A Multiple Factor Model for European Stocks

THOMAS G. STEPHAN*, RAIMOND MAURER AND MARTIN DÜRR*****

*) Deutscher Investment Trust (DIT),
D-60329 Frankfurt/Main, Mainzer Landstr.11-13, Germany
Telephone: 49 69 263 14306
Facsimile: 49 69 263 14937
E-Mail: thomas.stephan@dit.de

**) Johann Wolfgang Goethe University of Frankfurt/Main,
Chair for Investment, Portfolio Management and Pension Systems
D - 60054 Frankfurt/Main, Mertonstrasse 17, Germany
Telephone: 49 69 798 25227
Facsimile: 49 69 798 25228
E-mail: Rmaurer@wiwi.uni-frankfurt.de

***) Deutscher Investment Trust (DIT),
D-60329 Frankfurt/Main, Mainzer Landstr.11-13, Germany
Telephone: 49 69 263 14555
Facsimile: 49 69 263 14972
E-Mail: martin.duerr@dit.de

Abstract

We present an empirical study focusing on the estimation of a fundamental multi-factor model for a universe of European stocks. Following the approach of the BARRA model, we have adopted a cross-sectional methodology. The proportion of explained variance ranges from 7.3% to 66.3% in the weekly regressions with a mean of 32.9%. For the individual factors we give the percentage of the weeks when they yielded statistically significant influence on stock returns. The best explanatory power – apart from the dominant country factors – was found among the statistical constructs „success“ and „variability in markets“.

1. Introduction

Multiple factor models attempt to describe asset returns and their covariance matrix as a function of a limited number of risk attributes. Factor models are thus based on one of the basic tenets of financial theory: no reward without risk. In contrast to the Capital Asset Pricing Model (CAPM) first presented by SHARPE (1964), LINDNER (1965) and MOSSIN (1966) that uses the stock beta as the only relevant risk measure, empirical studies could not confirm this very restrictive statement. For instance, FAMA/FRENCH (1992) found price to book value ratio and market capitalisation (in contrast to beta) to have significant influence on stock returns. The Arbitrage Pricing Theory presented by ROSS (1976) already posited a more general multiple factor structure for the return generating processes. However, it neither specified the nature nor the number of these factors.

Starting with the studies of ROSENBERG, multiple factor models have been applied early in investment practice, mainly because they allow a differentiated risk-return-analysis. The applications of multiple factor models are various and are based on the analysis and prognosis of portfolio risk. Multiple factor models can give valuable insights especially in performance and risk attribution. They provide – used prospectively – for a better basis for portfolio construction (because of an improved risk prognosis) as well as the basis for deliberate deviations from a benchmark portfolio (cf. ALBRECHT/MAURER/MAYSER 1996).

The objective of this study is the conception and empirical application of a fundamental multiple factor model for a universe of 652 European stocks using fundamental descriptors. The model is estimated according to a cross-sectional approach and its explanatory power is verified over time. The article is structured as follows: The theoretical foundations that are relevant for the conception of a multiple factor model are presented in section 2. In section 3 we present the basic statistical techniques for the identification of multiple factor models. The underlying data, the methodology and the findings of the empirical work are presented in section 4. The conclusion in section 5 summarises the results.

2. The general structure of Multiple factor models

In the general form multiple factor models posit that the period returns of the different assets are "explained" by common factors in a linear model. The asset returns are influenced by the factors according to the sensitivity or exposure of a specific security to these factors. These sensitivities thus play the role of beta in the CAPM, which represents the exposure of a security to a whole-market-factor. In addition, the asset return is influenced by another component, the so-called specific return, which is assumed to be independent of the factor returns. A multiple factor model for the $i = 1, \ldots, n$ relevant securities of a market thus looks as follows:

$$R_i = \alpha_i + \beta_{i1} \cdot F_1 + \ldots + \beta_{ik} \cdot F_k + \varepsilon_i. \tag{1}$$

where

R_i	return to security i,
α_i, β_{ij}	sensitivity/exposure of security i to factor j,
$F_1, \ldots, F_k$	the *k* factors,
ε_i	specific return to security i

The variables R_i, $F_1, \ldots, F_k$ and ε_i are random variables. The variance of ε_i is denoted by σ^2_i, the covariance matrix of the factors by Φ. In order to make this model a useful instrument, however, some assumptions must be fulfilled. The most important is that the specific returns $\varepsilon_1, \ldots, \varepsilon_n$ are not correlated amongst each other. This implies that the correlation between the returns of two different securities is solely determined by their common dependence on the factors $F_1, \ldots, F_k$. It will appear later that this assumption makes the estimation of Σ, the covariance matrix of the securities, much easier. Another assumption is that the expected specific return $E(\varepsilon_i)$ is zero, i.e. that the entire expected asset return which is not caused by the factor returns is comprised in the component α_i. Finally it is assumed that the specific returns are independent of the factors.

Equation (1) holds in period *t*. The corresponding index has been left out for simplicity. It is important to mention, however, that all terms of equation (1) can change over time. This is obvious for the asset returns and the factors. For the time being we will also allow the sensitivities to take on different values in different periods.

In financial theory the covariance matrix of the asset returns Σ is of specific interest. We will therefore look at equation (1) in matrix notation to see the functionality for all assets at one time. R is the random vector of the returns with the elements $R_1, ..., R_n$. As the sensitivities are often referred to as loadings of the securities on the respective factors the matrix of these exposures is denoted by $\widetilde{L}$. The matrix $\widetilde{L}$ looks as follows:

$$\widetilde{L} = \begin{bmatrix} \alpha_1 & \beta_{11} & \cdots & \beta_{1k} \\ \vdots & \vdots & & \vdots \\ \alpha_n & \beta_{n1} & \cdots & \beta_{nk} \end{bmatrix}. \tag{2}$$

The factors are represented by the random vector $\widetilde{F}$ with the dimension $k + 1$ whereas the first element F_0 takes the constant value 1. The remaining elements are the random variables $F_1, ..., F_k$. The random vector of the specific returns is denoted by ε. As the specific returns are not correlated amongst each other, the covariance matrix of ε is a diagonal matrix that is denoted by Ω. The elements on the diagonal of Ω are $\sigma^2_{\varepsilon 1}, ..., \sigma^2_{\varepsilon n}$, the variances of the specific returns. Using these notations, the matrix notation of equation (1) becomes:

$$R = \widetilde{L} \cdot \widetilde{F} + \varepsilon, \tag{3}$$

or in detail:

$$\begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix} = \begin{bmatrix} \alpha_1 & \beta_{11} & \cdots & \beta_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_n & \beta_{n1} & \cdots & \beta_{nk} \end{bmatrix} \cdot \begin{pmatrix} 1 \\ F_1 \\ \vdots \\ F_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \tag{4}$$

Consider the covariance matrix of R: $\Sigma = \mathrm{Cov}(R) = \mathrm{Cov}(\widetilde{L} \cdot \widetilde{F} + \varepsilon)$ $= L \cdot \mathrm{Cov}(F) \cdot L^T + \mathrm{Cov}(\varepsilon) = L \cdot \Phi \cdot L^T + \Omega$. The matrix L is merley the matrix $\widetilde{L}$ with the first column (the α_i's) omitted. The random vector F is the vector $(F_1, ..., F_k)^T$. F_0 being constant, we get: $\mathrm{Cov}(\widetilde{L} \cdot \widetilde{F}) = L \cdot \mathrm{Cov}(F) \cdot L^T$. Under the assumptions of our multiple factor model the covariance matrix Σ thus becomes:

$$\Sigma = L \cdot \Phi \cdot L^T + \Omega. \tag{5}$$

It should be remarked that the assumptions of the model usually lead to a considerable reduction in dimension. A covariance matrix for n securities without any restriction has $\frac{1}{2} \cdot n \cdot (n + 1)$ different parameters. In equation (3) there are $n \cdot k$ parameters for matrix L, $\frac{1}{2} \cdot (k^2 + k)$ parameters for matrix Φ and n parameters for matrix Ω. Altogether this gives us $(n + \frac{1}{2} \cdot k) \cdot (k + 1)$ parameters. Given a ratio of factors to securities common in practice, this signifies a considerable reduction. If, for instance, we wanted to estimate the covariance matrix of 500 asset returns without any restriction, we would have to estimate 125,250 parameters. If, however, we make the assumptions of a multiple factor model with five factors, we only have to estimate 3,015 parameters which is only about 2.4% of the original number.

3. Methodologies to estimate multiple factor models

There are three different methodologies to estimate factor models:

1. time series analysis,
2. cross-section analysis and
3. statistical factor analysis.

Time series analysis is possibly the most intuitive approach for estimating a factor model. Following this methodology, the matrix of the loadings is estimated given the known values of the factors. The advantage of this approach is the control of the factors which can thus be interpreted easily. Typical factors that are considered relevant in many studies, as for instance in the studies of BERRY/BURMEISTER/MCELROY (1988), are the excess return of long term bonds, exchange rates, price changes of raw materials and inflation.

Usually a linear regression is performed with the additional assumption that the parameters α_i, $\beta_{1i}, \ldots, \beta_{ki}$ for a security i are constant over time. In the original formulation of the model, this constancy was not required. It is reasonable to assume, however, that the sensitivity of a stock to a certain factor changes over time, especially after mergers or larger restructurings. Even if no constancy of the parameters is assumed, time series based methodologies always need some time to adequately adapt to these kind of abrupt changes.

Cross-section analysis is certainly less intuitive than the time series approach. Here the exposures are taken as given. Then a regression is performed over *all* securities in *one* period and not over *one* asset over *all* periods. The matrix of the loadings serves as regressor matrix and the estimated parameter vector is interpreted as the vector of the factor values. This regression is performed for several periods in order to obtain time series for the factor values. Starting from these time series we can then estimate Φ, the covariance matrix of the factors. The obvious problem of the cross-section analysis is the sensitivities assumed to be known. The multiple factor models developed by BARRA (cf. ROSENBERG 1974 and ROSENBERG/MARATHE 1976) use fundamental descriptors, because sensitivities must correspond to the economic profile of a firm. The regression model in period *t* thus takes the following form:

$$\begin{pmatrix} r_{t1} \\ \vdots \\ r_{tn} \end{pmatrix} = \begin{bmatrix} \delta_{t11} & \cdots & \delta_{t1k} \\ \vdots & \cdots & \vdots \\ \delta_{tn1} & \cdots & \delta_{tnk} \end{bmatrix} \cdot \begin{pmatrix} f_{t1} \\ \vdots \\ f_{tk} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t1} \\ \vdots \\ \varepsilon_{tn} \end{pmatrix}. \tag{6}$$

δ_{tij} is the value of descriptor *j* for security *i* in period *t*. The estimate of the parameter vector is interpreted as the vector of factor values in period *t*. If the main purpose of the factor model is the simplified estimation of the covariance matrix Σ, an exact interpretation of the statistically obtained factors is not necessary. However, the selection of the descriptors often suggests a corresponding interpretation of the factors which is why this selection should already be based on logical and fundamental reflection.

The third common methodology for the estimation of factor models is the **statistical factor analysis**. Following this approach both the loadings and the factors are estimated simultaneously. Consequently both are statistical constructs that can on the one hand, “optimally” explain the past, but on the other hand can hardly be interpreted economically[1]. Factor analysis is a multivariate statistical methodology attempting to explain the covariance structure of observable random variables by the smallest possible number of linear combinations of these random variables (which are interpreted as factors). All variables of the model can be estimated based solely on the historical returns using the ML-methodology; however, only on the assumption of a multivariate normal distribution of asset returns. In addition, the assumption of constancy of the matrix *L* is required. Further information about the application of factor analysis is provided in RAO (1996). The advantage of factor analysis is the absolute “objectivity” of the approach. Neither the sensitivities nor the factors are defined in advance, but rather are estimated based on the data. However, factor analysis requires the constancy of parameters.

An empirical comparison between a fundamental multiple factor model estimated using the cross-sectional approach and a model estimated using factor analysis was performed by BECKERS/CUMMINS/WOODS (1993). As expected, in-the-sample, the statistical factor model had the highest explanatory power. The fundamental factor model can be seen as a general factor model with specific restrictions on the interpretability of the factors leading to sub-optimality in the ex-post explanation of the data. The out-of-the-sample properties, however, stand in favour of the fundamental risk model, at least in this study. This implies that a pure statistical factor model may lead to an over-fitting of the parameters to the data. This can be reduced by making (reasonable) a-priori assumptions on the respective interactions.

4 Empirical Study

4.1 Data base and experimental design

The objective of this empirical study is the estimation of a fundamental multiple factor model for a universe of European stocks using cross-section analysis. This methodology was preferred to the statistical factor analysis because the results are more easily interpretable and because the constancy of the sensitivities is not required. In addition, the study examines the explanatory power of the model over time.

The database consists of a selection of 656 European stocks from 12 EU member countries as well as Norway and Switzerland. These stocks, due to their market capitalisation and liquidity, form a standard working universe for European investment funds. For this universe the asset returns were calculated on a weekly basis and adjusted for capital measures. If the price of stock *i* in period *t* is denoted by p_{ti} the return r_{ti} for this period is calculated as follows:

$$r_{ti} = \frac{p_{ti} + \text{dividend} + \text{adjustment for capital measures} - p_{t-1,i}}{p_{t-1,i}}. \tag{7}$$

Time series for the weekly returns were calculated in local currency[2] from 01.01.1988 to 30.06.1998. Friday is the base day for the return calculation.

In the next step the descriptors must be determined for each stock and each period. The descriptors examined herein are variables that play a central role in fundamental equity research. The selection of the descriptors is based on the list of descriptors that has already proven its worth in the *BARRA* model. The following descriptors have been selected:

1. *SIZE*: The natural logarithm of the market capitalisation (market cap = number of shares multiplied by the current stock price).
2. *SUCCESS*: The natural logarithm of last year's return. The return is calculated as current stock price divided by the price one year ago. The current price is adjusted for capital measures.
3. BTOP (*book to price*): The equity value stated in the last balance sheet of the company divided by the current market capitalisation. This indicator is often used in equity research to identify "cheap" stocks.
4. ETOP (*earnings to price*): Predicted earnings of the current financial year predicted by analysts divided by the current market capitalisation. This descriptor is very similar to BTOP. As ETOP uses predicted future earnings, this indicator is more up to date than BTOP; on the other hand, however, it is less objective and precise.
5. VIM (*variability in markets*): The historical specific variance $\sigma^2_{\varepsilon i}$, i.e. a measure for the variability in the past.
6. YIELD (*dividend yield*): Last dividend divided by current market capitalisation.
7. PEG (*price to earnings to growth)*: The price earnings ratio (the reciprocal of ETOP) divided by earnings growth of the last four years. ETOP is thus relativised. PEG can give some insight into the development stage of the company.
8. PBROE (*price to book to return on equity*): The quotient of market cap to book value (the reciprocal of BTOP) is divided by the return on equity over the last four years. In analogy to PEG for BTOP.
9. RSI6M (*relative strength index, 6 month*): Relative strength index by LEVY. The logarithm of the quotient of the average stock price over the last week to the average price over the last six months.
10. REV1M: Revision of earnings on a one-month basis. The current earnings predictions divided by the earnings that were estimated by analysts one month ago. This indicator measures the changes in analysts' predictions.
11. REV3M: Revision of earnings on a three-month basis. The current earnings predictions divided by the earnings that were estimated by analysts three months ago.
12. ROE (*return on equity*): Ratio of the earnings of the current financial year to the equity capital of the firm.
13. CROE (*cash flow return on equity*): Ratio of the cash flow of the current financial year to the equity capital of the firm.
14. F1CV: The coefficient of variation of analysts' earnings estimates for the current financial year. This indicator measures the level of agreement among analysts about the near future of the firm.
15. F2CV: The coefficient of variation of analysts' earnings estimates for the following financial year.

First the database[3] must be "cleaned", i.e. extreme values must be identified and modified. This is done using the so-called *skipped Huber method*. For each observation of a variable the absolute deviation from the median is calculated. The median of the variables is calculated over all periods and all assets at a time. Afterwards the median of the deviations is calculated. The median of the observations plus/minus 5.2 times the median of the deviations serve as limits. Extreme values are then referred to the respective limits. This methodology does not differentiate between observations that take on such extreme values because of measurement errors and those that are real extreme values. This is considered acceptable and even necessary, because it can be assumed that measurement errors represent the majority of the cases and because extreme values can distort the results of the linear regression. The advantage of this methodology in comparison to the more common *winsorization* (cutting off all values beyond three times the standard deviation) is that the extreme values themselves do not influence the values that are used for their identification.

The adjustment for extreme values variables are then standardised. This is done according to the usual method, i.e. the mean value of all observations (over all periods and all assets) is subtracted from each observation. The resulting value is divided by the standard deviation of the observations (over all periods and all assets). If x_{it} is the value of observation i in period t the standardised values y_{it} are calculated as follows:

$$y_{\mathrm{it}} = \frac{x_{\mathrm{it}} - \bar{x}}{s_{\mathrm{x}}} . \tag{8}$$

$\bar{x}$ is the mean value and s_{x} the standard deviation of all observations. These figures are determined based on the data already adjusted for extreme values.

In the next step the filtered and standardised descriptors are aggregated to so-called risk indices like in the *BARRA* model. The attribution of the descriptors to the different risk indices and the selection of the risk indices that are to be used in the model were based on fundamental criteria.

In addition to the risk indices constructed for the descriptors we added, as in the *BARRA* model for European stocks, another risk index which is in fact nothing more than a blue chip dummy variable. As blue chips we consider the largest companies that together represent 10% of the total market capitalisation. If a stock is a blue chip, the variable takes on the value 1, otherwise it is 0.

In addition to the risk indices we include country and industry factors. For each country or industry a dummy variable is defined which takes the value 1 if the stock can be attributed to the respective country or industry.

4.2. Estimation of the multiple factor model

For every week a regression of the asset returns on the risk indices as well as on the country and industry dummy variables is performed. The regression model for period t thus takes the form:

$$\underbrace{\begin{pmatrix} r_{t1} \\ \vdots \\ r_{tn} \end{pmatrix}}_{r_t} = \underbrace{\begin{bmatrix} \delta_{t11} & \cdots & \delta_{t1u} & l_{t11} & \cdots & l_{t1v} & b_{t11} & \cdots & b_{t1w} \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \delta_{tn1} & \cdots & \delta_{tnk} & l_{tn1} & \cdots & l_{tnv} & b_{tn1} & \cdots & b_{tnw} \end{bmatrix}}_{L_t} \cdot \underbrace{\begin{pmatrix} f_{t1} \\ \vdots \\ f_{tk} \end{pmatrix}}_{f_t} + \underbrace{\begin{pmatrix} \varepsilon_{t1} \\ \vdots \\ \varepsilon_{tn} \end{pmatrix}}_{\varepsilon_t}. \tag{9}$$

r_t is the vector of the actual asset returns. The matrix L_t represents the values of all the risk indices as well as of the country and industry dummy variables, whereas $\delta_{ti1}, \ldots, \delta_{tiu}$ stand for the risk indices, $l_{ti1}, \ldots, l_{tiv}$ for the country dummy variables and $b_{ti1}, \ldots, b_{tiv}$ for the industry dummy variables for stock i. Consequenctly we can state that $u + v + w = k$. The dummy variables look as follows:

$$l_{tij} = \begin{cases} 1 \Leftrightarrow \text{Stock } i \text{ is from country } j \\ 0 \Leftrightarrow \text{otherwise} \end{cases} \quad \text{and } b_{tij} = \begin{cases} 1 \Leftrightarrow \text{Stock } i \text{ belongs to industry } j \\ 0 \Leftrightarrow \text{otherwise} \end{cases}.$$

The parameter vector f_t is interpreted as the vector of the factor values, the vector of the error terms ε_t represents the specific returns. However, the variances of the error terms ε_{it} cannot be considered to be constant over all assets. This heteroscedasticity can be accounted for by using a GLS- or weighted linear regressison model.

There are several empirical indications that the volatility (i.e. the variance) of a stock decreases with the size of the company increasing. This is why we assume, following the methodology of *BARRA* (cf. CONNOR/HERBERT 1998), that the square root of the market capitalisation is invertedly proportional to the variance of the error term. If this assumption is true, we can perform a weighted linear regression with the roots of market capitalisation as weights. The estimator for the parameter vector which is interpreted as the vector of the factor values then becomes (cf. for instance MONTGOMERY/PECK 1982, chapter 9.2):

$$\hat{f}_t = (L_t^T \cdot W_t \cdot L_t)^{-1} \cdot L_t^T \cdot W_t \cdot r_t. \tag{10}$$

W_t is the diagonal matrix of the weights. The elements on the diagonal of W_t are the roots of market capitalisation of the stocks, all the other elements are 0. Because of the incorporation of the country and industry dummy variables we face the problem of multi-collinearity, i.e. the matrix of the regressors L_t is not of full rank. A restriction on the industry factors is introduced to circumvent this problem:

$$\sum_{i=u+v+1}^{k} f_{ti} = 0. \tag{11}$$

Due to this restriction, general movements of the market are solely accounted for by the country factors.

The result of the weekly regressions is an estimation of the factor values. Having obtained the time series for the factor values, we can estimate the covariance matrix of the factors Φ_t using an exponential smoothing factor in order to give less weight to the observations of past periods. Each element φ_{tij} of the matrix Φ_t is thus estimated as follows:

$$\hat{\varphi}_{tij} = \frac{1}{t-1} \cdot \sum_{u=0}^{t-1} \frac{\lambda^u}{\Lambda} \cdot (\hat{f}_{t-u,i} - \bar{\hat{f}}_{\cdot i}) \cdot (f_{t-u,j} - \bar{\hat{f}}_{\cdot j}). \tag{12}$$

The parameter Λ is used to standardise the weights in such a way that their sum is t. Λ is calculated as follows:

$$\Lambda = \tfrac{1}{t-1} \cdot \sum_{i=0}^{t-1} \lambda^i . \tag{13}$$

For $\lambda = 1$ we get $\Lambda = 1$, so that equation (13) is simply the common unbiased estimation of φ_{tij}. The choice of the value of the exponential smoothing factor λ determines the weight that is attributed to past observations.

The explanatory power of the weekly regressions will be verified by a quality measure. We use the quality measure R^2_A calculated as follows:

$$R^2_A = 1 - \frac{n}{n-k} \cdot \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}. \tag{14}$$

According to KVÅLSETH (1985) this definition of a quality measure is the most appropriate among the many different definitions suggested in the literature.

In order to determine the covariance matrix $\Sigma_t = L_t \cdot \Phi_t \cdot L_t^T + \Omega_t$ a quantification of the diagonal matrix Ω describing the specific risk is still required. To estimate Ω this study follows the methodology suggested by CONNOR/HERBERT (1998). The matrix L_t consists of the risk indices as well as of the country and industry dummy variables in period t and is consequently known. An estimate for the matrix Ω_t is gained from the model for the specific risk. The covariance matrix of the factors Φ_t is estimated based on the results of the weekly regressions. Consequently, the covariance matrix of the assets in period *t* can be estimated as follows:

$$\hat{\Sigma}_t = L_t \cdot \hat{\Phi}_t \cdot L_t^T + \hat{\Omega}_t . \qquad (15)$$

4.3. Results

For the weekly regressions the 15 descriptors were aggregated to six risk indices. The attribution of descriptors to the respective risk indices was based solely on fundamental criteria. The risk indices were calculated as the unweighted arithmetic mean of the (standardised) descriptors contained therein. The attribution of the descriptors is shown in table 1.

Table 1

Risk Index	Descriptor(s)
SIZE	SIZE
SUCCESS	SUCCESS, RSI6M
VALUE	BTOP, ETOP, PEG, PBROE
VIM	VIM, REV1M, REV2M, F1CV, F2CV
YIELD	YIELD
PROFIT	ROE, CROE

The index BLUECHIP was added to these six risk indices so that there were seven risk indices in the regression.

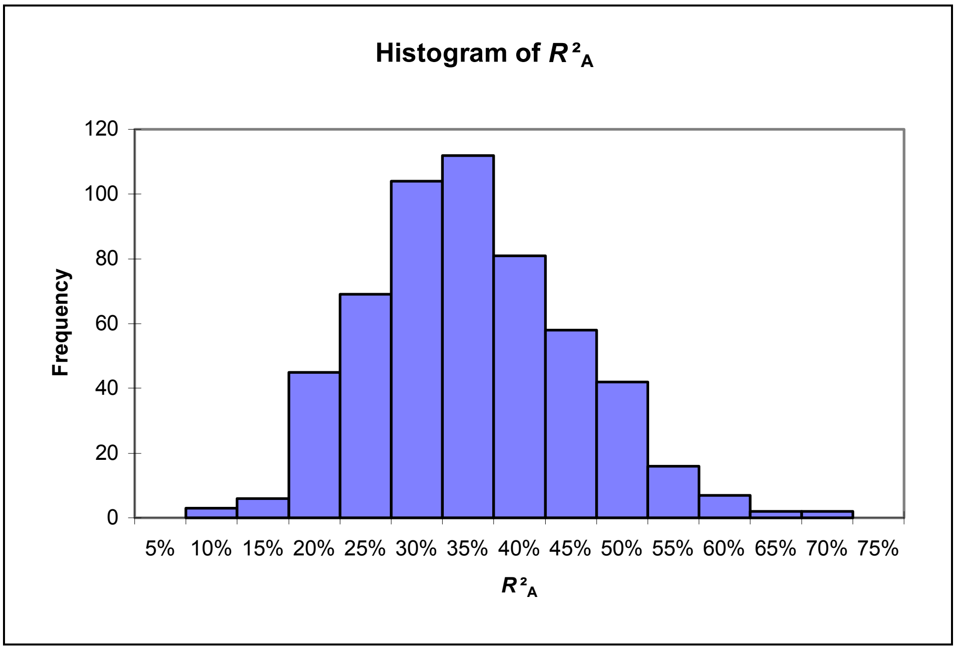
The stocks were attributed to countries according to the location of the headquarters of the respective companies. The firms were located in the following countries: Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Netherlands, Norway, Spain, Sweden, Switzerland and the United Kingdom. The attribution to the different industries was again based on fundamental criteria following the approach of MSCI. However, as this categorisation seems to be too detailed for the purpose of a factor analysis (there are more than 30 industries), the MSCI-industries were aggregated to ten sectors. These sectors are: finance, healthcare, raw materials, commercial services, consumer goods and retail, machine construction / electrical goods, media / leisure / software, multi-industry, oil / energy equipment and utilities / telecom.

Empirical studies of *BARRA* have shown (cf. CONNOR/HERBERT 1998) that continental European industries can be considered sufficiently similar to each other to be represented by only one single factor or one single dummy variable respectively. The British industries, however, are too different from the rest and must thus be represented by individual dummy variables. In this context Ireland has been classified as continental European. Each stock either belongs to one of the ten continental European industries or to one of the ten British industries. The restriction in order to avoid multi-collinearity has been applied both to the continental industries and to the British ones.

Consequently there are 41 regressors in the weekly regressions. These include seven risk indices, 14 country dummy variables, ten continental industry dummy variables and ten British industry variables. For every week between 1 January 1988 and 30 June 1998 a regression of the 41 regressors on the returns of the 656 stocks was performed. The actual number of asset returns in each regression, however, is slightly lower. This is due to the fact that not all of the stocks was already listed on the stock market in 1988 and that in some periods data for certain stocks were missing. The number of observations varies between 473 at the beginning of the examination period and 653 at the end. The mean of the numbers of observations is 581.

The first regression was done for 7 January 1988, the last one for 25 June 1998. 547 regressions were calculated altogether. The quality measure R^2_A was calculated according to equation (12) to determine the fit of the regressions. The values for R^2_A ranged from 7.3% to 66.3% in this study. The arithmetic mean was 32.9%, the median value 32.1%. The respective values in a similar study of *BARRA* (cf. CONNOR/HERBERT 1998) ranged from 9.6% to 56.3% with an arithmetic mean of 30.2%. Figure 1 shows a histogram of the values of R^2_A.

Figure 1



The horizontal line indicates the upper limit of the interval

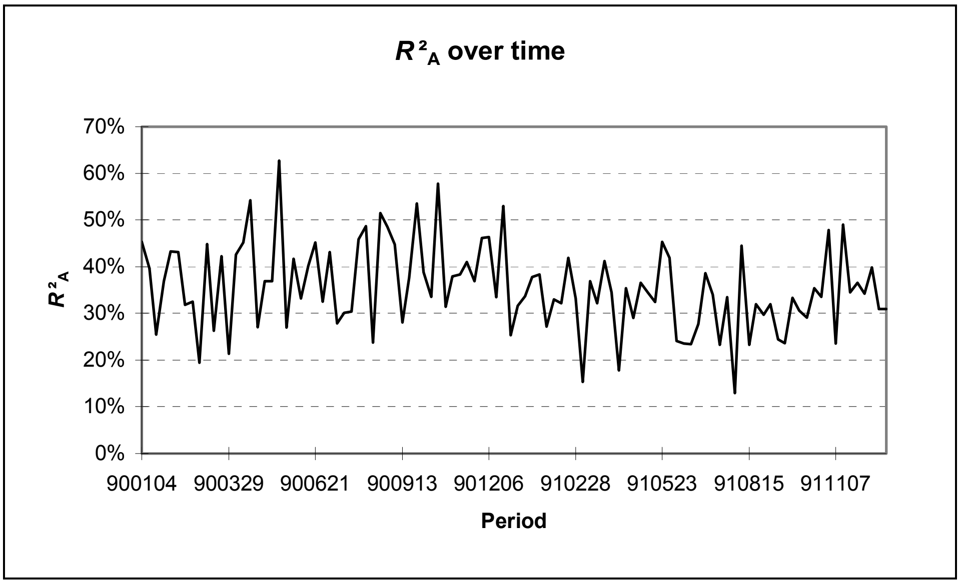
Table 2 shows selected quantiles of values of R^2_A.

Table 2

Quantile	R^2_A
Minimum	7,3%
10%	20,1%
25%	25,5%
Median	32,0%
75%	39,3%
90%	45,8%
Maximum	66,3%

The values of R^2_A over time are displayed in figure 2 and figure 3. Figure 2 shows the weekly values for the randomly selected period from 4 January 1990 to 26 December 1991.

Figure 2



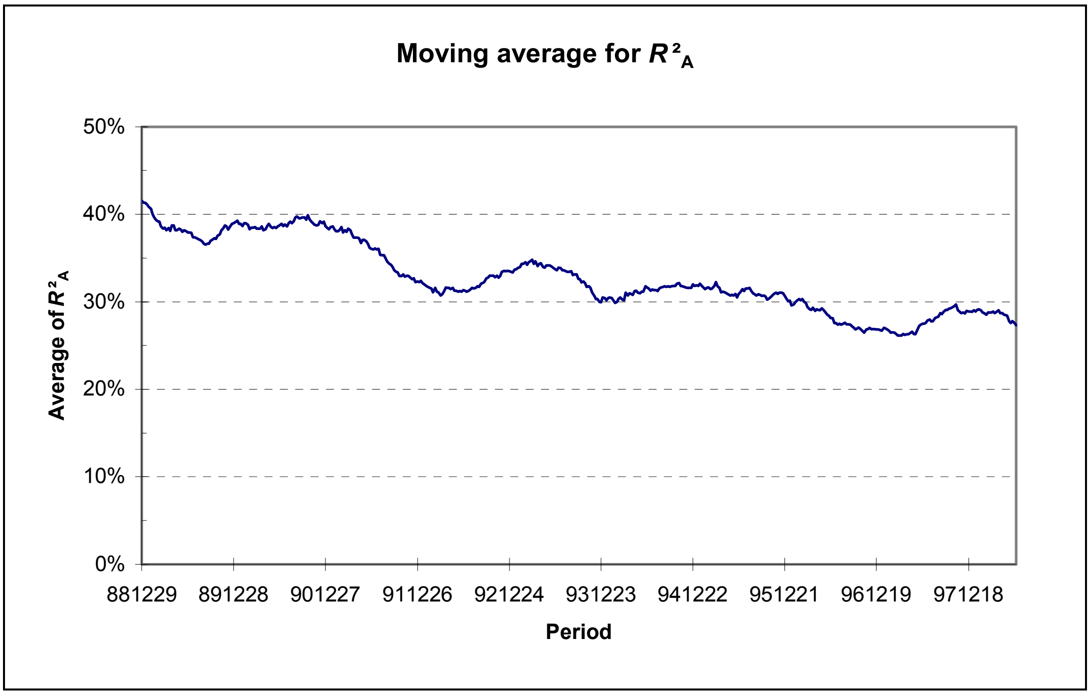
Date is indicated in the format YYMMDD

Looking at the entire examination period in this format would have been too confusing. In order to give an overview of the whole period we determined equally weighted 1-year moving averages for R^2_A. These were calculated following the according to formula:

$$\tilde{R}^2_{A,t} = \tfrac{1}{52}\sum_{i=0}^{51} R^2_{A,t-i} \, . \tag{16}$$

Figure 3 shows the values of $\tilde{R}^2_A$ over the entire available period:

Figure 3



Date is indicated in the format YYMMDD

The analysis shows that approximately 32% of total variance of the asset returns can be explained by the factor model. In comparison with other studies this value can be considered satisfactory. Moreover, figure 3 shows that the values of R^2_A have slightly decreased over time.

The following table shows how often a factor proved to be significant in the weekly regressions. This was examined using a simple t-statistic. However, it has to be kept in mind that the estimates of the parameter vector are interpreted as estimates of the factor values. If the zero-hypothesis of the t-test holds true, this only means that the value of this factor was equal to or near zero in this particular week. It does not signify that the factor is generally irrelevant. If, however, a factor is only rarely significant, it should be clarified whether it is still reasonable to keep this factor in the model. Table 3 shows how often (as percentage of all 547 regressions) a factor was significant at a 5% level.

Table 3

No.	Factor	Frequency	No.	Factor	Frequency
1	United Kingdom	70,6%	22	UK – Utilities / Telecom	25,4%
2	Italy	68,2%	23	Oil / Energy Equipment	22,1%
3	France	67,1%	24	Norway	21,6%
4	Germany	62,0%	25	Ireland	21,4%
5	Spain	61,6%	26	UK – Machine Construction / Electrical Goods	21,4%
6	Netherlands	52,5%	27	YIELD	18,6%
7	Switzerland	44,6%	28	UK – Commercial Services	18,1%
8	Belgium	43,3%	29	Utilities / Telecom	15,9%
9	Sweden	41,0%	30	VALUE	14,8%
10	SUCCESS	36,6%	31	Machine Constr. / Electr. Goods	13,9%
11	Austria	36,4%	32	UK-Media / Leisure / Software	13,7%
12	UK – Healthcare	32,7%	33	UK - Multi Industry	13,7%
13	Denmark	32,5%	34	Consumer Goods and Retail	13,2%
14	Finland	32,2%	35	Raw Materials	12,4%
15	Finance	29,3%	36	PROFIT	12,1%
16	UK – Raw Materials	28,5%	37	BLUECHIP	11,0%
17	UK – Finance	27,4%	38	Healthcare	10,8%
18	UK – Consumer Goods and Retail	27,4%	39	Commercial Services	10,4%
19	VIM	26,0%	40	Media / Leisure / Software	7,3%
20	UK – Oil / Energy Equipment	25,8%	41	Multi Industry	6,0%
21	SIZE	25,4%			

The factors in table 3 are arranged according to the frequency of their significance in the regressions. The country factors were significant most frequently. This is not surprising if we keep in mind that due to the construction of the matrix of the regressors, the intercept (i.e. general market movements) was integrated in the country dummy variables. Especially the factors of the large and important countries proved to be significant in more than 60% of all cases. As far as the risk indices are concerned, no clear picture emerges. The indices SUCCESS, VIM and SIZE are significant in a minimum of 25% of the regressions. The indices PROFIT and BLUECHIP, however, are at the lower end of the table. The situation for the industry dummy variables is no less diverse. It is striking, however, that the British industry factors are significant much more often than their continental counterparts which prove to be significant least frequently of all factors. In the end table 3 conforms to the satisfactory overall impression of the results of the regressions.

5 Conclusion

Multiple factor models are a powerful tool for the statistical formulation of return generating processes and can thus make detailed risk analysis and prognosis easier. There are three different methodologies for the statistical identification of the model: time series analysis with previously defined factor values, cross-section analysis with previously defined factor sensitivities and statistical factor analysis with simultaneous identification of loadings and factors. For reasons of more accurate interpretation we chose the cross-sectional approach was chosen for this study. The results of the estimations can be considered very satisfactory, as the multiple factor model explains more than 30% of the variance on average. Possible extensions of this study are the enlargement of the stock universe or the integration of bond portfolio analysis.

References

ALBRECHT, PETER; MAURER, RAIMOND; MAYSER, JÜRGEN (1996): ”Faktormodelle: Grundlagen und Einsatz im Management von Aktien-Portefeuilles”, *Zeitschrift für betriebswirtschaftliche Forschung* 48, No. 1: 3-29.

BECKERS, STAN; CUMMINS, PAUL; WOODS, CHRIS (1993): “The Estimation of Multiple Factor Models and their Applications: The Swiss Equity Market*”, Finanzmarkt und Portfolio Management* 7, No. 1: 24-45.

BERRY, MICHAEL A.; BURMEISTER, EDWIN; MCELROY, MAJORIE B. (1988): “Sorting out Risks Using Known APT Factors”, *Financial Analysts Journal* 44: 29-42.

CONNOR, GREGORY; HERBERT, NICK (1998): Regional Modelling of the Western European Equity Market, BARRA-Working Paper, London.

ELTON, EDWIN J.; GRUBER, MARTIN J. (1988): ”A Multi-Index Risk Model of the Japanese Stock Market”, *Japan and the World Economy* 1: 21-44.

FAMA, EUGENE; FRENCH, KENNETH R. (1992): ”The cross-sections of Expected Stock Returns”, Journal of Finance 47: 427-465.

GRINOLD, RICHARD C; KAHN, RONALD N. (1995): “Active Portfolio Management”, Chicago, IL, Probus

KVÅLSETH, TARALD O. (1985): “Cautionary Note About R^2”, *The American Statistican* 39, No. 4: 279-285.

LINTNER, JOHN (1965): ”The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets”, *Review of Economics and Statistics* 47, no. 1: 13-37.

MONTGOMERY, DOUGLAS C.; PECK, ELIZABETH A. (1982): Introduction to Linear Regression Analysis, New York, NY: Wiley.

MOSSIN, JAN (1966): ”Equilibrium in a Capital Asset Market”, *Econometrica* 34, no. 4: 768-783.

RAO, C. RADHAKRISHNA (1996): “Principal Component and Factor Analysis” in: Maddala, G. S. (Hrsg.); Rao, C. R. (Hrsg.): *Handbook of Statistics*, Vol. 14: 489-505, Amsterdam: North Holland.

ROSENBERG, BARR (1974): ”Extra-Market Components of Covariance in Security Returns”, *Journal of Financial and Quantitative Analysis* 9: 263-274.

ROSENBERG, BARR; MARATHE Vinay.(1976): ”Common factors in Security Returns: Micro economic determinants und macro correlates”, Proceedings of the seminar on the analysis of security prices, University of Chicago: 61-116.

ROSS, STEPHEN A. (1976): ”The Arbitrage Theory of Capital Asset Pricing”, *Journal of Economic Theory* 13, no. 3: 341-360.

Endnotes

---

[1] In some cases, however, it might be possible to gain some insights by analysing the correlation between the statistically determined factors and other indicators that are known and can therefore be interpreted, especially in the case of simple phenomena like a market factor or individual industry factors. ELTON/ GRUBER (1988) applied this approach in an empirical study.

[2] If the returns are calculated in local currency, currency covariances can be modelled separately for the tracking error prognosis. This, however, is not subject of this article.

[3] The accounting data were provided by Worldscope-Disclosure, the consensus estimates by IBES. For the reported data from Worldscope we had to consider a time lag for the publication of the accounts. We assumed a relatively conservative lag of six months from the end of the financial year until utilisation of the data in our factor model.