Semiparametric Modelling of the Cross-Section of Expected Returns in the German Stock Market*

Richard Stehle
Institut für Bank-, Börsen- und Versicherungswesen
Humboldt-Universität zu Berlin

Olaf Bunke & Volker Sommerfeld
Institut für Mathematik
Humboldt-Universität zu Berlin

December 1, 1997

Abstract

According to the Sharpe-Lintner capital asset pricing model, expected rates of return on individual stocks differ only because of their different levels of non-diversifiable risk (beta). However, Fama/French (1992) show that the two variables size and book-to-market ratio capture the cross-sectional variation of US stock returns better than other combinations of two variables. They report also that in the 1963-1990 period beta has virtually no explanatory power. This paper looks at a comparable data set for Germany for the time period 1968-1990. We analyze this data set in order to identify a "best" nonlinear model for the relationship between rates of return, beta, size and book-to-market. The model and corresponding regression estimates are chosen by "cross-validation" among a very rich class of parametric, semiparametric and nonparametric alternatives. The coefficients in the model are estimated each year.

The major result is that the parametric model proposed by Fama/French for US stock returns is almost the best one in Germany. The book-to-market-ratio turns out to be the variable with highest partial correlation with the stock return. In most of the annual regressions the corresponding coefficients have the correct sign and are statistically significant.

AMS subject classification: 62G07, 62J05, 62P05.

Keywords and phrases: Model selection, cross-validation, capital asset pricing model, German stock market, time series of cross-sectional data.

1 Introduction

According to the Sharpe-Lintner capital asset pricing model, expected rates of return on individual stocks differ only because of their different levels of non-diversifiable risk (beta). Banz (1981) presents empirical evidence that a negative relationship exists between firm size, measured by the market value of outstanding stocks, and the average rate of return on those stocks, even after adjusting for their different

[*] The research was carried out within the Sonderforschungsbereich 373 at Humboldt University Berlin The paper was printed using funds made available by the Deutsche Forschungsgemeinschaft

risk characteristics (size or small firm effect), which contradicts the Sharpe-Lintner model in its narrow sense. Fama/French (1992) conclude that the two variables size and book-to-market ratio capture the cross-sectional variation of US stock returns better than other combinations of two variables. They report also that in the 1963-1990 period non-diversifiable risk (beta) has virtually no explanatory power. In two subsequent studies Fama/French (1993, 1995) provide supporting evidence for their 1992 conclusions.

Actually, Fama/French use the logarithm of size and the logarithm of the book-to-market ratio as independent variables in their cross-sectional regression models, thus assuming a nonlinear relationship between the expected rate of return and the untransformed values of the independent variables. They use this specific form of a nonlinear relationship without giving a specific reason and without discussing alternative nonlinear functions.

This paper looks at a comparable data set for Germany for the time period 1968-1990. Instead of assuming a specific nonlinear relationship between rates of return, size and book-to-market, a wide range of possible nonlinear relationships is analyzed in a systematic fashion in order to identify a "best" nonlinear model. In our analysis we have to ensure a good fit to the data but also to take into consideration errors in estimating parameters or functions in the models. These errors cannot be neglected because of the finite number of observed stocks. We also analyze annual cross-sectional regressions in the traditional way and in the Fama/MacBeth (1973) framework. The results are compared with a regression, in which time independent coefficients are assumed, and with non- and semiparametric regression estimates.

The major result is that the model proposed by Fama/French is extremely close to the best nonlinear model involving beta and the two independent variables. Additionally, it is shown in this paper that this nonlinear parametric model performs better than models based on modern semi- and nonparametric estimators. In most of the annual regressions, the coefficient of the book-to-market ratio has the expected sign and is statistically significant. The coefficients of beta and size also have the correct sign, but they are statistically not significient. This is also an interesting new insight, since prior studies on the German stock market only focussed on size and found a strong size-effect in risk-adjusted returns.

The Fama/French results may be interpreted as support for a more general capital asset pricing model, in which size and book-to-market ratio are proxies for non-diversifiable risk and / or unknown risk variables. Unfortunately, the exact functional form of this more general capital asset pricing model is not known. Thus, the equation

$$E\,(R_{ti}) = f_t(\text{beta}_{ti},\ \text{size}_{ti},\ \text{book-to-market}_{ti}) \qquad (1.1)$$

is the starting point of our analysis. Since theoretical arguments, whether such a relationship should hold on a daily, monthly, quarterly, or annual basis, do not exist, we use annual rate-of-return observations as a point of departure. Beta, size, and book-to-market ratios are observed for each firm i at the beginning of each calendar year t, the rates of return R_{ti} are calculated for the following year $t+1$. As the structure of the relationship described in equation 1.1 may change over time, it is natural to estimate the regressions separately for each year.

Since we cannot observe expected rates of return, we use actual rates of return as dependent variable. The residuals may be interpreted as the deviations of the actual returns from their expected values. In an efficient capital market these deviations must have an expected value of 0. As a starting point we assume that the corresponding errors are uncorrelated and homoscedastic with respect to firms. That is, we start our analysis with the assumption

$$E(\epsilon_{ti}) = 0, \quad Var(\epsilon_{ti}) = \sigma_t^2 \tag{1.2}$$

where

$$\epsilon_{ti} := R_{ti} - f_t(\text{beta}_{ti}, \text{size}_{ti}, \text{book-to-market}_{ti})$$

In section 2, the methodology used is discussed in detail. In section 3, the empirical results are presented. An analysis which considers possible time dependent variances and correlations is beyond the scope of this paper. It requires the development of additional statistical procedures and is an objective of our current research.

2 Model Choice Methodology

One of the aims in analyzing time series of cross sectional data is to find a description of the dependence of a variable Y on other independent (explanatory) variables $X_1, \dots, X_k$ which are suspected to have an influence on Y. This dependence may be described by a regression function $f(X_1, \dots, X_k)$ which is interpreted as the expected value of Y for fixed values $X_1, \dots, X_k$ (see equation (1.1)). As the structure of this dependence could change (gradually or even suddenly) over time, it is natural to estimate the regression separately at each time period t, using all observations at t to estimate the parameters in a convenient model. We propose to estimate the regression function by the ATFR (Adaptive Time Dependent Fitting of Regression Models) procedure of Bunke (1997), which systematically tries numerous models of different forms and alternative transformations of the variables including observations of neighboring time periods within a given time horizon.

We consider variables $X_1, \dots, X_k, Y$ with the aim of identifying a dependence of Y on some of the variables $X_1, \dots, X_k$ which have influence on Y. For each time period $t = 1, \dots, T$ we have observations

$$X_{ti} = (X_{1ti}, \dots, X_{kti}),\ Y_{ti} \tag{2.3}$$

of these variables for firms indexed by $i = 1, \dots, n_t$. We assume that for fixed X_{ti} the observations Y_{ti} are realizations of uncorrelated random variables which follow a regression model

$$Y_{ti} = f_t(X_{ti}) + \epsilon_{ti}, \tag{2.4}$$

$$E(\epsilon_{ti}) = 0, \quad Var(\epsilon_{ti}) = \sigma_t^2. \tag{2.5}$$

The regression functions f_t and the variances σ_t^2 are unknown.

ATFR is a procedure for estimating the regression function f_t by an estimate $\hat{f}_t$ based on fitting in each time period t the same semiparametric model M, using observations of time period t and possibly of neighbouring time periods $t \pm 1, t \pm 2, \ldots, t \pm r$, within a "horizon" r. The model M and the horizon r are chosen by cross-validation with the aim of a small average mean squared error in estimating the regression function, that is by minimizing

$$MSE = \frac{1}{n} \sum_{t1}^{T} \sum_{i1}^{n_t} E \left| \hat{f}_t(X_{ti}) - f_t(X_{ti}) \right|^2 , \tag{2.6}$$

$$n = \sum_{t1}^{T} n_t . \tag{2.7}$$

The admitted values for the horizon r are $r = 0$ (only observations at t are used for fitting) and $r = 1, 2, \ldots, r_t$ where $r_t = \min[t-1, T-t]$.

The admitted models are elements of a rich class of parametric, nonparametric and semiparametric models. This class consists in 5 subclasses of models. Each model leads to corresponding estimates of regression functions f_t.

1. Parametric estimators

We consider parametric models M of the form

$$f_M(x|b) = T_0^{-1} \left[p_q(T_1[x_1], \ldots, T_k[x_k]; \tilde{T}|b) \right] \tag{2.8}$$

where $T_0, T_1, \ldots, T_k$ are transformations from a conveniently chosen class $\mathcal{T}_j$ of transformations and

$$\begin{aligned} & p_q(t_1, \ldots, t_k; \tilde{T}|b) && (2.9) \\ = \;& b_0 + \sum_{j_1 1}^{k} b_{j_1} T_{j_1}[t_{j_1}] + \sum_{j_1, j_2 1}^{k} b_{j_1 j_2} T_{j_1, j_2}[t_{j_1} t_{j_2}] + \cdots + \sum_{j_1, \ldots, j_q 1}^{k} b_{j_1 \cdots j_q} T_{j_1 \cdots j_q}[t_{j_1} \cdots t_{j_q}] \end{aligned}$$

is a nonlinear transformation of a polynomial of order q in the variables $t_1, \ldots, t_k$ with $\tilde{T} = (T_0, T_1, \ldots, T_k, T_{11}, T_{12}, \ldots, T_{kk}, \ldots, T_{1 \cdots 1}, \ldots, T_{k \cdots k})$, $T_{i_1 \cdots i_l} \in \mathcal{T}_j$. Each choice $M = (\tilde{T}, p_q)$ of transformations and of the polynomial p_q of order q determines a specific model M. Transforming 2.8 by T_0 corresponds to a model for the transformed dependent variable $T_0[Y]$ which is linear in the parameter vector b.

The heuristical background of such a model is the approximation of the regression of the variable $T_0[Y]$ on the transformed independent variables $T_1[X_1], \ldots, T_k[X_k]$ by a (possibly nonlinearly transformed) polynomial of order q. The choice $T_0 = T_1 = \cdots = T_k = T_{11} = \cdots = T_{k \ldots k} = I$ (identical transformation $I[x] = x$) is admitted in $\mathcal{T}_j$, that is a direct approximation of the original dependent variable by a polynomial in the original influential variables, but the possibility of alternative nonlinear variable transformations may lead to a better approximation and consequently to more accurate estimates of the regression. The classes $\mathcal{T}_j$ consists of the identical transformation I and the three standard transformations

$$T_j[x_j] = \begin{cases} \frac{1}{x_j a_j} \\ \ln[(x_j + d_j)/s_j)] \\ \exp(c_j x_j) \end{cases} \tag{2.10}$$

The constants a_j, c_j, d_j (see Bunke, Droge & Polzehl, 1995) have to be conveniently chosen (possibly $a_j = d_j = 0$, $c_j = 1$), while

$$s_j^2 = \frac{1}{n} \sum_{t,i} \left| X_{jti} - \frac{1}{n} \sum_{t,i} X_{jti} \right|^2 \tag{2.11}$$

denotes the empirical variance of the variable X_j. We remark that leaving out some of the terms in the polynomial 2.9 leads to a lower dimensional parametric model. In our model selection procedure we also admit a convenient hierarchical class of such models (see Bunke, 1997).

In the time period t the model $M = (T, q)$ is fitted by least squares using all observations of the time periods $t-r, t-r+1, \ldots, t, \ldots, t+r$, that is, minimizing

$$S_{M,r,t}(b) = \sum_{\tau t-r}^{tr} \sum_{i1}^{n_\tau} |Y_{\tau i} - f_M(X_{\tau i}|b)|^2. \tag{2.12}$$

with respect to b:

$$S_{M,r,t}(\hat{b}_{M,r,t}) = \min_b S_{M,r,t}(b). \tag{2.13}$$

2. Nonparametric kernel estimators
We also consider kernel estimates $\hat{f}_t^{K,h}$ (see Härdle, 1990) of the regression function defined by a kernel function K and a bandwidth h and based on the observations of the transformed variables $T_1[x_1], \ldots, T_k[x_k]$ and of Y in the time periods $t-r, \ldots, t+r$ within a horizon r.

We consider the following semiparametric extensions of the model 2.8.

3. Semiparametric estimates based on the additional estmation of a smooth link function g

$$f_M(x|b, g) = g\left[p_q(T_1[x_1], \ldots, T_k[x_k]; \tilde{T}|b)\right] \tag{2.14}$$

The estimation is performed iteratively by a kernel method for g and by least squares for the parameter b.

4. Estimates based on additive models

$$\begin{aligned} f_M(x|g) \quad = \quad & g_1(T_1[x_1]) + \cdots + g_k(T_k[x_k]) \\ & + g_{12}(T_1[x_1], T_2[x_2]) + \cdots + g_{k-1,k}(T_{k-1}[x_{k-1}], T_k[x_k]) \end{aligned} \tag{2.15}$$

where $g_1, g_2, \ldots, g_{k-1,k}$ are assumed to be smooth functions. Here we consider all hierarchical models with interactions up to order 2. We remark that leaving out some of the terms in 2.15 leads to models of lower complexity. In our model selection procedure we admit also a convenient hierarchical class of such models. The estimation is done by backfitting (see Hastie & Tibshirani, 1990) and leads to estimates

$$\hat{f}_t = f_M(x|\hat{g}_{M,r,t}). \tag{2.16}$$

5. Partially parametric estimates
We consider the combinations of the models 2.8 and 2.15:

$$f_M(x|b, g) = T_0^{-1}\left[p_q(T_1[x_1], \ldots, T_k[x_k]; \tilde{T}|b)\right] + g_1(T_1[x_1]) + \cdots + g_k(T_k[x_k]). \tag{2.17}$$

The estimation is done iteratively by backfitting and least squares.

If we select a certain model M or estimator in the above defined class and use in the time period t observations within a time horizon r to estimate the regression function f_t we arrive at an estimator $\hat{f}_t$, e.g. in the class 1. of parametric estimators

$$\hat{f}_t = f_M(x|\hat{b}_{M,r,t}), \quad M = (T, q) \tag{2.18}$$

(or e.g. in the class 2. of kernel estimates $\hat{f}_t = \hat{f}_t^{K,h}$ with selected kernel K and bandwidth h). The cross-validation criterion for such an estimator $\hat{f}_t$

$$C = C(M, r) = \frac{1}{n} \sum_{t1}^{T} \sum_{i1}^{n_t} |Y_{ti} - \hat{f}_t^i(X_{ti})|^2 \tag{2.19}$$

is defined by the values of estimates $\hat{f}_t^i$ of the regression function f_t at X_{ti}, where $\hat{f}_t^i$ is the estimator $\hat{f}_t$ leaving out the observation (X_{ti}, Y_{ti}). The criterion (2.19) is an almost unbiased estimate of the MSE (2.6) (up to a term independent of the model and the horizon). The adaptive regression estimate is given by a $(\hat{M}, \hat{r})$ minimizing the cross-validation criterion (2.19).

The quality of the selected model and horizon (or of the bandwidth h), that is of the estimator $\hat{f}_t$, may be also measured by a standardization of cross-validation with values in the interval $[0, 1]$:

$$B = B(Y, X_1, \ldots, X_k) = 1 - \frac{C(\hat{M}, \hat{r})}{S^2}$$

with

$$S^2 = \frac{1}{n} \sum_{t,i} |Y_{ti} - \overline{Y}|^2, \quad \overline{Y} = \frac{1}{n} \sum_{t,i} Y_{ti}.$$

This value has the same interpretation as a coefficient of determination and we call it "cross-validated coefficient of determination". In comparison to the usual coefficient of determination R^2 but also to the "adjusted coefficient of determination" (adjusted by "degrees of freedom") the coefficient B takes directly into consideration the errors committed in estimating the parameters for functions leading to the estimators $\hat{f}_t$. Therefore it gives a more realistic evaluation of the quality of the corresponding simultaneous selection of a model M and a horizon r.

The value B may be compared with the cross-validated coefficient of determination for any competitive model M and horizon r, e.g. a linear model M given by

$$f(x) = \sum_{j1}^{k} b_j x_j$$

and $r = 0$, that is, fitting only with the observations of the corresponding time period t. Such a comparison may be very useful to asses the goodness of some modification $\tilde{M}$ of the optimal model $\hat{M}$ which is especially simple, has only few terms or parameters, or which has a better economic interpretation than $\hat{M}$. This model $\tilde{M}$ will be acceptable if its cross-validated coefficient of determination is only slightly smaller than the optimal one.

It is possible to assess separately the influence of each variable X_j on the dependent variable Y by the cross-validated partial coefficient of determination:

$$\begin{aligned} B_j &= B(Y, X_j | X_1, \ldots, X_{j-1}, X_{j1}\ \ , \ldots, X_k) \\ &= \frac{\hat{B}(Y, X_1, \ldots, X_k) - \hat{B}(Y, X_1, \ldots, X_{j-1}, X_{j1}\ \ , \ldots, X_k)}{1 - \hat{B}(Y, X_1, \ldots, X_{j-1}, X_{j1}\ \ , \ldots, X_k)}. \end{aligned} \tag{2.20}$$

where $\hat{B}(Y, X_1, \ldots, X_k)$ denotes the maximal value of $B(Y, X_1, \ldots, X_k)$, that is, the cross-validated coefficient of determination for the optimal simultaneous choice of a model (that is, of the estimator) and of a horizon.

æ æ

3 Empirical Results

In this study we use annual data for all non-financial German firms listed on the Frankfurt stock exchange in the time period 1968-1990.[1] For each firm included in the sample we calculated the annual rates of return on the firms common stocks, the beta values (equity betas), the market value of the equity, and the market-to-book ratio. The annual rates of return include dividends, stock splits, rights issues etc.. German stockholders not only obtain a cash dividend, but also a tax credit in the amount of the corporate income tax on distributed profits, which is included in the rate of return calculation. Thus rates of return are calculated from the perspective of a German stockholder facing a marginal income tax rate of 0 %. In a first step, monthly rates of return were calculated, which susequently were compounded. Betas are normally based on the 60 monthly rates of return prior to the date to which they are assigned. To obtain the market values of the equity of the included companies, the market value of all outstanding common stocks were added, because in Germany common stocks and preferred stocks are close substitutes. Similarly, the book-to-market ratio was calculated. The number of data sets differs slightly from year to year, the total number of observations is 3329. The independent variables are the beta value β, the market value MV and the book-to-market ratio B/M of a firm at the end of year t. The dependent variable is the asset return R in the year $t+1$.

A good model is one that describes the regression function $E[R|\beta, MV, B/M] = f(\beta, MV, B/M)$ quite well in the sense of a large cross-validated coefficient of determination. On the other hand, this model should not have too many explanatory variables and a simple structure in order to have a sound economic interpretation. Therefore, we have run the ATFR procedure for different situations. We have chosen the polynomial order in the model 2.9 as $q = 3$, but additionally we limited the maximum number of terms in the model formula to 4 and 12, respectively. For these cases we selected (in the ATFR sense) best models in

[1] Banks and insurance companies are excluded because the typically have much higher debtequity ratios than nonfinancial firms which may effect the relationship among the variables we include in this study

the class $\mathcal{M}_1$ of parametric models using transformations *with* and *without* transformation constants a_j, c_j, d_j, s_j. That is, additionally to the case with transformation constants we selected the model using the transformations

$$\begin{aligned} T_1[x] &= x \\ T_2[x] &= \frac{1}{x} \\ T_3[x] &= \exp(x) \\ T_4[x] &= \ln(x) \end{aligned}$$

where the transformation T_4 for the dependent variable Y_{ti} was omitted because of the frequent negativity of asset returns.[2] In this way we obtain four models. The reason for this multiplicity is that we are interested in a simple model that has a sound economic interpretation, which is better achieved by a model without transformation constants. We obtained that the cross-validated coefficients of determination corresponding to these four models differ only very slightly. Additionally, we compared these models for the time horizons $r = 0, 1, 2$. *We get the ”almost best parametric model”*[3]

$$E\,R_{ti} = b_{1,t} + b_{2,t}\beta_{ti} + b_{3,t}\ln(MV_{ti}) + b_{4,t}\ln(B/M_{ti}) \tag{3.1}$$

with the characteristics (averaged over all years)

	model 3.1	model 3.2	best model
cross validation value	0.0849	0.0852	0.0844
coefficient of determination	0.1056	0.0886	0.1129
cross-validated coefficient of determination	0.0482	0.0448	0.0538

and coefficients given in table 3.1. The selected best time horizon is $\hat{r} = 0$ which means that the best fitting procedure consists in estimating the coefficients solely with the data of the corresponding year.

Table 3.2 contains the corresponding standard deviations.[4] It turns out that in most of the years book-to-market is statistically significant at a 5 per cent level. Table 3.1 analyzes the data also by the procedure suggested by Fama/MacBeth (1973). Once performed the OLS $\hat{b}_{1,t}, \ldots, \hat{b}_{4,t}$ of the coefficients of the independent variables beta, market value and book-to-market ratio, in a second step they analyze the time series of these estimated coefficients. Following this procedure, it turns out that only the book-to-market ratio coefficient is statistically significant at a 5 per cent level (t-value = 2.91). A more quantitative evaluation by P-values is given in the appendix.

The mean of the coefficients (written at the end of table 3.1) have the following simple economic interpretation. If we increase the beta value of a firm from 1 to 2

[2] Additionally we fitted 1 R as dependent variable in order to include the transformation T_4 in the model choice procedure Thereby note that for small R it holds that ln1 $R \approx R$

[3] The best parametric model is

$E\,R_{ti}\;\; b_{1,t}\;\; b_{2,t}\beta_{ti}\;\; b_{3,t}\,\ln MV_{ti}\;\; b_{4,t}\,\ln B/M_{ti}\;\; b_{5,t}\,\ln MV_{ti}\ln B/M_{ti}\;\; b_{6,t}\,\ln\{\beta_{ti}\ln MV_{ti}\ln B/M_{ti}\}.$

[4] Significance levels and standard deviations are calculated under the assumption of an adequate normal homoscedastic regression model 31

(which corresponds to a higher correlation of the rate of return of this firm with the rate of return of the market portfolio), the rate of return of that firm (averaged over all 23 years) increases by 1.77%. If we increase the market value from 10 million to 1 billion DM then the rate of return of the firm's stock decreases by

$$0.011(\ln(1,000,000,000) - \ln(10,000,000)) = 0.011 \ln(100) = 0.051,$$

that is, it decreases by 5.1%. If we increase the book-to-market ratio from 1/2 to 1 then the rate of return of a corresponding asset increases by 4%.

The influence of the different variables is described by the corresponding partial coefficients of determination

model (3.1)	β	$\ln(MV)$	$\ln(B/M)$
partial coefficient of determination	0.0191	0.0331	0.0457
cross-validated partial coefficient of determination	0.0102	0.0212	0.0343

Recall that we choose this "almost best parametric model" because the best parametric model has a complex structure and does not allow a sound economic interpretation.

As an alternative method to asses the influence of beta we fitted model (3.1) without the independent variable β:

$$R_{ti} = b_{1,t} + b_{2,t} \ln(MV_{ti}) + b_{3,t} \ln(B/M_{ti}). \tag{3.2}$$

The characteristics of this model are given above in comparison with the characteristics of model (3.1).

The coefficients of model (3.1) vary quite significantly over time. For a stability analysis, we fitted model (3.1) for the data of all 23 years together, that is, we use pooled data of all 23 years. The obtained results are given in the following table (t-values in parentheses).

coefficient of Intercept	0.2818	(6.02)
coefficient of β	-0.0053	(-1.29)
coefficient of $\ln(MV)$	-0.0073	(-0.83)
coefficient of $\ln(B/M)$	0.0975	(9.23)
cross validation value	0.1234	
coefficient of determination	0.0278	
cross-validated coefficient of determination	0.0248	

As the cross-validated coefficient of correlation for these averaged data is small in comparison with the averaged coefficient 0.0482, one could interprete this as that there is a real time dependent structure, that means that the variations in the coefficients are not only caused by random perturbations. The higher t-values for the above model with pooled data in comparison with the t-values appearing in the Fama/MacBeth procedure (see the last line of table 3.1) can be explained as follows. In the latter case we admit different coefficients for each year. Thus we obtain a time series of estimated coefficients which vary strongly over the different years. On the other hand, in the pooled data model we force the coefficients to be constant

year	Intercept	β	$\ln(MV)$	$\ln(B/M)$
1968	0.8114*	-0.0792	-0.0308	0.0884
	(5.0205)	(-0.8401)	(-1.7686)	(1.7177)
1969	0.0577	-0.1511*	-0.0058	0.0538*
	(0.6821)	(-3.3534)	(-0.6794)	(2.1147)
1970	0.1214	0.1076	-0.0133	-0.0726*
	(1.2415)	(1.9574)	(-1.3797)	(-2.3700)
1971	2.1086*	0.2127	-0.1554*	0.1971*
	(6.6676)	(1.3045)	(-5.0036)	(2.2808)
1972	0.0672	-0.0394	-0.0124	-0.0118
	(0.6159)	(-0.8332)	(-1.2558)	(-0.4512)
1973	-0.4825*	-0.0648	0.0437*	0.1204*
	(-4.0819)	(-1.4092)	(4.1097)	(4.4362)
1974	-0.3582*	0.1248*	0.0473*	0.1839*
	(-2.5728)	(2.0800)	(3.8035)	(4.9345)
1975	-0.0223	-0.0595	0.0051	0.1000*
	(-0.2089)	(-1.2699)	(0.5193)	(3.2568)
1976	0.4359*	0.0743	-0.0196	0.1207*
	(3.1423)	(1.1928)	(-1.5250)	(3.0251)
1977	0.7663*	0.0871	-0.0526*	-0.0001
	(6.0182)	(1.5865)	(-4.5399)	(-0.0020)
1978	0.0027	-0.0632	0.0000	0.0332
	(0.0222)	(-1.3580)	(-0.0047)	(0.9834)
1979	-0.1574	0.0025	0.0139	-0.0978*
	(-1.1598)	(0.0535)	(1.1877)	(-2.6434)
1980	-0.2517	-0.0206	0.0231	-0.0328
	(-1.8259)	(-0.4210)	(1.9143)	(-0.9595)
1981	0.1397	-0.0060	0.0057	0.0624
	(0.8610)	(-0.0985)	(0.3928)	(1.6471)
1982	0.3226	0.0995	-0.0023	0.1520*
	(1.6366)	(1.6146)	(-0.1336)	(3.5788)
1983	0.5328*	0.0384	-0.0324*	0.0731
	(3.0610)	(0.7268)	(-2.1599)	(1.6547)
1984	-0.2110	0.2479*	0.0418*	0.1526*
	(-1.0802)	(3.3296)	(2.4920)	(3.5123)
1985	0.0439	0.2239*	-0.0153	-0.1773*
	(0.1619)	(2.1112)	(-0.6519)	(-2.7319)
1986	0.3264*	-0.0957	-0.0316*	0.0369
	(1.8527)	(-1.5597)	(-2.0765)	(0.9460)
1987	0.7774*	0.0386	-0.0235	0.1909*
	(3.8514)	(0.8810)	(-1.4708)	(5.1455)
1988	1.0019*	-0.0210	-0.0306	0.0748
	(3.0311)	(-0.1894)	(-1.1709)	(1.0497)
1989	0.5366*	-0.2011*	-0.0299	0.0187
	(2.6106)	(-3.1427)	(-1.8814)	(0.4440)
1990	-0.2213	-0.0493	0.0221	0.0587
	(-1.2837)	(-0.8983)	(1.6644)	(1.8913)
mean	0.276	0.0177	-0.011	0.0576*
standard deviation	0.568	0.116	0.041	0.095
t-value	2.33	0.729	-1.276	2.916

Table 3.1: Coefficients of model 3.1 which is selected by the ATFR procedure (t statistics in parenthesis). The values designed with * are significant at a 5 per cent level.

year	*Intercept*	β	$\ln(MV)$	$\ln(TOB)$
1968	0.1616	0.0942	0.0174	0.0515
1969	0.0846	0.0451	0.0085	0.0255
1970	0.0978	0.0550	0.0096	0.0306
1971	0.3162	0.1630	0.0311	0.0864
1972	0.1091	0.0473	0.0099	0.0262
1973	0.1182	0.0460	0.0106	0.0271
1974	0.1392	0.0600	0.0124	0.0373
1975	0.1068	0.0468	0.0098	0.0307
1976	0.1387	0.0623	0.0129	0.0399
1977	0.1273	0.0549	0.0116	0.0392
1978	0.1209	0.0465	0.0102	0.0337
1979	0.1357	0.0476	0.0117	0.0370
1980	0.1378	0.0490	0.0121	0.0342
1981	0.1623	0.0605	0.0145	0.0379
1982	0.1971	0.0616	0.0175	0.0425
1983	0.1741	0.0528	0.0150	0.0442
1984	0.1953	0.0745	0.0168	0.0434
1985	0.2708	0.1060	0.0234	0.0649
1986	0.1762	0.0614	0.0152	0.0390
1987	0.2018	0.0438	0.0160	0.0371
1988	0.3305	0.1110	0.0262	0.0712
1989	0.2055	0.0640	0.0159	0.0422
1990	0.1724	0.0549	0.0133	0.0310

Table 3.2: Standard deviations of the estimated coefficients in model 3.1

over the whole period of 23 years. Thus, we neglect existing structural changes over time. That is, in order to ensure a simple structure of the model, we use a wrong model. Hence, the smaller t-values in the former case are more realistic than the bigger ones in the pooled data case.

Note that model (3.1) gives a remarkable improvement in comparison with the linear model

$$E\,R = b_1 + b_2\beta + b_3 MV + b_4 B/M. \tag{3.3}$$

Fitting model (3.3) for each year we get the following characteristics (averaged over all years).

cross validation value	0.0940
coefficient of determination	0.0548
cross-validated coefficient of determination	0.0

For comparison, a semiparametric analysis was performed while estimating the optimal model from the classes $\mathcal{M}_2$ to $\mathcal{M}_5$, which are described in *2* to *5*, respectively. The corresponding characteristics are

	C	B	R^2
2. kernel estimator	0.0860	0.0359	0.1020
3. smooth link funktion	0.0870	0.0322	0.1120
4. additive model	0.0966	0.0	0.1023
5. semiparametric additive model	0.0929	0.0	0.1121

That is, the functional relationship between the stock returns R and the explanatory variables β, MV and B/M is well approximated by the obtained "almost optimal" parametric model (3.1). Semiparametric modelling does not lead to better results, because the nonparametric part seems to increase essentially the estimation error.

References

Banz, R.W. (1981). The Relationship Between Returns and Market Value of Common Stocks. *Journal of Financial Economics* **9**, 3-18.

Bunke, O. (1992). Semiparametric Modelling for a Variable Depending on Time and Explanatory Variables. In P. G. M. van der Heyden et al. (editors), *Statitical Modelling*, Verlag J. Eul, Köln, 115-126.

Bunke, O. (1997). Semiparametric Estimation and Prediction for Time Series Cross Sectional Data. *Discussion Paper*, Sonderforschungsbereich 373, Humboldt University, Berlin, to appear.

Bunke, O., Droge, B. & Polzehl, J. (1995). Model Selection, Transformations and Variance Estimation in Nonlinear Regression. *Discussion Paper* **52**, Sonderforschungsbereich 373, Humboldt University, Berlin.

Droge, B. (1992). On a Computer Program for the Selection of Variables and Models in Regression Analysis. In V. Fedorov, W. G. Müller and I. N. Vuchkov (editors), *Model-Oriented Data Analysis*, Springer, Heidelberg, 181-192.

Fama, E.F. & French, R.F. (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance* **47**, 427-465.

Fama, E.F. & French, R.F. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* **33**, 3-56.

Fama, E.F. & French, R.F. (1995). Size and Book-to-Market Factors in Earnings and Returns. *Journal of Finance* **50**, 131-155.

Fama, E.F. & MacBeth, J.D. (1973). Risk, Return and Equilibrium: Empirical Tests. *Journal of Political Economy* **81**, 607-636.

Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press.

Hastie, T.J. & Tibshirani, R.J. (1990). *Generalized Additive Models*, Chapman & Hall.

Loughran, T. (1996). Book-to-Market Across Firm Size, Exchange and Seasonality: Is There an Effect? University of Iowa, unpublished manuscript.

Appendix

Usual and cross-validated coefficients of determination for each year of model 3.1

year	usual	cross-validated
1968	0.08714	0.0278
1969	0.1469	0.1048
1970	0.0489	0.0
1971	0.2075	0.1058
1972	0.0375	0.0
1973	0.2205	0.1729
1974	0.2982	0.2607
1975	0.0764	0.0046
1976	0.0990	0.0384
1977	0.1348	0.0761
1978	0.0177	0.0
1979	0.0609	0.0049
1980	0.0385	0.0
1981	0.0203	0.0
1982	0.1138	0.0559
1983	0.0621	0.0
1984	0.2383	0.1886
1985	0.0723	0.0
1986	0.0864	0.0485
1987	0.189	0.1046
1988	0.0195	0.0
1989	0.1123	0.0540
1990	0.0406	0.0

P-values of the estimated coefficients in model 3.1

The following table contains the p-values which are the probabilities that the absolute value of the estimated coefficients is bigger than the corresponding t statistic. A p-value smaller than 0.01 corresponds to a significant regression coefficient at a 1 per cent level (”strongly significant”). A p-value between 0.01 and 0.05 corresponds to a ”significant” coefficient and a p-value between 0.05 and 0.1 to a ”weakly significant” one.

year	*Intercept*	β	$\ln(MV)$	$\ln(TOB)$
1968	0.0000	0.4022	0.0790	0.0879
1969	0.4963	0.0010	0.4980	0.0362
1970	0.2165	0.0523	0.1699	0.0191
1971	0.0000	0.1942	0.0000	0.0241
1972	0.5390	0.4062	0.2112	0.6525
1973	0.0001	0.1610	0.0001	0.0000
1974	0.0111	0.0394	0.0002	0.0000
1975	0.8348	0.2062	0.6043	0.0014
1976	0.0020	0.2350	0.1295	0.0030
1977	0.0000	0.1149	0.0000	0.9984
1978	0.9823	0.1767	0.9963	0.3271
1979	0.2481	0.9574	0.2370	0.0092
1980	0.0701	0.6744	0.0577	0.3390
1981	0.3907	0.9217	0.6951	0.1018
1982	0.1040	0.1087	0.8939	0.0005
1983	0.0027	0.4686	0.0325	0.1003
1984	0.2820	0.0011	0.0139	0.0006
1985	0.8716	0.0365	0.5155	0.0071
1986	0.0659	0.1209	0.0396	0.3457
1987	0.0002	0.3799	0.1436	0.0000
1988	0.0029	0.8501	0.2436	0.2956
1989	0.0100	0.0020	0.0619	0.6577
1990	0.2014	0.3706	0.0983	0.0607