, p.105] -formula[83] or the Gordon [1962, p.45, eq.(4.6)] -growth model[84], its origin goes back at least as early as

---

[82] Note that the identity $E_0(\underline{D}_1) = D_0(1+g)$ in the well-known constant growth model eq.(3.63) implies that the current dividend D_0 serves as the spanning point for the future expected dividends or that the current dividend D_0 did not deviate from its formerly expected value. The last terms on the RHS of eqs.(3.63) and (3.64) are less restrictive in this respect.

[83] Except for the use of continuous instead of discrete compounding, Gordon & Shapiro's formula is identical to that of Williams. It is then puzzling (and incorrect) that Gordon & Shapiro [1956, p.104] state that "the models he [Williams] developed were arbitrary and complicated so that the problem of growth remained among the phenomena dealt with qualitatively".

[84] Again, except for the use of continuous instead of discrete

Makeham [1868, p.198 fn].[85)]

By rewriting eq. (3.63), the expected (required) total return on the stock can be expressed as

$$(3.65) \qquad k = \frac{E_0(\underline{D}_1)}{P_0} + g$$

where the first term in this decomposition is the expected dividend yield and g is the expected price return $E_0(\underline{P}_1)/P_0-1$ for the next year.[86)] Analogously, from eq.(3.64) it follows that the certainty equivalent dividend yield $CE_0(\underline{D}_1)/P_0$ and the certainty equivalent price return $g' = CE_0(\underline{P}_1)/P_0-1$ sum to the risk free rate r.

Modelling cash flow patterns

The aspect of cash flow growth, separating bonds and common stocks, deserves more attention. Depending on the magnitude (and uncertainty) of the growth rate implied by the projected dividend stream, we can say that common stock can be situated in a continuum where at one side the common stock predominantly possesses the characteristics of a straight, default risk free bond and where at the other side its characteristics of earnings growth potential dominate.[87)] These growth stocks offer a problem. If the dividend growth rate is larger than the risk adjusted discount rate, eq.(3.63) represents a diverging geometric series, yielding an infinitely large stock price. Yet we can observe stocks that have (had) a growth rate in excess of the discount rate, but do (did) not trade against infinitely high prices. Durand [1957, p.349] noticed

---

compounding, Gordon's formula (4.6) is identical to that of Williams [1938]. However, Gordon [1962, p.46 fn.4] insists that Williams "did not get as far as Eq.(4.6). (...) The first statement of Eq.(4.6) as a stock valuation model (...) was contained in Gordon and Shapiro [1956]..."

85) In a footnote on p.199, an editor of the Journal terms Makeham's result "curious" and illustrates it further. For a detailed historic overview of the development of the model, we refer to Soldofsky [1966].

86) That the growth rate of the stock price equals the growth rate of the expected dividends can easily be derived from expressing eq. (3.63) for P_1 and then dividing this by the original equation for P_0. Nonetheless, Horvath [1993, p.276] denies this, but his argument rests on faulty reasoning.

87) The essence of this growth is not mere expansion, but the existence of investment opportunities that yield more than the discount rate. See Walter [1956, pp.31-32], Durand [1957, p.355], Miller & Modigliani [1961, p.417] and Merrett & Newbold [1982, p.6] on this point. The nature of growth is best highlighted by the 'investment opportunities approach'; cf. Miller & Modigliani [1961, pp.416ff] and Mao [1966].

the "remarkable analogy between the appraisal of growth stocks and the famous Petersburg Paradox" and offered several explanations for this 'growth paradox'. Two of the explanations for finite prices of growth stocks were in fact already provided by Clendenin & Van Cleave [1954]:

- it is expected that the growth rate will decrease to a 'normal' level after some time;
- investors use a much higher discount rate for more distant dividends.

Still, because of the problem involved in estimating the length of the 'above normal' growth period (to which the stock price is very sensitive) and because of the problem of evaluating the large but uncertain dividends in the very remote future, growth stocks "seem to represent the ultimate in difficulty of evaluation" (Durand [1957, p.362]).

Stepping from the positive approach (i.e. explaining finite prices for growth stocks) to the normative approach of valuing growth stocks, Durand [1957, p.359] realized "the difficulty (...) to find some reasonable, objective basis for setting up an appropriate schedule of [discount] rates". Clendenin & Van Cleave [1954, p.373] use an arbitrary scheme of discount rates, but Burrell [1960, p.75] criticizes the approach of increasing discount rates for compensating the uncertainty of remote dividends. Instead, attention focussed on enhancing the reality of the constant growth model by specifying more realistic dividend growth patterns.

A cross-section of the variety of proposed (and more or less mechanical or arbitrary) dividend growth models is presented in Table 3.1.[88)]
Of all proposed models, a 3-phase growth model is the most appealing and widely used. This model is a **blend** between the models of Molodovsky, May & Chottiner [1965] and Brigham & Pappas [1966]. The model assumes first a period of above normal growth, then a transition phase in which the growth rate gradually declines to the 'normal' level, and finally a steady growth state in which the normal growth rate is maintained in perpetuity. As it reflects some life cycle of the firm and the limited investment opportunities that offer a return higher than the cost of capital, the model is more realistic than a constant growth model, or a 2-stage growth model (with an abrupt change of the growth rate). At the same time, however, the calculations involved are much more elaborate and complicated.[89)] In particular, there exists no simple closed-form

---

88) We refer to Wendt [1965] and Bierman & Hass [1971] for other overviews. In addition we note that several functional forms for cash flow patterns, as developed in the field of capital budgeting, could equally well be applied to stocks. The interpretation of the Laplace transform as a present value function, for example, was initially proposed by Grubbström [1967] and rediscovered by Buser [1986].

89) Cf. Fuller [1979] for the formulas for the standard 3-phase growth model. By its ability to cope with many growth stages, the model of Brooks & Helms [1990] is most general, but computationally quite complex. See also Bower [1992] on this model.

Table 3.1: Alternative patterns of dividend growth in dividend discount models.

	growth rate				
	high, constant	declining to next level	average ('normal')	zero	no dividends
Clendenin & Van Cleave [1954][b]	×			×	× (> 60 yrs)
Burrell [1960]	×			×	
Miller & Modigliani [1961, pp. 421-422 fn.15][e]	×			×	
Holt [1962][c]	×		×		
Malkiel [1963]	×		×[d]		
Eiteman [1968]	×		×[d]		
Molodovsky, May & Chottiner [1965][c]	×	linear		×	
Williams [1938, pp.89-94][a]	×	asymptotically		×[f]	
Mao [1966][a,e]	×	×		×	
Brigham & Pappas [1966]		-linear -de-/accelerated -S-shaped	×		
Bauman [1969]		linear	×		
Bierman, Downes & Hass [1972]		asymptotically	×[f]		
Malkiel [1979]		exponentially	×		
Brooks & Helms [1990]	multiple stages of constant growth		×		

[a] According to a logistic curve.
[b] Because the evaluation horizon is truncated after 60 years, this is not really a 3-phase model.
[c] Growth of earnings instead of dividends is considered.
[d] Finite horizon model. Expected end-of-horizon price (after high growth period) is estimated by multiplying grown earnings with a normative price-earnings ratio for standard (average-growing) firms.
[e] Investment opportunities are considered.
[f] Because of the asymptotics, this growth rate is never reached.

expressions for the theoretical stock price in a 3-phase model.[90)]

A present value model under flexible dividend growth

Ideally, we are looking for a present value model model formulation that gives enough flexibility to incorporate a realistic pattern of dividend growth rates without at the same time becoming untractable. A suchlike model must offer the possibility to incorporate investment analysts' dividend forecasts over the next, say, 3 to 5 years as well as the possibility to specify a transition phase in which the growth rates converge towards some 'steady state' economy-wide growth rate. The restriction for such a formulation would be the description of the growth rate curve by means of only a limited number of parameters, so that a simple closed-form expression for the present value formula can be derived.

As a compromise between keeping the appealing simplicity of a closed-form present value formula and the reality of a variable dividend growth model, we derived a flexible dividend growth model. For the formal underpinnings of the model, we refer to Appendix 3.C. The basic idea is that there exists a long term normal growth rate of expected dividends that is reached after an initial period of (variable) abnormal growth:

$$P_0 = \sum_{t=1}^{T} \frac{Div_0 (1+g_t)^t}{(1+k)^t} + \sum_{t=T+1}^{\infty} \frac{E_0 (Div_T)(1+g)^t}{(1+k)^t} \tag{3.66}$$

where T is the length of the period where the actual growth rates g_t differ from the long term growth rate g. Next, the stock price is decomposed in the present value of steady state growth opportunities and the present value of abnormal growth opportunities as follows. For the period of above (below) normal initial growth, in each period a specific dividend amount is added to (substracted from) the dividend stream that grows at the long term growth rate g. The dividend components stemming from above or below normal growth are transferred back in time and summed to the next expected dividend. This transformed first dividend then grows in perpetuity at the normal rate g. This transformed dividend stream is equivalent to the original dividend stream in that it results in the same stock price.

After some exercises, the closed-form present value formula for the stock price finally becomes:

[90)] One possible exception is Bierman, Downes & Hass' [1972] model, but the authors themselves were the first to recognize that the dividend time path is still severely restricted. Fuller & Hsia [1984] provide a simple approximation to the 3-phase model; see however Appendix 3.C on this point.

$$(3.67) \qquad P_0 = \frac{D_0(1 + g)\{1+G\}}{k - g} = \frac{D_0(1 + g')\{1+G'\}}{r - g'} \qquad \text{for } k > g,\ r > g'$$

where G (G') is the adjustment factor for the first expected (certainty equivalent) dividend. This is a simple expression, but can at the same time embrace any flexible dividend growth pattern.

Comparing eq.(3.67) with (3.63), it is tempting to interprete (1+G) as a proportionality factor between the price of a growth stock and the price of a stock that does not possess above or below average growth (but that is identical for the rest). However, as there is no reason why the expected return k on both stocks would be the same, this interpretation is incorrect.

In can be argued that the simplicity of (3.67) is more appearent than real, because G follows from a rather complicated expression. Although Appendix 3.C provides a simple approximation to G in order to estimate k for the general case, the simplicity of (3.67) lies in the approximation of the elasticities of the stock price, especially with respect to k. For that goal, the factor G (and G') is treated as exogeneous from k (or r) and g (or g').

3.5.2 From present value model to factor model: bonds

3.5.2.A Sources of bond returns

As Fisher [1906, p.284] states:

> "We now see that the value of capital [i.e. the present value of a cash flow stream] actually changes through any one of four causes: (1) Through the effect of discount; that is, while no income is being received, the value of the capital will rise along a discount curve. (2) Through the periodic detachment of income; that is, at times when income or outgo occurs, the capital will be directly decreased by the amount of the income and increased by the amount of the outgo reached and passed. (3) Through unexpected changes in the rate of interest; that is, when such changes occur causing revaluations of the future by discounting it at a new rate, the value of the capital will change correspondingly — increasing if the rate of interest falls, and decreasing if it rises. (4) Through unforeseen changes in expected income."

This paragraph is immediately followed by the remark that "[t]he fourth case is the one of most practical importance" (Fisher [1906, p.284]). This applies particularly to stocks and will be discussed in section 3.5.3. As the present value model is most obvious for fixed income securities, we first consider the step from present value model to factor model for bonds.

The first three causes for a change in price, and hence a return, apply for fixed-coupon, default free straight bonds.[91] The current market value of such a bond is:

$$(3.61') \qquad B_0 = \sum_{t=1}^{T} \frac{CF_t}{(1+{}_0r_t)^t} = \sum_t PV_t$$

where PV_t denotes the present value of the cash flow to be received at time t. As we focus on a factor model as a risk model, we will neglect the bond return as a result of a reduction of the remaining time to maturity. Under the assumptions made sofar, then, changes in the term structure of interest rates are the 'only' sources of risk for bonds. This is the starting point for section 3.5.2.B. In section 3.5.2.C, we reframe the results in a factor model context.

3.5.2.B Bond returns and interest rates

In this section, we briefly highlight some results from immunization theory in which the duration and (hyper-) convexity characterictics of bonds are used to relate bond returns to changes in interest rates.

The instantaneous bond return as a result of infinitesimally small instantaneous changes in the term structure of interest rates is:

$$(3.68) \qquad \frac{dB}{B} = \sum_t \frac{dPV_t}{B} = \sum_t \frac{dPV_t}{PV_t} \cdot \frac{PV_t}{B}$$

Inserting

$$(3.69) \qquad \frac{dPV_t}{PV_t} = -t \, \frac{\dot{d_0r_t}}{1+{}_0r_t}$$

in eq. (3.68), yields the relation between the instantaneous bond return and small proportional changes in the spot rates:

$$(3.70) \qquad \frac{dB}{B} = \sum_t (-t) \frac{PV_t}{B} \cdot \frac{d_0r_t}{1+{}_0r_t}$$

As several spot rates within the term structure of interest rates can change, the resulting interest rate risk is multi-dimensional and

[91] We will adhere to Fisher's [1906] decomposition. A somewhat different classification of sources of bond returns is provided by McEnally [1981]. In a performance context, Fong, Pearson & Vasicek [1983] present yet another return decomposition.

expression (3.70) is not tractable. Things can be simplified, however, by specifying elasticity functions that relate the proportional changes in the spot rates to relative changes in some set of reference rates. Choosing, for simplicity, a single elasticity function $\psi(t)$, relating proportional changes in spot rates to relative changes in the reference rate r, we get:[92)]

$$(3.71) \qquad \frac{d_0 r_t}{1 + {}_0 r_t} = \psi(t) \frac{dr}{1 + r}$$

Incorporating eq. (3.71) in (3.70) yields:

$$(3.72) \qquad \frac{dB}{B} = - \left[\sum_t t \frac{PV_t}{B} \psi(t) \right] \frac{dr}{1 + r}$$

Eq. (3.72) relates bond price variability to 'interest rate' variability. The term between brackets is generally denoted as a **duration measure**: the sensitivity of the (instantaneous) bond return for infinitesimally small relative changes in one plus 'the' interest rate. Although in many cases is referred to 'the' duration of a bond, specific definitions of duration depend on the specifications of the process that is assumed to drive interest rate changes.[93)]

When $\psi(t) = 1$ for all t, the term between square brackets in (3.72) represents the Fisher & Weil [1971] duration of the bond. Simplifying further by assuming a flat term structure (so ${}_0r_t = r$, $\forall t$), the term between square brackets reduces to the well-known Macaulay [1938] duration D:

$$(3.73) \qquad D = \frac{\sum_t tCF_t (1+y)^{-t}}{B} = - \frac{dB(y)}{dy} \cdot \frac{(1+y)}{B(y)}$$

where the notation B(y) indicates that the bond price B is considered as a function of its yield to maturity.[94)] Although Macaulay [1938] used this duration in a literal sense (as the weighted average time to maturity or 'mean term') to characterize bonds in an adequate way, the measure is relevant in its role as an 'interest rate' sensitivity. The

[92)] Cf. Maloney & Yawitz [1986].

[93)] For example, the observation that short term rates fluctuate more than long term rates could be incorporated by specifying $\psi(t) = \psi^t$, $0<\psi<1$, and setting r equal to the short rate ${}_0r_1$. Substituting this in (3.72) then yields the bond price elasticity with respect to the specified, instantaneous changes in the interest rates. For an alternative process, see Khang [1979].

[94)] In mathematical terms, the bond price can be represented as a function of the yield. This does not deny the fact that the yield to maturity is implied by the bond price given the magnitude and the timing of the cash flows.

last equality in eq.(3.73) identifies this duration measure as the negative of the bond's (point) elasticity with respect to its internal yield to maturity[95].

The link between bond price variability and duration was developed by Hicks [1939, pp.186-188] and Samuelson [1945, p.19 fn.1]. The relationship was rediscovered by Fisher [1966] (although in a different context), Hopewell & Kaufman [1973] and Weil [1973], who extended the role of duration as a proxy for interest rate risk. Since then, many papers are devoted to this subject.[96]

Although simple formulations for a security's elasticity with respect to changes in interest rates (like the Macaulay and Fisher-Weil duration measures) are very popular, Ingersoll, Skelton & Weil [1978] and Cox, Ingersoll & Ross [1979] note that these measures only apply for the assumed type of change(s) in the term structure that underlies the development of the duration measure. The possibility that the actual (instantaneous) change in the term structure does not correspond to the asumed type of change is the source of 'stochastic process risk' (Bierwag, Kaufman & Toevs [1983a]), invalidating the duration measure as a risk measure.

3.5.2.C Multi-factor models for bond returns

Theoretical developments have developed along three lines. On the one hand, the univariate analysis was extended to incorporate non-linearities. On the other hand, the univariate analysis was extended to incorporate multiple reference rates. Finally, a truly multi-factor approach emerged when, in addition to interest rates, other characteristics are considered.

Incorporating non-linearities

As a point elasticity, duration applies to infinitesimally small changes in the yield to maturity and thus represents the first term in a Taylor series expansion. As the relationship between bond price and yield to maturity is non-linear, one can only expect a reasonable first order approximation for small yield changes. For evaluating changes in bond prices when yield changes are larger, additional terms in the expansion must be considered. The first order approximation can be extended to a second order approximation by including the convexity. The convexity, ½C, is defined as half of the second order elasticity of the bond price

95) The more commonly used 'modified duration' is a semi-elasticity and relates the instantaneous bond return to a percent-point change in the yield: $(dB/B)/(dy)$.

96) Cf. the contributions to Bierwag, Kaufman & Toevs [1983b] and the references in Bierwag [1987].

with respect to its yield[97] and it represents the bond's return sensitivity for second order relative yield changes:

$$(3.74) \qquad \tfrac{1}{2}C = \tfrac{1}{2}\cdot\frac{\sum_t t(t+1)CF_t(1+y)^{-t}}{B} = -\tfrac{1}{2}\cdot\frac{d^2B(y)}{dy^2}\cdot\frac{(1+y)^2}{B(y)}$$

Using duration and convexity, the immediate bond return can be approximated as:

$$(3.75) \qquad \frac{\Delta B}{B} \approx -D\cdot\frac{\Delta y}{1+y} + \tfrac{1}{2}C\cdot\left[\frac{\Delta y}{1+y}\right]^2$$

Even higher order elasticities ('hyper-convexities') can be considered. In practice, the use of convexity adds limited information.[98] As the knowledge of a bond's duration allows a better approximation of its price reaction to yield changes than is suggested in the literature, the information value of convexity is somewhat limited even from a theoretical point of view. This can be inferred from the results in our Appendix 3.D.[99] In that Appendix, we show that the non-linearities that arise from the discounting effect can adequately be captured by a using only the first order elasticity in a simple non-linear formulation.

Incorporating multiple reference rates and multi-factor models

The (hyper-) convexity analyses focus on the approximation of the non-linear **single-argument** bond price function B(y). By concentrating on the yield or 'the' interest rate, the results are limited to at best changes in the **level** of the term structure. Another route of analysis, in contrast, stresses the multi-dimensional character of interest rate risk by explicitly recognizing that more than one rate can change. By specifying functional forms for the term structure, changes in its **slope** and **form** can be analyzed. The applicability of the analysis can be enhanced by specifying a factor model of the term structure.[100] As a compromise between tractability and generality, it is then assumed that

---

97) See for example Fabozzi & Fabozzi [1989, p.71].

98) This applies even more to hyper-convexities. To my knowledge, there are no practical applications of hyper-convexities. However, these hyper-convexities should not be confounded with the higher order terms in the 'duration vector' as derived by Chambers, Carleton & McEnally [1988]. Assuming that polynomials in time are adequate to describe the functional form of the term structure, these duration measures were derived from the bond price's sensitivities to changes in the polynomial coefficients. In this sense, their model belongs to the class of multi-factor duration models, as described below.

99) See Hallerbach [1994a] for a further elaboration of this point.

100) Cf. Sharpe [1985, pp.326-329] and Litterman & Scheinkman [1991].

there exist some reference rates that drive all other relevant spot rates. This returns us to eq. (3.72), but then in multivariate form.[101] In case of a two-factor model, for example, a perfect relationship between (changes in) all interest rates and two reference rates r_1 and r_2 is assumed.[102] In this way, the single duration measure D is replaced by multiple 'partial durations' D_j, each representing the elasticity of the bond price for changes in the corresponding source j of risk:[103]

$$(3.72') \quad \frac{dB}{B} = - D_1 \frac{dr_1}{1 + r_1} - D_2 \frac{dr_2}{1 + r_2}$$

For the practical implications of these specifications, the early study by Cooper [1977] is interesting. Allowing for changes in both the level and the shape of the term structure, he analyzes the sensitivity of security returns to interest rate changes of both types. Cooper demonstrates that returns on portfolios, consisting of securities with moderate remaining terms to maturity (up to 7 years, e.g.), can be significantly affected by changes in the shape of the term structure.[104] The longer maturity range of the term structure, however, is almost flat and hence more characterized by changes in the interest rate levels than by changes in the shape. So, analyzing the interest rate sensitivity of securities with longer maturities, changes in the shape of the term structure will be less and less important.

Towards multi-attribute models

Specifying multiple (partial) duration measures for interest rate risks is only one extension of the framework. Indeed, there is no reason why the analysis should be restricted to interest rate risks at one side and straight, default free, fixed coupon bonds on the other side. To start from the latter, general multi-factor models for bond prices have been

[101] In this sense, the Taylor expansion eq.(3.75) does not represent a multi-factor model. Cf. Bierwag, Kaufman & Latta [1988].

[102] Mostly a short rate and a long rate, or a long rate (for the level) and the difference between the long and a short rate (for the slope of the term structure). Cf. Cooper [1977, p.712] and Haugen [1993, p.369], for example.

[103] For an overview of models to describe (changes in) the term structure of interest rates, we refer to Dobson, Sutch & Vanderford [1976], Cooper [1977] and Vasicek & Fong [1982], and more recently, Nelson & Siegel [1987] and Bliss & Ronn [1989]. Ingersoll [1983, pp.177ff] develops a two-factor model. See also Brennan & Schwartz [1983] and Nelson & Schaefer [1983] for factor models.

[104] I.e., the greater volatility of the term structure over the short maturity range compared to the long maturity range.

developed.[105] In these models, that are better referred to as 'multi-attribute models', bond prices are linked to both market-wide factors and bond-specific characteristics. Market-wide factors include the term structure of default free interest rates (accounting for the time value of money) and yield spread factors. The yield factors reflect premiums for default risk (relating to the particular bond sector or quality rating), liquidity premium (for the degree in which the bond issue is traded) and a current yield factor (relating to the different taxation of capital gains and interest income). Bond-specific characteristics include cash flows, adjusted for embedded options, as well as exposures relating to the yield spread factors. As this multi-dimensional relationship is estimated empirically, the difference between the market price and the fitted price also is a bond-specific characteristic. Given the model, a change in the bond price can now be explained in terms of changes in the term structure of default free interest rates and in unexpected changes in the cash flows (due to embedded options), the factors and/or the exposures (for example, due to a change in the rating).[106]

Another route for extending the framework is obvious by now and twofold:
- investigating the interest rate risk characteristics ('duration and convexity') of stocks and other securities (financial assets);
- specifying other sources of risk for these securities.

This will be done in the next section.

3.5.3 From present value model to factor model: stocks

Inspired by the analysis of interest rate sensitivities of bonds, a profound attention is paid in the literature to estimate 'equity duration', as a measure for the (negative of the) sensitivity of common stock returns to interest rate changes. By analogy to bond duration analysis, a frequently employed approach for gauging equity duration starts from a present value model of the stock. Section 3.5.3.A discusses the theoretical relationship between stock returns and interest rates.

As remarked before, changes in the term structure of interest rates are the 'only' sources of risk for default free straight bonds. Stocks, however, have an additional risk component, stemming from changes in the investors' expectations of the risky future dividends. After all, cash flow growth is the most outstanding feature separating bonds and common stocks. This additional risk component can be decomposed and differentiated to multiple sources; this is shown in

[105] See Kahn [1991] and Kahn & Gulrajani [1993] and the references cited therein.

[106] In contrast to the multi-factor models as described above, the term structure of default free interest rates is incorporated completely.

section 3.5.3.B.

As is apparent from section 3.5.2, the present value model is used in the context of comparative statics. That is, in the present value function (i.e. the price function) the discount rate and the growth rate of the (expected) dividends are assumed to be constant over time. Then, the effect of a ceteris paribus change in these variables on the present value is evaluated. This comparative static approach can be contrasted with the dynamic model as developed by Campbell & Shiller [1988, 1989]. Their 'dividend ratio model' entails a linear approximation to the relationship between log returns, log dividends and log prices over time, and can be considered as a dynamic Gordon growth model. In their model, the log dividend-price ratio is related to the present value of one-period interest rates, adjusted for dividend growth, discounted at a constant rate. Their analysis, however, focusses on predicting future discount rates and future dividend growth rates (cf. chapter two, section 2.5.2). For forecasting purposes, then, the adequate time series modelling of the relationships between the variables is crucial. This forecasting approach, where stock returns are related to explanatory variables that are known in advance, can be contrasted with what Campbell [1991, p.158] calls "the contemporaneous regression approach", where stock returns are related to contemporaneous innovations in variables which might plausibly affect the stock market. As we follow the latter approach and strive to maintain a close link with the models that are used in investment practice, we will continue to use the comparative statics approach as set out for bonds.

3.5.3.A Stock returns and interest rates[107]

In the literature, equity duration is defined as the negative of the elasticity of the theoretical stock price with respect to a change in the risk adjusted discount rate k. This is analogous to the definition of duration as a proxy for a bonds's interest rate risk, as discussed in section 3.5.2.B. Starting from the Gordon growth model eq. (3.63) for the stock price at time t, this elasticity can be computed as:[108]

$$(3.76) \qquad \frac{dP_t/P_t}{dk/(1+k)} = -\frac{1+k}{k-g}$$

Solving eq. (3.63) for (1+k) and 1/(k-g) respectively, and multiplying both expressions, we get:

[107] This section draws on Hallerbach [1994c].

[108] Cf. Durand [1957, p.358] or Boquist, Racette & Schlarbaum [1975, p.1363].

$$(3.77) \qquad \frac{dP_t/P_t}{dk/(1+k)} = -\ (\ 1 + 1/\delta_t\ )$$

where $\delta_t \equiv D_t/P_t$, the current dividend yield of the stock.[109] According to eq. (3.77), the equity duration equals one plus the reciprocal of the current dividend yield.[110]

Starting from eq. (3.64) in the certainty equivalent framework, we obviously get the same result:

$$(3.78) \qquad \frac{dP_t/P_t}{dr/(1+r)} = -\ \frac{1 + r}{r - g'} = -\ (\ 1 + 1/\delta_t\ )$$

The certainty equivalent approach to measure the interest elasticity of common stocks is employed by Haugen & Wichern [1974] and Casabona, Fabozzi & Francis [1984].

The equity duration paradox

Typically, the values of this theoretical derived equity duration range from 20 to 50 or even as high as 80 (Lanstein & Sharpe [1978]). From data provided by Farrell [1985, p.23], for example, and using eq. (3.77), the duration of the S&P-Index ultimo 1984 can be estimated as 25! These values are much larger than empirically estimated durations, that are not likely to exceed 10 or 15.[111] This is the so-called 'equity duration paradox'.

To resolve this paradox, the 'conventional' theoretical approach, leading to eqs. (3.77-78), can be critized on several grounds. One criticism is the assumption of a flat term structure of interest rates. The equity duration, estimated on the basis of eqs. (3.77-78), then provides only an indication for the degree in which the stock price responds to parallel shifts in the term structure. However, the practical implication of this conclusion will be limited.[112].

---

[109] Williams & Pfeifer [1982] provide a different derivation of this expression.

[110] In case of continuous compounding, we get $(dP_t/P_t)/dk = -1/(k-g) = -P_t/E_t(\underline{D}_{t+1})$. The equity duration then equals the reciprocal of the **expected** dividend yield.

[111] See chapter four, sections 4.3.2 and 4.3.4.

[112] In the long maturity range, the term structure of interest rates is almost flat and thus predominantly exhibits changes in its level. This implies that returns on assets with longer remaining terms to maturity (such as stocks, which have an infinite maturity) will be predominantly affected by changes in the interest rate level and not by changes in the shape of the term structure. Cf. section 3.5.2.C, where we referred to Cooper's [1977] results.

There is however one snake in the grass. In case the shape of (footnote continued on next page)

There are two more serious sources of bias. The cause of these biases is the fact that the conventional equity duration measure treats the cash flow stream, accruing to the investor, as fixed coupons. The time pattern as well as the sensitivity of the expected dividend stream for interest rate changes is thus ignored. These 'dividend effects' could explain the large differences between theoretical durations and empirically estimated durations. We first discuss the exogeneous dividend growth bias that results when the relationship between the growth rate of the expected (or certainty equivalent) dividends and the interest rate is ignored. Next, we discuss the dividend growth pattern bias, that stems from neglecting the impact of abnormal growth patterns on the equity duration.

The exogeneous dividend growth bias

Allowing for a relationship between the growth rate of the expected dividends and the interest rate, we have from eq. (3.64):

$$\frac{dP_t/P_t}{dr/(1+r)} = -\frac{1+r}{r-g'} + \left[\frac{1+r}{1+g'} + \frac{1+r}{r-g'}\right]\frac{\partial g'}{\partial r} \tag{3.79}$$

To simplify this expression, we first rewrite eq. (3.64) in terms of the current dividend yield $\delta_t \equiv D_t/P_t = (r-g')/(1+g')$. Solving δ_t for $(1+r)$ and $1/(r-g')$ and multiplying then gives $(1+r)/(r-g') = 1 + 1/\delta_t$, the familiar expression for the conventional equity duration. Solving δ_t in turn for $(1+r)/(1+g')$ yields $(1+r)/(1+g') = 1 + \delta_t = \delta_t(1+1/\delta_t)$. Substituting these results in (3.79), we get:

$$\frac{dP_t/P_t}{dr/(1+r)} = -\{\,1 + 1/\delta_t\,\}\left[\,1 - (1+\delta_t)\frac{\partial g'}{\partial r}\right] \tag{3.80}$$

where the expression between braces is the conventional equity duration. According to this formula, the adjusted theoretical equity duration equals the conventional equity duration, multiplied by an adjustment factor (between the large brackets) that depends on the sensitivity $\partial g'/\partial r$ of (certainty equivalent) dividend growth to interest rate changes.[113] For a positive (negative) sensitivity, the conventional duration measure will overestimate (underestimate) equity duration.

---

the term structure of interest rates reveals information about the future course of economic activity, it can have an impact on the growth rate of the expected dividends. This is then an additional feature of the exogeneous dividend growth bias, to be discussed below.

[113] Note that in a risk adjusted discount rate framework, we have to substitute $\partial g/\partial k$ for $\partial g'/\partial r$ to get the relevant elasticity.

The dividend growth pattern bias

The dividend growth pattern bias stems from neglecting the impact of abnormal growth patterns on the equity duration. Note that this in fact entails a criticism of the validity of the underlying constant growth model for the theoretical stock price.

In section 3.5.1.B, we discussed some present value models for stocks. Depending on the growth characteristics of the projected dividend stream, common stocks can be classified in a range from bond-like stocks to 'growth engines'. A 3-phase dividend discount model is appealing and widely used to take into account the influence of the growth pattern on the theoretical stock price. As for more or less realistic dividend growth patterns there do not exist simple closed-form expressions for the theoretical stock price, neither do exist simple closed-form expressions for the implied theoretical equity duration. Haugen & Wichern [1974] consider a two-stage growth scenario, but even in this more simple case their formulas become so complex that they can only be evaluated numerically.

As argued in section 3.5.1.B, the flexible dividend growth model (as derived in Appendix 3.C) is a compromise between keeping the appealing simplicity of a closed-form present value formula and the reality of a variable dividend growth model. A long term normal ('steady state') growth rate g' of certainty equivalent dividends serves as a pivot for the abnormal growth rates during some initial period of time. The stock price is decomposed in the present values of steady state growth opportunities and abnormal growth opportunities, and the first dividend is adjusted for the latter growth opportunities. This yields a transformed dividend stream, equivalent to the original dividend stream in that it results in the same stock price. It is a simplification, because the transformed (certainty equivalent) dividends grow at a constant rate g'.

From the closed-form present value formula (3.67) in certainty equivalent form, the elasticity of the stock price with respect to the interest rate is easily derived as:

$$(3.81)\qquad \frac{dP_t/P_t}{dr/(1+r)} = -\frac{1+r}{r-g'}$$

This expression is exactly the same as in the constant growth case. Note however, that both the dividend yield δ_t and the growth rate g' in eqs. (3.67) and (3.81) are different from δ_t and g' in eq. (3.64)! The analogy with the constant growth case can be maintained by working with the current dividend yield, based on the transformed dividend. From eq. (3.67) we have $\delta_t(1+G') = (r-g')/(1+g')$. Solving $\delta_t(1+G')$ for (1+r) and

$1/(r-g')$ and multiplying then immediately gives:[114)]

$$(3.82)\qquad \frac{dP_t/P_t}{dr/(1+r)} = -\left[1 + \frac{1/\delta_t}{1+G'}\right]$$

where G' is the adjustment factor for the first certainty equivalent dividend. Compared to the conventional equity duration, we see that the degree of (certainty equivalent) abnormal growth G' serves as a correction factor for the current dividend yield. Initial above normal growth ($G'>0$) decreases duration, whereas initial below normal growth ($G'<0$) increases duration. So, for initial above (below) normal growth, the conventional duration measure in eq. (3.77-78) overestimates (underestimates) equity duration.

Combining the biases

We conclude by noting that a theoretical present value model can only provide insights into the interest rate sensitivity when the two 'dividend effects' are taken into account. The conventional expression for equity duration is not adequate. For a complete picture, we combine the effect of both biases in the equity duration.

For the case in which the interest rate sensitivity of the dividend correction factor G' is assumed to be zero ($\partial G'/\partial r = 0$), we simply substitute in eq. (3.80) the current dividend yield based on the transformed dividend, $\delta_t(1+G')$, for δ_t and get:

$$(3.83)\qquad \frac{dP_t/P_t}{dr/(1+r)} = -\left[1 + \frac{1/\delta_t}{(1+G')}\right]\cdot\left[1 - \{1 + \delta_t(1+G')\}\frac{\partial g'}{\partial r}\right]$$

However, this assumption may be unwarranted, since by definition, G' depends on g'.[115)] So even when the abnormal growth opportunities themselves display no interest rate sensitivity, the dividend correction factor exhibits an interest rate sensitivity of $\partial G'/\partial r = -G'/(1+g')$. As readily follows, this reduces (3.83) to:

---

114) As we used some approximations (see Appendix 3.C), the formula (3.82) will not yield exact results. We did however perform some simulations with two-stage growth scenarios, which showed that (3.82) yields a surprisingly good approximation to the (adjusted) theoretical equity duration.

115) From Appendix 3.C it follows that in the certainty equivalent case we have $G' \equiv \Gamma(g'^{*}-g')/(1+g')$, where g'^{*} is the reference rate for abnormal (CE) growth. When $\partial(g'^{*}-g')/\partial r = 0$, we have $\partial G'/\partial r = -G'/(1+g')$.

$$(3.84) \qquad \frac{dP_t/P_t}{dr/(1+r)} = -\left[1 + \frac{1/\delta_t}{(1+G')}\right]\cdot\left[1 - (1+\delta_t)\,\frac{\partial g'}{\partial r}\right]$$

We see that only the growth correction factor between the first square brackets is maintained. The correction factor for the interest rate sensitivity of the growth rate has now disappeared.

In chapter four, section 4.3.6, we will investigate the empirical relevance of correcting for the two biases in conventional equity duration.

3.5.3.B Multi-factor models for stock returns

Following bond analyses, one could extend the univariate equity duration model to a multivariate version by incorporating additional terms, representing the non-linear effects of the relationship. However, as argued in section 3.5.2.C, a truncated univariate Taylor series expansion is not a multi-factor model. Furthermore, one could cope with non-linearities by specifying the simple non-linear model as proposed in Appendix 3.D to this chapter.

Instead, the intrinsic differences between bonds and stocks suggest to consider additional factors. As put forward in section 3.5.1.B, cash flow growth (and the accompanying risks) is the most outstanding feature separating bonds and common stocks. It can then be recognized that the cash flows (the earnings, and hence the dividends) are generated by economic variables. This was noted in an early stage, for example by Samuelson [1937, p.478], who in the context of the present value model remarks: "the income stream depends upon a multitude of parameters, each representing a relevant economic variable or condition." In a more explicit form, Brennan [1973], Treynor & Black [1976], Turnbull [1977] and Gooding [1978] relate uncertain future cashflows to underlying (but unnamed) economic variables. Oxelheim & Wihlborg [1987,1991] extended the analysis by considering the nature and identity of the economic influences on a firm's cash flows. For a recent analysis and extension, we refer to Goedhart [1994].

To model the influences of economic variables on the cash flow stream in the present value model, one can relate the discount rate and the growth rate(s) of expected dividends to underlying economic variables. This was already suggested by Robichek & Cohn [1974, p.442]. Estep, Hanson & Johnson [1983] and Estep, Clayman, Johnson & McMahon [1984] explicitly link the growth rate g in the constant growth model to various economic factors $\{\delta_j\}_j$ by specifying:

$$(3.85) \qquad g = g_0 + \sum_{j=1}^{k} \gamma_j\,\delta_j$$

The obvious extension of (3.76) in the RADR form then is:

$$(3.86) \qquad dP_t/P_t = \left[\frac{\partial P_t/P_t}{\partial k/(1+k)} \right] \cdot \frac{dk}{1+k} + \sum_{j=1}^{k} \left[\frac{\partial P_t/P_t}{\partial g/(1+g)} \right] \cdot \frac{\gamma_j}{1+g} \, d\delta_j$$

where the terms between square brackets denote the elasticities of the stock price with respect to the discount rate and growth rate. One factor candidate that affects both the discount rate and the growth rate(s) is inflation.[116)] Because of the Fisher hypothesis, (expected) inflation is part of the risk free rate, but inflation will also influence the value of expected future cash flows.

An extension of this approach is presented in Hallerbach [1993a], to which we refer for further details.

The advantage of this present approach for analyzing sensitivities is that the future cash flow stream is taken as a starting point. Analyzing the characteristics of this cash flow stream (for example by fundamental analysis) will then yield candidates for the variables that are likely to influence changes in the (expected) levels of the cash flows. It is then a challenge to fundamental analysis to extend the use of dividend discount models from the expected return context to the combined risk and return context.

3.5.4 The empirical validity of present value models

Present value models clearly reflect that financial markets are forward looking. The question now arises whether the expectations regarding the future are rational and can be related to the present value framework.

In a number of studies, collected in his book, Shiller [1989] found that stock prices exhibited excess volatility when compared to the fluctuations in the (future) dividend stream. This is a serious criticism of present value models, but the assumptions underlying Shiller's empirical work have been criticized in turn by Marsh & Merton [1986], for example. In a theoretical context, Black [1986] developed his noise trader hypothesis. According to this hypothesis, market participants cannot distinguish between true news and noise. Summers [1986] also disputes that stock prices represent rational assessments of fundamental values.

We can argue then that the **level** of market prices may not reflect 'true' or rationally assessed fundamental values, but that the **changes** in these prices convey information about changes in expectations about future cash flows. Like the dividend discount model is used in terms of

[116)] Cf. Jahnke [1975], Joehnk & Petty [1980], Estep, Hanson & Johnson [1983] and Leibowitz, Sorensen, Arnott & Hanson [1989].

relative or comparative values of stocks[117] (without pretending to indicate the true value), it can be used in terms of changes in stock values relative to their former value. However, in the light of noise trading and excess volatility, the ability to allocate a large proportion of price fluctuations to changes in fundamental or economic variables will be limited. This is indeed confirmed by studies by Rol [1988] and Fama [1990] (see also chapter four, section 4.4 on this point). Still, it is worthwile to investigate the proportion of return variability that **can** be explained, realizing that 'market fads' and bursts of irrationality cloud the relationship between market values and fundamental or economic variables.

3.6 SUMMARY AND CONCLUSIONS

In this chapter, we highlighted the importance of multi-factor models in portfolio analysis by elaborating on their statistical and economical aspects. In the four sections, we successively dealt with the following aspects of factor models:
- the principle of conditioning;
- the reduction of complexity;
- the concept of risk; and
- their derivation and economic-intuitive content.

The principle of conditioning

Starting from a general setting, we assumed an unspecified price generating function, relating security prices to common underlying state variables or factors. The latter variables describe the securities' relevant economic environment. We showed how to derive a general multi-factor representation of security returns and indicated the necessary assumptions for arriving at (approximately) linear factor representations. We distinguished between factor models in unconditional and conditional expectations, and between discrete time and continuous time model formulations (see Appendix 3.A). Special attention was paid to the nature of the error term in a linearized model.

We provided two interpretations for the link between security returns and their economic environment. One view starts from the economic environment and considers factor models as a mechanism describing how perceived changes in this environment generate security returns. The complementary view starts from the security returns and considers a factor model as a vehicle to condition these returns on

[117] Cf. the relative value argument in Williams [1938, pp.186-187] and the "comparative value concept" of Bauman [1969, p.109].

changes in the environment. In this respect, we distinguished between complete, partial and exhaustive conditioning. In the latter case, the conditioning factors account for all interrelationships between the security returns. Under complete conditioning, the total (true) set of generating factors is considered.

Incomplete information about the return generating process only allows partial conditioning. In a stochastic context, we paid special attention to the linearization of factor models and the interpretation of the resulting sensitivity coefficients. We indicated the importance of linear(ized) factor models in a portfolio context.

Employing the minimum mean squared error criterium, we derived the return expectation, conditional on the factors, as the 'best' approximation to the underlying return generating process. This implies that a factor model, linearized in the set of selected factors, not only can be specified bluntly by assumption but could also be defended on the ground of joint elliptically (spherically) distributed returns and factors. Alternatively, a linearized factor model can be defended as a linear least squares approximation to the unknown true underlying return generating process.

We next confronted the linear (first order differential) approximation with the linear least squares approximation. In the former case, the securities' sensitivity coefficients are gradients, representing local approximations to the return generating function (as derived from the price function). In the latter case, the sensitivity coefficients are (multiple) regression coefficients, representing global approximations to the return generating function. We explored the differences between the two sets of coefficients in detail (see Appendix 3.B), and provided bounds on the differences between them. As we adopt a different set of assumptions, this analysis supplements White's [1980]. In addition, assuming normality of the factors (which is less restrictive than joint normality of returns and factors), we derived the intuitively appealing result that the least squares coefficients equal the expected value of the gradients.

Finally, we viewed linearized factor models from the perspective of non-stationarity and omitted variables bias.

The reduction of complexity

In further developing multi-factor models as statistical concepts, we argued that conditioning reduces two fundamental problems attached to the investment decision: the information problem and the combination problem. These complex problems arise from the interrelatedness of security returns, making their entire joint distribution relevant for portfolio analysis. As simplifying distribution assumptions, however, factor models tackle both problems.

The information problem centers around assessing portfolio return distributions. Difficult in its own right, the problem is aggravated by the necessity of estimating the required inputs for the portfolio problem on the level of the individual securities. We first discussed the information problem on the portfolio level and provided the link between conditioning and decomposition. A multi-factor model actually increases the dimensionality of a portfolio's return, alleviating probability assessments. On the securities' level, the interactions between the security returns obstruct the appraisal of their joint distribution. For the sake of tractability, the representation of this distribution needs to be simplified. Here, we linked exhaustive conditioning by factor models with hierarchical structures. This allows to shift focus from the joint distribution of security returns to the joint distribution of the factors. Now, the dimensionality of the joint return distribution is greatly reduced, enhancing probability assessments. Having conditioning related to the decomposition principle and hierarchical structures, we also provided links with sensitivity analysis and probabilistic modelling and furnished other examples of conditioning.

The combination problem entails the difficulty of aggregating the relevant security characteristics into portfolio characteristics. This problem is closely related to the information problem on the securities' level. Again, the issue is how to account for the joint distribution in a tractable way, but now the focus moves from assessing this distribution to processing it. A multiplicity of portfolios can be composed and the principal problem is how to convolute the securities' distributions into portfolio distributions. Also in this context, the factor approach proved to be invaluable.

To accentuate the underlying idea, we shifted attention from the full probability distributions to their summary measures (moments) on the argument that statistical moments are either relevant by themselves, or apposite for an (approximate) representation of the full probability distribution.

The former case involves a direct use of the moments. We elaborated on multi-moment portfolio analysis and concluded that a factor model representation is indispensable for its feasibility. Instead of summarizing the securities' joint return distribution in terms of their moments and comoments, this representation is simplified by relating security returns to factors. As a result, the joint distribution can be summarized in an approximate way by the (co-)moments of the factors and the securities' factor sensitivities. From this information, the moments of any portfolio's return distribution can readily be obtained.

The latter and more general case entails an indirect use of the moment information. However, the procedure is essentially the same as in the former case, except that it is supplemented with techniques that allow the (approximate) construction of any portfolio's full return distribution from its derived moments. This opens the way for inferring more detailed probabilistic information on the portfolio level (like the probability of below-target returns &c.).

The concept of risk

In the market model, in (multi-) index models as well as in (multi-) factor models, the link between security returns and the corresponding set of conditioning variables materializes in the form of sensitivity coefficients or betas. We explored in what circumstances and to what extent these sensitivities can be regarded as risk measures. Doing so, we distinguished between the portfolio approach and the conditioning approach.

The tenet of portfolio theory is that individual securities must be evaluated in a portfolio context. Given the interactions between security returns and the effects of diversification, this is especially relevant for risk attributes. Judging the relevance of security attributes by their contribution to portfolio attributes constitutes the portfolio approach.

Under fairly general conditions, we first derived that in an optimal portfolio all relevant security attributes are linear. Furthermore, we stipulated a generalized portfolio optimality principle, according to which the properly weighted average of a security's marginal contributions to the attributes of an optimal portfolio is the same for all securities. As a corollary, this result extends Roll's [1977] critique beyond the mean-variance framework. In the context of multi-moment portfolio analysis (encompassing mean-variance analysis as a nested case), we showed how the coefficients in a polynomial market model (including the standard and the quadratic market model as nested cases) can be interpreted as risk measures. Crucial for this interpretation is that the security returns are related to the (optimal) portfolio in which they are incorporated.

In the conditioning approach, the conditioning variable in the market-, index- or factor models is not restricted to the optimal portfolio, but can be chosen freely. We discuss under what conditions the sensitivities resulting from conditioning on market indices or on factors can be interpreted as risk measures. In these two general forms of the conditioning approach, the interpretation of sensitivity coefficients as risk measures does not critically depend on their evaluation in a portfolio context.

As a special case, however, the third form does depend on both a diversified portfolio context and exhaustive conditioning. By relating security returns to a set of common factors, a factor model allows the decomposition of these returns into a systematic part and an unsystematic part. In very general terms, the risk accompanying a portfolio's return is related to the shape of its return distribution.

value function can be moulded in terms of current expectations with respect to future realizations of the economic variables. These information variables concretisize the state variables and now act as arguments of the price function. As a revision of factor anticipations results in security price changes, the return generating process can be represented in the form of a multi-factor model, linking changes in factor anticipations (information variables) to security returns.

We provided a general overview of present value models and discussed the step from present value model model to multi-factor model for both bonds and stocks. The specification of an appropriate present value function is crucial. While maintaining all flexibility on the side of future cash flow patterns, we derived a realistic yet tractable present value model (see Appendix 3.C). From this closed form price function, a multi-factor model can readily be derived. Accentuating on the relationship between stock returns and interest rates, we observed two biases in the conventional expression for a stock's theoretical interest rate sensitivity and reformulated an alternative expression. In order to cope with non-linearities, especially when these arise from the discounting effect, we developed a simple non-linear model formulation (see Appendix 3.D). While using only first order elasticities, non-linearities can adequately be captured.

The advantage of the present value approach is two-fold. An explicit present value model indicates what factors are relevant, and the resulting expressions of the factor elasticities can be examined, providing insight in the nature of the relations between returns and factor movements as well as in their expected intertemporal stationarity. It is then implicitly assumed that present value models provide a valid description of security prices. Although there is evidence to reject the empirical validity of present value models when applied in an absolute sense, it can be argued that these models can adequately account for the securities' relative prices and, most importantly, for the changes in their prices.

APPENDIX 3.A: A MULTI-FACTOR MODEL IN CONTINUOUS TIME

In discrete time, the validity of the general linear MFM in conditional expectations (eq.(3.19)) crucially depends on the intrinsic linearity of the relationship between security prices (or returns) and (changes in) factors. In continuous time, when $\Delta t \to dt$, all relationships are (locally) linear. So even if the price function $P(\cdot)$ is a non-linear function of the state variables $\underline{\boldsymbol{\theta}}$, in the limit of continuous time (i.e. to order dt), their instantaneous contemporaneous changes will be perfectly correlated.[1] In this Appendix, we first present the mechanics of deriving a continuous time multi-factor model. This rests on the application of results from continuous time analysis (mainly Itô's lemma), as clarified by Merton [1982], for example. We then make a comparison with the discrete time result.

Let us assume that the state variables follow continuous-time Markov stochastic processes:

$$(3.A.1) \quad \underline{d\theta}_{jt} = \alpha_j(\underline{\theta}_j, t)\ dt + s_j(\underline{\theta}_j, t)\ \underline{dz}_{jt} \qquad \forall\ j \in k$$

Here, $\underline{dz}_j$ is a one-dimensional basic Wiener or Brownian motion process with

$$(3.A.2) \quad \underline{dz}_{jt} = \underline{\varepsilon}_{jt}\sqrt{dt}$$

where $\underline{\varepsilon}_j$ is a zero-mean unit-variance stochastic variable. Hence, $\underline{d\theta}_j$ follows a generalized Wiener process (Itô process) where $\alpha_j(\underline{\theta}_j, t)$ is the instantaneous conditional expected change (drift coefficient) per unit time of $\underline{\theta}_j$ and $s_j(\underline{\theta}_j, t)$ is its instantaneous conditional standard deviation (diffusion coefficient) per unit time.[2]

We start from the expansion of the price function (3.2) in eqs.(3.4-5) and take the limit as $\Delta t \to dt$:

$$(3.A.3) \quad dP_t = \sum_j \frac{\partial P(\cdot)}{\partial \theta_j} d\theta_j + \frac{\partial P(\cdot)}{\partial t} dt + \tfrac{1}{2} \sum_i \sum_j \frac{\partial^2 P(\cdot)}{\partial \theta_i \partial \theta_j} d\theta_i d\theta_j + o(dt)$$

where the partial derivatives are evaluated at time t and where the

---

1) Cf. Merton [1982, p.32].

2) Mostly, it is explicitly assumed that $\underline{\varepsilon}_j$ follows a normal distribution. However, as Merton [1982, p.36] remarks, its distributional characteristics can be chosen almost arbitrarily, because in the limit of continuous trading, its change over a large number of infinitesimally trading intervals will be asymptotically normally distributed by the central limit theorem. So, "the generality gained by not making the [normality] assumption is more apparent than real" (Merton [1973, p.873 fn.16]).

asymptotic order symbol o(dt) denotes terms of smaller order than dt. As $\Delta t \rightarrow dt$, these terms can be neglected. In the expansion, the second order terms $d\theta_i d\theta_j$ have been retained because they have a term of order dt and thus cannot be ignored (Itô's lemma)[3]. This term equals $\underline{\varepsilon}_i\underline{\varepsilon}_j dt$, but as its variance is of order $(dt)^2$, the term becomes non-stochastic and equal to its expected value as $\Delta t \rightarrow dt$. Hence,

$$(3.A.4)\quad \underline{dz}_i\underline{dz}_j = \rho_{ij}dt$$

where ρ_{ij} is the instantaneous correlation coefficient between the Wiener processes $\underline{dz}_i$ and $\underline{dz}_j$.

Incorporating (3.A.1) in (3.A.3), recognizing that $\partial P(\cdot)/\partial t = 0$, and collecting terms now yields:

$$(3.A.5)\quad dP_t = \left[\sum_j \frac{\partial P(\cdot)}{\partial \theta_j} \alpha_j(\underline{\theta}_j, t) + \tfrac{1}{2} \sum_i \sum_j \frac{\partial^2 P(\cdot)}{\partial \theta_i \partial \theta_j} \rho_{ij}\, s_i(\underline{\theta}_i, t) s_j(\underline{\theta}_j, t) \right] dt + \sum_j \frac{\partial P(\cdot)}{\partial \theta_j} s_j(\underline{\theta}_j, t)\, d\underline{z}_j$$

The instantaneous security return then follows the process:[4]

$$(3.A.6)\quad \underline{dP}_t / P_t = \alpha(\underline{\theta}, t)\, dt + s(\underline{\theta}, t)\, \underline{dz}$$

with:

$$(3.A.7)\quad \alpha(\underline{\theta}, t) = \sum_j \frac{\partial P(\cdot)/P(\cdot)}{\partial \theta_j} \alpha_j(\underline{\theta}_j, t) + \tfrac{1}{2} \sum_i \sum_j \frac{\partial^2 P(\cdot)/P(\cdot)}{\partial \theta_i \partial \theta_j} \rho_{ij}\, s_i(\underline{\theta}_i, t) s_j(\underline{\theta}_j, t)$$

and

$$(3.A.8)\quad s(\underline{\theta}, t)\, \underline{dz} = \sum_j \frac{\partial P(\cdot)/P(\cdot)}{\partial \theta_j} s_j(\underline{\theta}_j, t)\, d\underline{z}_j$$

$\alpha(\underline{\theta}, t)$ and $s(\underline{\theta}, t)$ are the instantaneous expected return and the

---

[3] All other terms of $\underline{d\theta}_i\underline{d\theta}_j$ are o(dt) and can be neglected: $\underline{dz}_j dt = 0$ and $(dt)^2 = 0$. For Itô's lemma, see for example Merton [1971, pp.375-376] and Merton [1973, 1982].

[4] The log price change can be derived by multiplying both sides of eq.(3.A.6) with P_t and applying Itô's lemma again:

$$\underline{d\ln P}_t = [\, \alpha(\underline{\theta}, t) - \tfrac{1}{2} s(\underline{\theta}, t)^2 \,]\, dt + s(\underline{\theta}, t)\, \underline{dz}$$

instantaneous standard deviation of the security return, conditional on the levels of the state variables at time t. It can clearly be seen how the conditional expected and the unanticipated components of $\underline{\boldsymbol{\theta}}_t$ propagate into the respective components of the security return.

We can now translate this into a linear MFM in conditional expectations:

$$(3.A.9) \quad \underline{r}_t = E_t(\underline{r}_t) + \sum_j b_{jt}[\underline{d\theta}_{jt} - E_t(\underline{d\theta}_{jt})]$$

This expression resembles its discrete time linear analogon eq.(3.19), but its linearity is more apparent than real. This expression explicitizes that, although the price function $P(\cdot)$ may be non-linear, linearity is preserved by letting the sensitivities b_{jt} change **continuously** over time. In this way, eq.(3.A.9) describes tangency lines to the non-linear function $\underline{r}_t$ of $[\underline{d\theta}_{jt} - E_t(\underline{d\theta}_{jt})]$.

Forcing the sensitivities to be constant over time, $b_{jt} = b_j$, introduces an error term in (3.A.9) that contains the non-linear effects of the changes in $\underline{\boldsymbol{\theta}}_t$. Furthermore, it implies that b_j will depend on the (range in the) specific values that $\underline{\boldsymbol{\theta}}_t$ takes on. Considering only a subset of the state variables (partial conditioning) adds the (linear and non-linear) effects of omitted factors to the error term. The effects of non-stationarities and omitted variables are discussed in section 3.2.4.

APPENDIX 3.B: TAYLOR SERIES APPROXIMATION VERSUS LEAST SQUARES APPROXIMATION

This Appendix confronts the linear least squares approximation and the first order Taylor series approximation, as discussed in sections 3.2.2 and 3.2.3. This Appendix is self-contained for the reader's convenience.

Starting point is the general assumption that a security's return is generated by some function $\Phi(\cdot)$ of k factors $\underline{\boldsymbol{\delta}}$:

(3.B.1) $\underline{r}_t = \Phi(\underline{\boldsymbol{\delta}}_t) + \underline{\xi}_t$

The additive error term $\underline{\xi}$ accounts for the influence of various other factors. It has zero mean and constant finite variance and satisfies:

(3.B.2) $E(\underline{\xi}_t | \underline{\boldsymbol{\delta}}_t) = 0$

For brevity, we will omit the time subscripts in the following.

Two linear approximations

We can specify two linearized versions of eq.(3.B.1). A truncated first order Taylor series approximation around the expected values of the factors takes the form:

(3.B.3) $\underline{r} = a' + \sum_j b'_j \underline{\delta}_j + \underline{e}'$

where the term $\underline{e}'$ reflects the approximation error as well as the true stochastic error. The sensitivities are the gradients, evaluated at $E(\underline{\boldsymbol{\delta}})$:

(3.B.4) $$b'_j = \frac{\partial\Phi\{E(\underline{\boldsymbol{\delta}})\}}{\partial\underline{\delta}_j} \qquad (j=1,..,k)$$

and the intercept readily follows as:

(3.B.5) $$a' = \Phi\{E(\underline{\boldsymbol{\delta}})\} - \sum_{j=1}^{k} \frac{\partial\Phi\{E(\underline{\boldsymbol{\delta}})\}}{\partial\underline{\delta}_j} E(\underline{\delta}_j)$$

The (ordinary) linear least squares (OLS) approximation yields:

(3.B.6) $\underline{r} = a + \sum_j b_j \underline{\delta}_j + \underline{\varepsilon}$

where the zero-mean error $\underline{\varepsilon}$ includes both the approximation error and the true stochastic error $\underline{\xi}$:

$$(3.B.7) \quad \underline{\varepsilon} \equiv \Phi(\underline{\boldsymbol{\delta}}) - a - \textstyle\sum_j b_j \underline{\delta}_j + \underline{\xi}$$

In this approximation, the parameters a and $\{b_j\}_{j=1,..,k}$ are chosen to minimize $E(\underline{\varepsilon}^2)$. By standard statistical arguments, it follows that these parameters are the multiple regression coefficients and that

$$(3.B.8) \quad E(\underline{\varepsilon} \cdot \underline{\delta}_j) = 0 \qquad (j=1,..,k)$$

Expressions for the least squares coefficients follow from inserting eqs.(3.B.6) and (3.B.8) in the covariance between the security return and the factors:

$$(3.B.9) \quad \mathrm{Cov}(\underline{r}, \underline{\delta}_j) = \sum_{i=1}^{k} b_j \, \mathrm{Cov}(\underline{\delta}_i, \underline{\delta}_j) \qquad (j=1,..,k)$$

Solving each of these k equations for the i-th partial derivative at i=j yields:

$$(3.B.10) \quad b_j = \frac{\mathrm{Cov}(\underline{r}, \underline{\delta}_j) - \sum\limits_{i=1; i \neq j}^{k} b_i \, \mathrm{Cov}(\underline{\delta}_i, \underline{\delta}_j)}{\mathrm{Var}(\delta_j)} \qquad (j=1,..,k)$$

and for the intercept

$$(3.B.11) \quad a = E(\underline{r}) - \textstyle\sum_j b_j E(\underline{\delta}_j).$$

Expressing least squares coefficients in terms of gradients

In order to clarify the relationship between the first order Taylor series approximation (3.B.3) and the linear least squares approximation (3.B.6), we relate the linearized least squares MFM (which is imposed on the return) to the non-linear general MFM (3.B.1) (which drives the return). It is then necessary to make some additional assumptions on (3.B.1). We assume that the factors $\{\delta_j\}_{j=1,..,k}$ are jointly normally distributed and that they are uncorrelated with the error term:[1)]

---

1) Note that assumption (3.B.12) is weaker than (3.B.2). When the error term accounts for a multitude of other return generating factors, we could invoke the central limit theorem and defend the normality of the error term. In that case, the factors and the error would be independent. However, as the error term in eq.(3.B.1) is additive, (footnote continued on next page)

$$(3.B.12)\quad E(\underline{\xi}\cdot\underline{\delta}_j) = 0$$

We start again from the covariance between the return and factor j, but now on the basis of eq.(3.B.1):

$$(3.B.13)\quad \begin{aligned} Cov(\underline{r},\underline{\delta}_j) &= Cov\{\Phi(\underline{\boldsymbol{\delta}}),\underline{\delta}_j\} + Cov(\underline{\xi},\underline{\delta}_j) \\ &= Cov\{\Phi(\underline{\boldsymbol{\delta}}),\underline{\delta}_j\} \end{aligned}$$

Using the multivariate version of Stein's lemma[2), we can decompose this covariance as:

$$(3.B.14)\quad Cov(\underline{r},\underline{\delta}_j) = \sum_{i=1}^{k} E\left[\frac{\partial\Phi(\cdot)}{\partial\underline{\delta}_i}\right] Cov(\underline{\delta}_i,\underline{\delta}_j) \qquad (j=1,..,k)$$

Again, solving each of these k equations for the i-th partial derivative at i=j yields:

$$(3.B.15)\quad E(\partial\Phi/\partial\underline{\delta}_j) = \frac{Cov(\underline{r},\underline{\delta}_j) - \sum_{i=1;i\neq j}^{k} E(\partial\Phi/\partial\underline{\delta}_i)\ Cov(\underline{\delta}_i,\underline{\delta}_j)}{Var(\delta_j)} \qquad (j=1,..,k)$$

The eqs.(3.B.10) and (3.B.15) are consistent only when:

$$(3.B.16)\quad b_j = E\left[\frac{\partial\Phi(\cdot)}{\partial\underline{\delta}_j}\right] \qquad (j=1,..,k)$$

(and where the intercept a is defined as before in eq.(3.B.11)).
So the least squares coefficients represent the expected values of the gradients, $E[\Phi'(\cdot)]$, whereas the Taylor series coefficients equal the gradients evaluated at the spanning point, $\Phi'[E(\cdot)]$.

---

the linear independence assumption will suffice for our purpose. Also note that we do not impose any distributional restrictions on the returns.

2) When $\underline{x}$ and $\underline{y}$ are bivariate normally distributed and $g(\cdot)$ is some at least once differentiable function, then according to Stein's lemma we have under some mild regularity conditions that $Cov\{g(\underline{x}),\underline{y}\} = E\{g'(\underline{x})\}Cov(\underline{x},\underline{y})$. This covariance property appears disguised in Rubinstein [1973b, pp.613-614] and is derived explicitly by Rubinstein [1976, pp.421-422]. Losq & Chateau [1982, pp.784-786] have extended Stein's lemma to the multivariate version we use in the text:

$$Cov\{g(\underline{x}_1,\ldots,\underline{x}_k),\underline{y}\} = \sum_{j=1}^{k} E\{\partial g(\cdot)/\partial\underline{x}_j\}\ Cov(\underline{x}_j,\underline{y})$$

Relating the coefficients from both approximations

Comparing eqs.(3.B.4) and (3.B.16), we expect (not so much to our surprise) the least squares coefficients $\{b_j\}$ to differ from the Taylor series coefficients $\{b'_j\}$ whenever $\Phi(\cdot)$ is a non-linear function. When the signs of the third partial derivatives of $\Phi(\cdot)$ are persistent (i.e., the sign of each third derivative remains the same over all possible values of the corresponding factor), then it readily follows from Jensen's inequality that:

(i) $b_j \geq b'_j$ when $\partial^3\Phi(\cdot)/\partial\underline{\delta}_j^3 \geq 0$

(ii) $b_j \leq b'_j$ when $\partial^3\Phi(\cdot)/\partial\underline{\delta}_j^3 \leq 0$

for $j=1,..,k$. In case *(i)*, we have a positive third derivative everywhere, so the function $\Phi(\cdot)$ is characterized by increasing convexity (when $\Phi''(\cdot) \geq 0$) or decreasing concavity (when $\Phi''(\cdot) \leq 0$) in the direction of $\underline{\delta}_j$.[3] Hence, the corresponding least squares coefficient is greater than the Taylor series coefficient. In case *(ii)*, $\Phi(\cdot)$ is decreasingly convex (when $\Phi''(\cdot) \geq 0$) or increasingly concave (when $\Phi''(\cdot) \leq 0$) in the direction of δ_j, so the corresponding least squares coefficient is smaller than the Taylor series coefficient.[4]

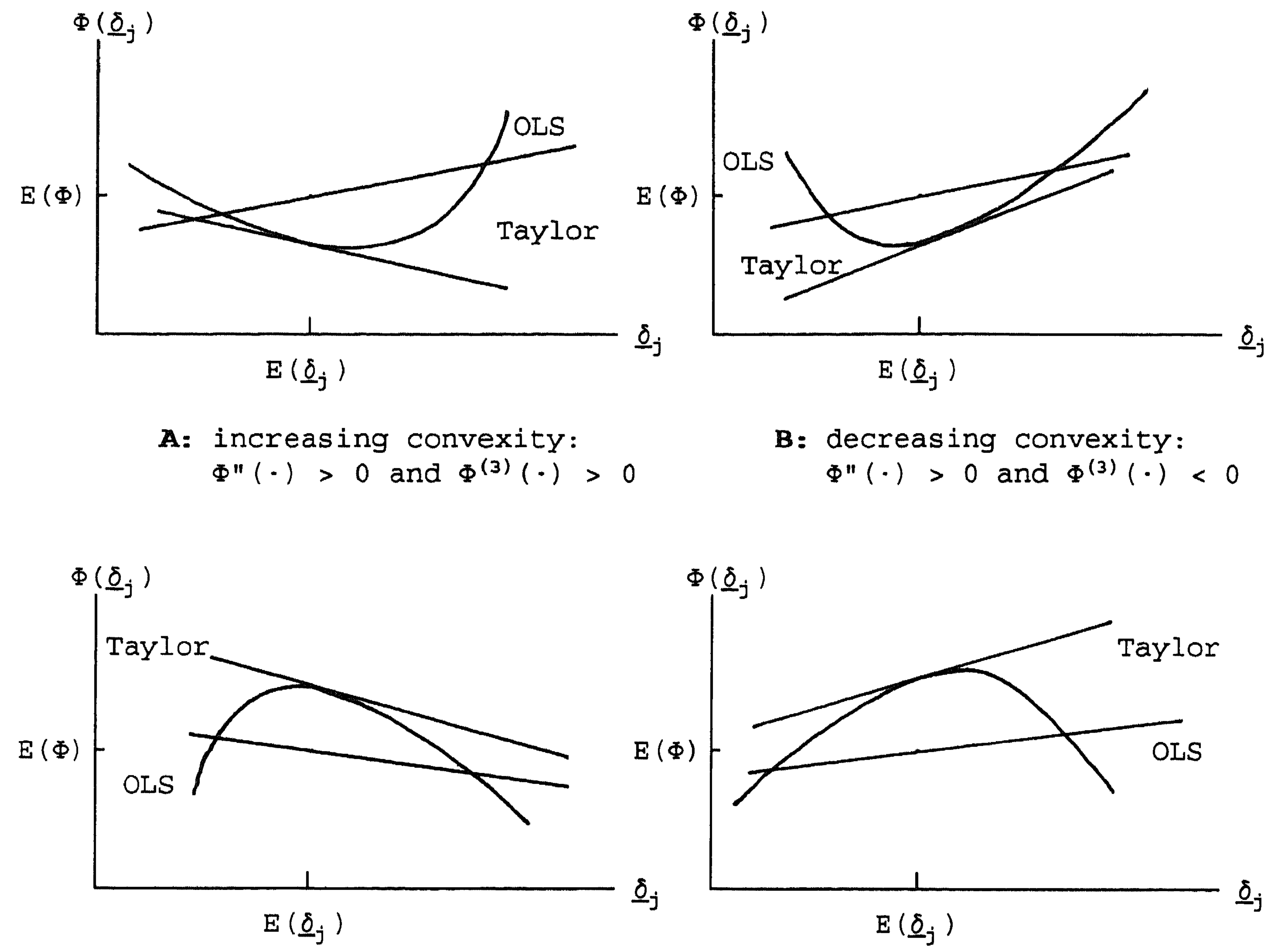
The four situations are depicted in Figure 3.B.1. Note that this Figure treats only situations of either concavity or convexity; the derived results are more general and also apply to **mixed concave / convex** functions. According to the assumed normality of the factor $\underline{\delta}_j$, its values are distributed symmetrically around its expected value. The latter value serves as spanning point for the Taylor series approximation. The OLS approximation, in contrast, is tilted towards the segment where the function exhibits relatively more convexity or concavity. Furthermore, the OLS approximation passes through the means $(E(\underline{\delta}_j), E\Phi(\cdot))$ whereas the Taylor approximation is tangent to the underlying function.

---

[3] Note that mixed convexity / concavity is allowed. When, in the case of $\Phi^{(3)}(\cdot) > 0$, we have that $\Phi''(\cdot) < 0$ switches to $\Phi''(\cdot) > 0$, this implies that decreasing concavity goes over in increasing convexity.

[4] When we limit ourselves to **strictly** increasing or decreasing convexity or concavity, then the equality signs in *(i)* and *(ii)* and in the text above must be omitted.

Figure 3.B.1: First order Taylor series approximation versus linear least squares (OLS) approximation. Univariate case for strictly increasing or decreasing convexity or concavity.



A: increasing convexity: $\Phi''(\cdot) > 0$ and $\Phi^{(3)}(\cdot) > 0$

B: decreasing convexity: $\Phi''(\cdot) > 0$ and $\Phi^{(3)}(\cdot) < 0$

C: decreasing concavity: $\Phi''(\cdot) < 0$ and $\Phi^{(3)}(\cdot) > 0$

D: increasing concavity: $\Phi''(\cdot) < 0$ and $\Phi^{(3)}(\cdot) < 0$

Bounds on the differences between the coefficients

We can illustrate and refine this result for the univariate case (thus also covering the multivariate case where the $\underline{\delta}_j$'s are independent). We start with the weak assumptions that the factor has a finite fourth moment and is uncorrelated with the (true) error term.
Incorporating a second order Taylor series expansion of $\Phi(\cdot)$ around $E(\underline{\delta}_j)$ in eq. (3.B.13) gives:

$$(3.B.17)\quad \mathrm{Cov}(\underline{r},\underline{\delta}_j) = \mathrm{Cov}\{\Phi(\underline{\delta}_j),\underline{\delta}_j\} = E\{\Phi(\underline{\delta}_j)(\underline{\delta}_j - E(\underline{\delta}_j))\}$$
$$= \mathrm{Var}(\underline{\delta}_j)\Phi'[E(\underline{\delta}_j)] + \tfrac{1}{2}E\{[\underline{\delta}_j - E(\underline{\delta}_j)]^3\}\cdot\Phi''[E(\underline{\delta}_j)]$$
$$+ (1/6)E\{[\underline{\delta}_j - E(\underline{\delta}_j)]^4\cdot\Phi^{(3)}[\underline{\vartheta}]\}$$

where $\underline{\vartheta}$ (a function of $E(\underline{\delta}_j)$ and $\underline{\delta}_j$) lies between $\underline{\delta}_j$ and $E(\underline{\delta}_j)$. Assuming that $\underline{\delta}_j$ follows a symmetric distribution, the third central moment $E\{[\underline{\delta}_j - E(\underline{\delta}_j)]^3\}$ is zero. Assuming further that the third derivative $\Phi^{(3)}(\cdot)$ is either uniformly positive or uniformly negative and that there exist numbers m_j and M_j such that either $-\infty < m_j < \Phi^{(3)}[\underline{\delta}_j] < 0$ or $0 < \Phi^{(3)}[\underline{\delta}_j] < M_j < \infty$, we can rewrite (3.B.17) as:

$$(3.B.18) \quad \Phi'[E(\underline{\delta}_j)] < \frac{\mathrm{Cov}(\underline{r}, \underline{\delta}_j)}{\mathrm{Var}(\underline{\delta}_j)} < \Phi'[E(\underline{\delta}_j)] + \frac{E\{[\underline{\delta}_j - E(\underline{\delta}_j)]^4\}}{6 \cdot \mathrm{Var}(\underline{\delta}_j)} M_j \qquad \text{for } \Phi^{(3)}[\underline{\delta}_j] > 0$$

and

$$(3.B.19) \quad \Phi'[E(\underline{\delta}_j)] > \frac{\mathrm{Cov}(\underline{r}, \underline{\delta}_j)}{\mathrm{Var}(\underline{\delta}_j)} > \Phi'[E(\underline{\delta}_j)] + \frac{E\{[\underline{\delta}_j - E(\underline{\delta}_j)]^4\}}{6 \cdot \mathrm{Var}(\underline{\delta}_j)} m_j \qquad \text{for } \Phi^{(3)}[\underline{\delta}_j] < 0$$

Again, the first term of (3.B.18-19) is the Taylor coefficient b'_j from (3.B.4) and the second term is the OLS slope estimate b_j from (3.B.10). In this more general case of a symmetric (but not necessarily normal) distribution we see that the relationship between b_j and b'_j is still governed by *(i)* and *(ii)* above. Moreover, the difference between the coefficients depends on the second and fourth moments of the independent variable and the rate of change in the convexity or concavity of $\Phi(\cdot)$.

For the more specific normal distribution, the fourth central moment equals $E\{[\underline{\delta}_j - E(\underline{\delta}_j)]^4\} = 3[\mathrm{Var}(\underline{\delta}_j)]^2$. Also, the OLS slope b_j is now as defined by (3.B.16). Using this information and the relationship between $E[\Phi'(\cdot)]$ and $\Phi'[E(\cdot)]$ conform Jensen's inequality, we get from eqs.(3.B.18-19):

(i') $\quad b'_j < b_j < b'_j + \frac{1}{2}\mathrm{Var}(\underline{\delta}_j) \cdot M_j \qquad$ for $\Phi^{(3)}[\underline{\delta}_j] > 0$

(ii') $\quad b'_j > b_j > b'_j + \frac{1}{2}\mathrm{Var}(\underline{\delta}_j) \cdot m_j \qquad$ for $\Phi^{(3)}[\underline{\delta}_j] < 0$

From this result, we see that the discrepancy between the two coefficients depends on the variance of the independent variable and the rate of change in the convexity and/or the concavity of $\Phi(\cdot)$.

This is consitent with White [1980, p.154]. At the one hand his analysis is more restrictive because he considers functions that are either concave or convex, thus requiring uniform signs of the second derivatives. At the other hand his analysis is more general in that he allows the distribution to be skewed. In general, skewness will always introduce inconsistency between the two coefficients.

Resume

We can summarize these results as follows. When the return generating process is non-linear, we will make an approximation error by using a linearized MFM. This approximation error very likely generates discrepancies between the gradients of the Taylor series approximation and the linear least squares coefficients. Under the assumptions made (especially the normality of the factors), we showed that the OLS coefficients can be interpreted as the expected values of the corresponding gradients of the unknown function. The discrepancies between the two sets of coefficients are due to non-linearities related to the third and higher order (partial) derivatives of $\Phi(\cdot)$, and (because of the assumed symmetry of the distribution) not to its second order derivatives. This implies that not concavity or convexity per se, but **differences** in concavity and/or convexity over the ranges of the independent (i.e. state) variables are relevant.

The results concerning the relationship between the global least squares coefficients and the local Taylor coefficients can be generalized in an approximate way to the less restrictive setting of symmetric distributions for the independent variables (factors).[5]

[5] It is difficult to say exactly how the interpretation of the OLS coeffcients as the expected values of the gradients carries over to the situation where the factors are distributed approximately normal, or where they follow (finite-moment) symmetric distributions. Here, we do not explore this issue further.

APPENDIX 3.C: A SIMPLE FLEXIBLE GROWTH MODEL

Ideally, we are looking for a present value model formulation that gives enough flexibility to incorporate a realistic pattern of dividend growth rates without at the same time becoming untractable. Desirable features of such a model would be:

- the possibility to incorporate investment analyst's dividend forecasts over the next, say, 3 to 5 years;
- the possibility to specify a transition phase in which the growth rates gradually decline (incline) until some 'steady state' economy-wide growth rate is reached.

The restrictions for a suchlike formulation would be:

- the description of the growth rate curve by means of only a limited number of parameters;
- the possibility to derive a simple closed-form expression for the present value formula.

In this Appendix, we propose a simplified flexible growth model. In the derivation, we assume a constant risk adjusted discount rate. Later, we will restate the result for the certainty equivalent framework.

Growth patterns and the stock price

We assume that dividend forecasts are available over the next T periods. After time T, it is expected that the dividend growth rate has reached the rate of growth in the economy as a whole. Dividend growth will continue from then in perpetuity at that 'normal' rate g. The current (theoretical) stock price then is:

$$P_0 = \sum_{t=1}^{T} \frac{D_0\,(1+g_t)^t}{(1+k)^t} + \sum_{t=T+1}^{\infty} \frac{E_0\,(\underline{D}_T)\,(1+g)^t}{(1+k)^t} \qquad \text{for } k > g \tag{3.C.1}$$

Figure 3.C.1: The curve of projected dividend growth rates g_t.
g^* = above normal growth rate; g = normal growth rate.

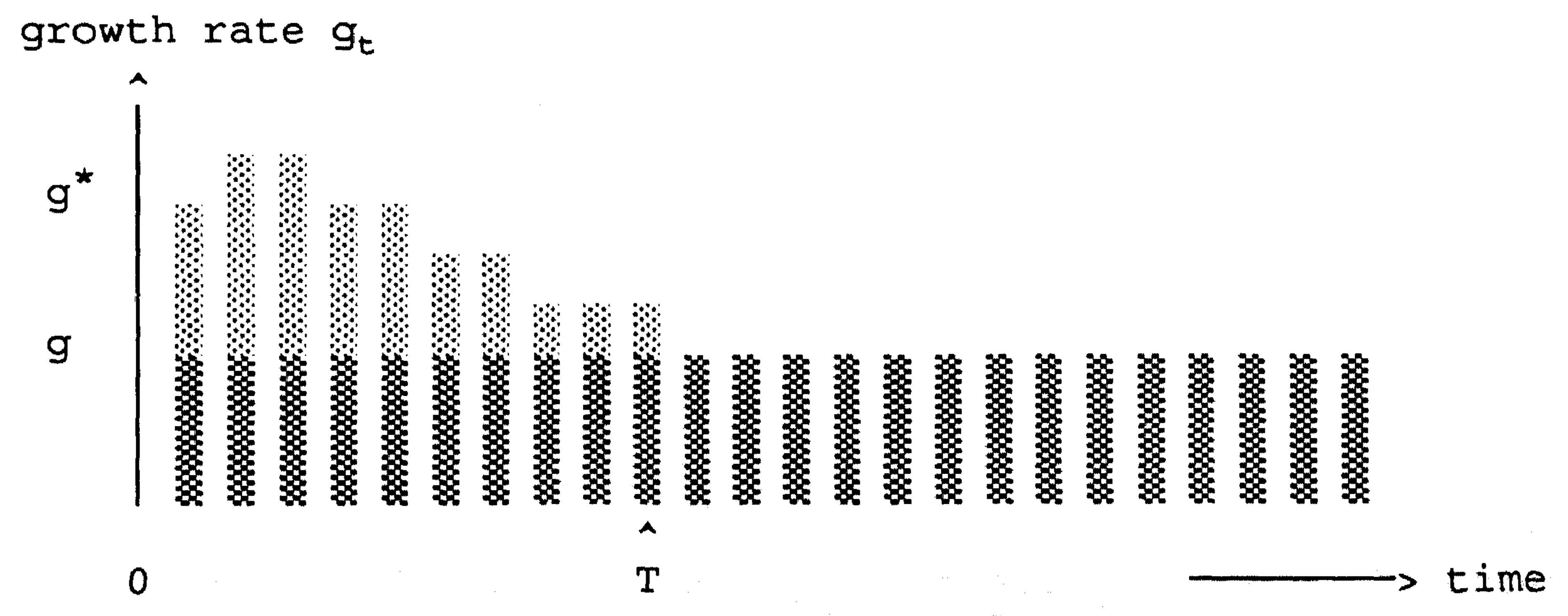


Figure 3.C.1 shows an example of the growth rates over time. These growth rates are defined by:

(3.C.2) $$D_t/D_0 = (1 + g_t)^t$$

For notational ease, expected dividends are denoted by D.

A simplified representation

We assume that the specific dividend forecasts or growth rates are known for the above normal growth period [0,T]. The basic idea, now, is to use one single forward growth rate g^* as a **spanning point** for all other growth rates during the period of above normal growth [1,T]. This reference rate can be specified exogenously, or estimated from the growth rates $\{g_t\}_{t=1,\ldots,T}$, implied by the dividend forecasts. Next, any abnormal growth rate g_t can be decomposed in the normal growth rate g and the effects of excess growth g^*-g. For period 1 we have:

(3.C.3) $$D_1 = (1 + g_1)\ D_0 = (1 + g)\ D_0 \left[1 + \frac{g_1 - g}{1+g} \right]$$

$$= (1 + g)\ D_0 \left[1 + d_1 \frac{g^* - g}{1+g} \right]$$

where d_1 denotes the relative excess growth multiplier for period 1. Eq.(3.C.3) can now be solved for d_1:[1)]

(3.C.4) $$d_1 = \frac{g_1 - g}{g^* - g}$$

The first dividend can now be decomposed as:

(3.C.5) $$D_1 = (1 + g)\ D_0 + (1 + g)\ D_0\ d_1 \left[\frac{g^* - g}{1+g} \right]$$

The first term on the RHS is the next period's expected dividend under normal growth; the second term reflects the above normal growth in period 1.

For any period $t \leq T$, we extend this as follows:

1) Using, for example, the first growth rate as the reference rate for above normal growth ($g^* = g_1$), it follows that d_1=1.

$$(3.C.6)\quad D_t = (1+g_t)^t D_0 = (1+g)^t D_0 \left[1 + \frac{g_t - g}{1+g}\right]^t$$

$$= (1+g)^t D_0 \left[1 + \sum_{\tau=1}^{t} d_\tau \frac{g^* - g}{1+g}\right]$$

$$= (1+g)^t D_0 + \sum_{\tau=1}^{t} (1+g)^t D_0\, d_\tau \frac{g^* - g}{1+g}$$

Starting from t=1, the relative excess growth multipliers $\{d_t\}_{t\le T}$ can now be obtained recursively.

From eq.(3.C.6), we observe that each period's expected dividend consists of a 'normal' component $(1 + g)^t D_0$ that grows at the normal rate g in perpetuity. Furthermore, in each period $t\le T$ an 'excess growth' dividend component $(1+g)^{t-1} D_0\, d_t (g^*-g)$ is introduced and added to the dividend stream. The latter dividend component is a function of the excess growth rate (g^*-g) and grows at the normal rate g in perpetuity. The present value of the excess growth dividend component introduced in period $t\le T$ is:

$$(3.C.7)\quad \frac{1}{(1+k)^t} \sum_{\tau=1}^{\infty} \frac{D_0 d_t (g^*-g)(1+g)^{t+\tau-1}}{(1+k)^{\tau-1}}$$

$$= \frac{\dfrac{(1+g)^{t-1}}{(1+k)^{t-1}} D_0 d_t (g^*-g)}{k - g}$$

The current stock price now consists of the present value of the infinite series of normal dividend components, plus the sum of the present values of each of the T excess growth dividend components as given by (3.C.7):

$$(3.C.8)\quad P_0 = \frac{(1+g)D_0}{k-g}\left[1 + \frac{g^*-g}{1+g} \sum_{t=1}^{T} d_t \left[\frac{1+g}{1+k}\right]^{t-1}\right]$$

Defining the weighted sum of the relative excess growth multipliers:

$$(3.C.9)\quad \Gamma \equiv \sum_{t=1}^{T} d_t \left[\frac{1+g}{1+k}\right]^{t-1} ,$$

and the adjustment factor G for the initial dividend,

$$(3.C.10) \qquad G \equiv \Gamma \frac{g^{*}-g}{1+g}$$

eq.(3.C.8) can be reformulated as:

$$(3.C.11) \qquad P_0 = \frac{(1+g)\, D_0\, \{1+G\}}{k-g} \qquad \text{for } k > g$$

We see that the original dividend stream is transformed into an equivalent stream by adjusting the first (expected) dividend. The transformed stream grows in perpetuity at the normal rate g; it is equivalent to the original stream one because it implies the same stock price.[2)]

An approximation to the discount rate k

Eq. (3.C.8) is exact, but still quite impressive. One approximation that could be applied is to choose a relatively high reference rate $g^{*} \geq g_t$ ($t \leq T$) and to restate eq. (3.C.6) as:

$$(3.C.6') \qquad D_t = (1+g_t)^t\, D_0 = (1+g)^t\, D_0 \left[1 + \frac{g^{*}-g}{1+g} \right]^{\sum^{t}_{\tau=1} d_\tau}$$

$$\sim (1+g)^t\, D_0 + \sum_{\tau=1}^{t} (1+g)^t\, D_0\, d_\tau\, \frac{g^{*}-g}{1+g}$$

Using logarithms, the excess growth multipliers d_t can readily be computed recursively from the equalities in the first line. To obtain the second line, we used the first term of a binomial expansion to get rid of the exponent. As $(g^{*}-g)/(1+g)$ will be small, we expect this to be a reasonable approximation. In particular, because of the choice of g^{*}, this will yield a small understatement of the d_t's.

Next, Γ is approximated as the **unweighted** sum of the excess growth multipliers:

2) Fuller & Hsia [1984] introduced a simplified present value formula that approximates a standard three-phase discount model and that bears resemblence with our formula (3.C.11). Our variable Γ equals in their model the number of H periods from now through halfway the transition period in the three-phase model. As they do not allow for more flexibility in dividend growth, their 'H-model' could be seen as a special case of our model. However, as Fuller & Hsia do not present the derivation of their model, we cannot decide on this issue.

$$\text{(3.C.9')} \qquad \Gamma = \sum_{t=1}^{T} d_t$$

As g < k for convergence, the terms between square brackets in eq. (3.C.9) are smaller than one. Hence, neglecting the weights will yield an overstatement of Γ. As the approximation errors in the d_t's and Γ work in opposite directions, they tend to compensate each other. When starting from the actual stock price, using the two-step approximated Γ in eq.(3.C.10) and the approximated G in (3.C.11) will yield an approximation to k. From (3.C.11), it readily follows that:

$$\text{(3.C.12)} \qquad 1 + k = (1 + g) \cdot [1 + \delta(1 + G)]$$

where $\delta \equiv D_0/P_0$ is the curent dividend yield of the stock. The equality sign must be replaced by '$\sim$' when the approximated value of G is used. Some calculations showed indeed that the approximation is adequate for a wide range of growth patterns.

An approximation to the stock price elasticities

Although the resulting stock price formula (3.C.11) is relatively simple, we see that the components $\{d_t\}_{t \leq T}$ and Γ, and hence G, are functions of g and k. For deriving simple, closed-form expressions for stock price elasticities, it is desired that the correction factor for the first dividend does not depend on the normal growth rate and the discount rate. We therefore introduce approximations by neglecting the functional relationship between G and k and g and assuming that:

$$\text{(3.C.13)} \qquad dG/dg = 0 \qquad \text{and} \qquad dG/dk = 0$$

The derivations in section 3.5.3 then readily follow.

The certainty equivalent case

In the certainty equivalent framework, the current stock price can be expressed as:

$$\text{(3.C.1')} \qquad P_0 = \sum_{t=1}^{T} \frac{D_0 (1+g'_t)^t}{(1+r)^t} + \sum_{t=T+1}^{\infty} \frac{CE_0(\underline{D}_T)(1+g')^t}{(1+r)^t} \qquad \text{for } r > g'$$

Hence,

$$\text{(3.C.11')} \qquad P_0 = \frac{(1 + g')\, D_0 \,\{1 + G'\}}{r - g'} \qquad \text{for } r > g'$$

For equivalence with the risk adjusted discount rate approach, the certainty equivalent growth rates g′ are defined as $(1+g'_t) = [(1+g_t)\cdot(1+r)/(1+k)]$. This implies that G′ = G. However, for providing consistency with the general certainty equivalent framework and for indicating a certainty equivalent context, we will use G′ in this case.

For computing elasticities, also in this case we introduce approximations by neglecting the functional relationship between G′ and r and g′ and assuming that:

(3.C.13′) $dG'/dg' = 0$ and $dG'/dr = 0$

APPENDIX 3.D: MODELLING NON-LINEARITIES WITH FIRST ORDER ELASTICITIES

As a first order elasticity, duration is often applied to relate changes in bond or stock prices to changes in the discount rate or 'the' interest rate. As the relationship between price and discount rate is non-linear, convexity (and even hyper-convexity) characteristics can be used to improve the linear first order approximation. This Appendix draws on Hallerbach [1994a] and proposes an alternative approach by deriving a rational approximation. While our specification is based only on a first order elasticity, it implicitly uses quasi-convexity and quasi-hyperconvexity characteristics. These 'quasi-convexities' greatly improve the accuracy of the approximation and provide exact results for perpetuities as well as for one-period cash flows. Starting from the Taylor series approximation, we derive a rational approximation, discuss its accuracy and implications, and present applications.

The standard Taylor series approximation

Consider the continuous and analytic price function B(r), expressing the current price of security B as a function of the discretely compounded discount rate ('the' interest rate) $r>0$.[1] As there is only one argument, the Taylor series expansion eq.(3.4) of the price function B(r') when the initial rate r changes to r' becomes:

$$(3.D.1) \qquad B(r') = \sum_{i=0}^{\infty} (1/i!)(\Delta r)^i B^{(i)}(r)$$

with $\Delta r \equiv r'-r$, and where the superscripts between brackets denote the derivatives of the price function with respect to r, evaluated at r. Subtracting B(r) from both sides of eq.(3.D.1) and dividing by B(r) yields the univariate analogon of eq.(3.6), i.e. the instantaneous return as a result of a change in the discount rate:

$$(3.D.2) \qquad \frac{\Delta B(r)}{B(r)} = \sum_{i=1}^{\infty} \phi_i \left[\frac{\Delta r}{1+r}\right]^i$$

where ϕ_i denotes the i-th order elasticity of the price B with respect to r:[2]

---

1) This mathematical representation involves no causality. Relevant is only that the discount rate r is consistent with the current price and the magnitude and the timing of the cash flows. The price function actually is a present value function.

2) Alternatively, ϕ_i is the sensitivity parameter (or response coefficient) of the **return** for i-th order relative changes in one plus the rate r.

$$(3.D.3) \qquad \phi_i \equiv (1/i!)\, B^{(i)}(r)\, \frac{(1+r)^i}{B(r)} \qquad i \geq 1$$

The Macaulay [1938]-duration D equals $-\phi_1$, the convexity ½C equals ϕ_2 and the hyper-convexity $^1/_6$H is $-\phi_3$, &c. Hence:

$$(3.D.4) \qquad \Delta B/B = -D\cdot[\Delta r/(1+r)] + ½C\cdot[\Delta r/(1+r)]^2 - {}^1/_6H\cdot[\Delta r/(1+r)]^3 + \ldots$$

Considering the intrinsic non-linearity of a price function in the discount rate, one can only expect a reasonable accurate approximation for small rate changes when using only the first order elasticity. For evaluating price changes when rate changes are larger, higher order elasticities must be used to improve the first order approximation. An alternative, however, is to use a first order elasticity in a simple non-linear form, derived from a rational approximation.

Rational approximation and 'quasi-convexity'

In many cases, for a given amount of computational effort, a function can be approximated with greater accuracy by the use of rational functions, rather than by the use of polynomials. We here apply the Padé rational approximation to the price function.[3)]

Under a Padé rational approximation of order (m,k), the approximate representation of the price function B(r') with $r' \equiv r+\Delta r$ takes the form of the rational function R(r'):

$$(3.D.5) \qquad R(r') = R_{m,k}(r') = \frac{P_m(r')}{Q_k(r')}$$

where $P_m(r')$ and $Q_k(r')$ are polynomials of degree m and k, respectively:

$$(3.D.6) \qquad \begin{aligned} P_m(r') &= a_0 + a_1(r'-r) + \ldots + a_m(r'-r)^m \\ Q_k(r') &= b_0 + b_1(r'-r) + \ldots + b_k(r'-r)^k \end{aligned}$$

We assume that $P_m(r')$ and $Q_k(r')$ have no common zero and are reduced to their lowest degree by the cancellation of common factors. Clearly, we require that $Q_k(r') \neq 0$. More in particular, we require that $b_0 \neq 0$ in

---

3) Cf. for example Ralston [1965, Ch.7.3] or Young & Gregory [1973, Ch.6.12]. Other applications of the Padé rational approximation include Ijiri [1972], who provides approximations to compounding and discounting formulas, and Phillips [1982], who introduces the technique in econometrics to approximate probability density functions. Hallerbach & Kremer [1993] consider the rational approximation in exploratory data analysis, as an alternative to neural networks.

order to prevent singularities for $(r'-r) = 0$. Without loss of generality, we can normalize by setting $b_0 = 1$.[4]

The Padé rational approximation can be seen as a kind of generalization of a Taylor series expansion. Given m, k and r, the (m+k+1) free coefficients of $P_m(r')$ and $Q_k(r')$ are chosen such that at the spanning point r:

1. their values are equal: $R(r) = B(r)$; and
2. the first (m+k) derivatives of R(r) are equal to the corresponding derivatives of B(r): $R^{(j)}(r) = B^{(j)}(r)$ for $j=1,\ldots,m+k$.

This implies that the coefficients of the polynomials can be chosen such that the constant term and the terms involving $(r'-r)$, $(r'-r)^2$, ... , $(r'-r)^{m+k}$ all vanish in the following expression:[5]

$$(3.D.7) \qquad \sum_{i=0}^{\infty} (1/i!)(r'-r)^i B^{(i)}(r) Q_k(r') - P_m(r')$$

It is suggested that the most accurate approximations are obtained for rational fractions with numerator and denominator of (nearly) equal degrees. As we look for an approximation of price changes involving only duration, we choose a (0,1)-order Padé approximation.[6] For $m = 0$, $k = 1$ and $b_0 = 1$, eq.(3.D.7) becomes:

$$(3.D.8) \qquad [B(r) + \Delta r B'(r) + \tfrac{1}{2}(\Delta r)^2 B''(r) + \ldots.][1 + b_1(\Delta r)] - a_0$$

Collecting the constant terms and the terms involving Δr, we have:

$$(3.D.9) \qquad [B(r) - a_0] + \Delta r[B(r)b_1 + B'(r)] + (\Delta r)^2[\ \ldots..\ ] + \ldots..$$

From this equation, we see that $B(r')-R(r')$ and its first derivative (the terms involving Δr) vanish when:

$$(3.D.10) \qquad \begin{aligned} a_0 &= B(r) \\ b_1 &= -B'(r)/B(r) \end{aligned}$$

Substitution of eq.(3.D.10) in (3.D.9) finally gives the approximation for the price when the discount rate changes:

---

[4] Other ways to normalize the rational parameters is to set $\sum_j b_j^2 = 1$ or $b_k = 1$.

[5] The (m+k+1) free coefficients are thus determined by:

$$\sum_{i=0}^{\infty} (1/i!)(r'-r)^i B^{(i)}(r) Q_k(r') - P_m(r') = O((r'-r)^{m+k+1})$$

[6] Note that a (m,0)-order Padé approximation is equivalent to a (m)-order Taylor series approximation.

$$(3.D.11)\qquad B(r') \approx \frac{B(r)}{1 - [B'(r)/B(r)]\Delta r}$$

Dividing both sides of (3.D.11) by the current price B(r) and recognizing the duration $D = -\phi_1$ yields the expression of one plus the instantaneous security return:

$$(3.D.12)\qquad \frac{B(r')}{B(r)} = 1 + \frac{\Delta B}{B} \approx \frac{1}{1 + D\cdot[\Delta r/(1+r)]}$$

The principle of the derivation can be reversed by expanding (3.D.12). Recognizing the sum formula for an infinite geometric series with common ratio $-D\Delta r/(1+r)$, we can rewrite (3.D.12) as:

$$(3.D.13)\qquad \Delta B/B \approx -D\cdot[\Delta r/(1+r)] + D^2\cdot[\Delta r/(1+r)]^2 - D^3\cdot[\Delta r/(1+r)]^3 + \ldots$$

Comparing eq.(3.D.13) with (3.D.4), we see that the approximation formula (3.D.12) implicitly uses D^2 as a measure for convexity $\frac{1}{2}C$, D^3 as a measure for hyper-convexity $\frac{1}{6}H$, and so on. In the approximation, eq.(3.D.3) is thus replaced by $\phi_i = \{-D\cdot(1+r)\}^i$. As all convexity measures are exact functions (powers) of the duration D, we call these measures **'quasi-convexities'**. These quasi-convexities allow for a close approximation of the convex response of the security price to changes in the discount rate.[7)]

As an intuitive digression on the derivation of eq.(3.D.12), consider the following. A Taylor series expansion of a function F(x) for large values of the argument x will provide poor approximations when only a limited number of terms is used. If F(x) is to be approximated for large x, $x \neq 0$, we can substitute $z \equiv 1/x$ in the function and construct a Taylor series expansion for the function $G(z) = F(1/z)$ instead. For large x, z will be very small, so that only few terms in the expansion will be sufficient to arrive at a comforting degree of accuracy.

For the approximation of the price function B(r') we can apply a similar procedure. Note that the price fuction B(r) is a rectangular hyperbola in (1+r), but a polynomial in 1/(1+r). For the approximation, we start from the reciprocal of the function B(r). A first order Taylor series expansion of $B(r')^{-1}$ around r takes the form:

$$(3.D.14)\qquad B(r')^{-1} \approx B(r)^{-1} + B(r)^{-1}\cdot D\cdot\Delta r/(1+r)$$

where D is duration as before. Multiplying both sides with B(r) and taking the reciprocals of both sides, we finally get eq.(3.D.12). The

7) For the sake of completeness, we mention that eq.(3.D.12) has an asymptote for $\Delta r = -(1+r)/D$; in general practice, this will not be relevant.

(0,1)-Padé approximation can thus be seen as a reciprocal first order Taylor series expansion.

Furthermore, denoting the instantaneous return $\Delta B/B$ as r_B, we can rewrite eq.(3.D.12) as:

$$(3.D.15) \quad r_B/(1+r_B) \sim - D\cdot\Delta r/(1+r)$$

This transformation of the approximation is now linear in duration (as indeed is the actual relation between $r_B/(1+r_B)$ and $\Delta r/(1+r)$, approximately).

The accuracy of the approximation

For fixed cash flow streams at either end of the maturity spectrum (perpetual or one-period), the discount rate elasticities of all orders are exact powers of duration. As a result, eq.(3.D.12) applies exact, for example for the constant dividend growth model. Although the flexible growth model, as derived in Appendix 3.C, is an infinite-horizon model, eq.(3.D.12) will not yield exact results because of the relationship between the dividend adjustment factor G and the discount rate. However, the approximation will be very good.

For intermediate maturity ranges, a security's actual and quasi-convexity characteristics will not coincide. Consequently, the formula will no longer yield exact results. However, the approximation is very good and for some cases even superior to the second order Taylor series (duration and convexity) approximation.[8]

Applications and implications

Eq.(3.D.12) or (3.D.15) can be used to relate price changes to changes in the discount rate. Relating to the discussion in Appendix 3.B, either gradients evaluated in the spanning point can be used, or (non-)linear regression estimation can be applied to eq.(3.D.12) or (3.D.15).[9]
In fact, as the approximation curve follows the curve of the actual non-linear relationship quite closely, we expect that the least squares estimation of the first order elasticity in eq.(3.D.12) will be much less sensitive to the particular range of discount rates in the sample. Indeed, the ability of the approximation to follow the actual

---

8) For details, we refer to Hallerbach [1994a]. We there also derive a (1,1)-rational approximation. Given the knowledge of duration as well as convexity, the approximation of instantaneous returns on finite-maturity bonds is then virtually exact. This applies even for the case of major yield changes, where the conventional second order approximation deteriorates very rapidly.

9) Note, however, that using eq.(3.D.15) in a linear regression context implies a specific assumption about the error term.

relationship does not come from a changing gradient, but stems from the non-linear use of a constant first order elasticity.

The approximation eq.(3.D.12) will be applied in chapter four, section 4.3.5, for modelling the asymmetry in interest rate sensitivities of stock returns.