Semi-parametric Estimation of a Characteristic-based Factor Model of Stock Returns

By

Gregory Connor and Oliver Linton

DISCUSSION PAPER 346
March 2000

FINANCIAL MARKETS GROUP
AN ESRC RESEARCH CENTRE

LONDON SCHOOL OF ECONOMICS

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

ISSN 0956-8549-346

Semiparametric Estimation of a Characteristic-based Factor Model of Stock Returns*

Gregory Connor
London School of Economics
Email: g.connor@lse.ac.uk

Oliver Linton
London School of Economics and Yale University
Email: o.linton@lse.ac.uk

February 10, 2000

Abstract

This paper develops a new estimation procedure for characteristic-based factor models of stock returns. It describes a factor model in which the factor betas are smooth nonlinear functions of observed security characteristics. It develops an estimation procedure that combines nonparametric kernel methods for constructing mimicking portfolios with parametric nonlinear regression to estimate factor returns and factor betas. Factor models are estimated for UK and US common stocks using book-to-price ratio, market capitalization, and dividend yield.

*We would like to thank seminar participants at London Business School, London School of Economics, and Oxford University for helpful comments.

1 Introduction

In a series of important papers, Fama and French (hereafter denoted FF), building on earlier work by Banz (1981), Basu (1977), Rosenberg, Reid and Lanstein (1985) and others, demonstrate that there are large return premia associated with size and value. Size is defined as the log of market capitalization; value is defined as the book-to-price ratio or a related valuation ratio such as the earnings-to-price ratio. These size and value return premia are evident in US data for the period covered by the CRSP/Compustat database (FF (1992)), in earlier US data (Davis (1994), and in non-US equity markets (FF (1998), Hodrick, Ng and Sangmueller (1999)).

FF (1993,1995,1996,1998) contend that these return premia can be ascribed to a rational asset pricing paradigm in which the size and value characteristics proxy for assets' sensitivities to pervasive sources of risk in the economy. Haugen (1995) and Lakonoshik, Shleifer and Vishny (1994) argue that the observed value and size return premia arise from market inefficiencies rather than from rational risk premia associated with pervasive sources of risk. They argue that these characteristics do not generate enough nondiversifiable risk to justify the observed premia. Similarly, MacKinlay (1995) argues that the return premia are too large relative to the return volatility associated with hedge portfolios designed to capture these characteristics, and this creates a near-arbitrage opportunity in the FF model. Daniel and Titman (1997) argue that the factor returns associated with the characteristics are partly an artifact of the FF estimation methodology. A key issue in this debate is the magnitude and significance of the pervasive factors associated with these corporate characteristics. To address this issue properly requires estimating a factor model of returns in which the factors are linked to these characteristics. This paper develops a new methodology for estimating this type of factor model.

Rosenberg (1974) proposes a simple and direct approach to estimating a characteristic-based factor model. He assumes that the factor betas in the model are equal to observed security characteristics or to known functions of observed security characteristics. So, for example, the factor beta associated with the value factor could be assumed to be equal to the firms' book-to-price ratio (suitably scaled). Given that the factor betas are directly observed in this way, the factor returns can be estimated by cross-sectional linear regression of equity returns on these characteristic-based factor betas.

A weakness of Rosenberg's approach is the need to assume that the observed characteristics are reasonable proxies for the factor betas. Consider, for example, FF's (1995,1996) contention that the book-to-price ratio is related to the security's sensitivity to an economy-wide "distress" factor. Using the book-to-price ratio as a proxy for the factor beta in a characteristic-based factor model requires more this. It requires that returns respond linearly to this "distress" factor in direct proportion to their book-to-price ratios.

FF (1993) develop an alternative approach to estimate the factor returns associated with characteristics. They sort securities according to their size and value characteristics and construct two-dimensional fractile portfolios. They use differences between the returns on large-size and small-size fractile portfolios (adjusted for the value characteristic) as an estimate of the size factor. Analogously, the difference between high-value and low-value fractile portfolios, adjusted for the size characteristic, serves as an estimate of the value factor. These factor return estimates do not require any strong assumptions on the relationship between factor betas and characteristics, only monotonicity and smoothness. However, there is no obvious way to generate standard errors for the estimates. Also, in order to estimate the factor betas, a set of time-series regressions must be run with the estimated factor returns as explanatory variables. This gives rise to an errors-in-variables problem in the estimated factor betas.

This paper develops an estimation methodology that combines elements of Rosenberg's and FF's procedures. It uses a characteristic-based factor model like Rosenberg's, except that the factor betas are unknown, nonlinear functions of observed characteristics, in the spirit of FF. The estimation methodology has two steps. The first step uses nonparametric kernel methods to construct factor-mimicking portfolios associated with a set of chosen values of the characteristics. The second step uses parametric nonlinear regression, with the collection of first step portfolio returns as the independent variable, to estimate the factor returns and factor beta functions. This new methodology allows for a more general functional form than Rosenberg's, and facilitates a range of approximate (asymptotic) statistical results not available with FF's procedure. It gives simultaneously estimated, consistent and asymptotically normal estimates of the factor returns and the factor beta functions, and approximate standard errors for all estimated parameters.

The model is applied to UK and US equities using three security characteristics: the book-to-price ratio, the log of market value, and dividend yield. First we estimate a two-characteristic model using the book-to-price ratio and log of market value and compare our estimates to FF-type estimates for the same data sets. Then we add dividend yield, and find that it is an additional source of pervasive factor risk in both the US and UK.

Section 2 presents the new estimation methodology. Section 3 applies it to the data. Section 4 summarizes the paper and suggests some further extensions and applications of the approach.

2 Methodology

2.1 Description of the Factor Model

We assume that there is an asymptotically large number of securities, indexed by $i = 1, \ldots, n$, and asset returns are observed for a fixed number of time periods $t = 1, \ldots, T$. We assume that the following characteristic-based factor model generates returns:

$$r_{it} = f_{zt} + \sum_{j=1}^{k} g_j(c_{ij}) f_{jt} + \varepsilon_{it}, \tag{1}$$

where r_{it} is the return to security i at time t; f_{zt}, f_{jt} are the factor returns; $g_j(c_{ij})$ the factor betas, c_{ij} the security characteristics, and ε_{it} the mean zero asset-specific returns. The factor returns f_{jt} are linked to the security characteristics by the factor beta functions $g_j(\cdot)$. We assume that $g_j(\cdot)$ is a smooth time-invariant function of the characteristics, but we do not assume a particular functional form. The zero-characteristic factor f_{zt} captures that part of common return not related to the characteristics; all assets have unit exposure to this factor. The observed security characteristics and asset returns are used to estimate the factor returns and factor beta functions.

There are two indeterminancies in the functions $g_j(\cdot)$. The first is additive. One can add an arbitrary constant a to any of the functions $g_j(\cdot)$ and subtract af_{jt} from f_{zt}, and the predictions of the returns model 1 are unchanged. To eliminate this indeterminacy, we impose the condition $g_j(0) = 0$ for all j, without loss of generality.

The second indeterminacy is multiplicative. One can multiply any $g_j(\cdot)$ by any non-zero constant and f_j by the reciprocal of the same constant and the predictions of the returns model (1) are unchanged. We assume that $g_j(1) \neq 0$ for each j. Without loss of generality we set $g_j(1) = 1$.

The identification constraints $g_j(0) = 0$ and $g_j(1) = 1$ are given intuitive content by the choice of units of c_{ij}. Let c^*_{ij} denote the raw characteristic j for asset i. We rescale the raw characteristics linearly so that the cross-sectional average of c_{ij} equals zero and the cross-sectional standard deviation equals one. The constraint $g_j(0) = 0$ means that the factor return f_{zt} is the common-factor-related return of an asset with "average" characteristics. The constraint $g_j(1) = 1$ means that over the interval $[0, 1]$ measured in units of standard deviation the increase in factor beta equals one.

2.2 Kernel-based Portfolio Weights for Factor-Mimicking Portfolios

In this subsection we present a new method for creating factor-mimicking portfolios, based on non-parametric kernel methods. Our purpose in developing this new technique is the estimation of characteristic-based factor models, but there are other potential applications. For example, the technique could be used for the construction of benchmark portfolios in event studies or in performance measurement of managed portfolios.

Our methodology is partly founded on the earlier work of FF (1993) and we very briefly summarize their approach. FF rank securities by two characteristics, size and book-to-price (BTP), and perform a bivariate sort of the securities into fractiles. They use three fractiles for BTP and two for size, so the bivariate sort gives a total of six fractiles: large size/high BTP, large size/medium BTP, large size/low BTP, small size/high BTP, small size/medium BTP, small size/low BTP. They group the assets into capitalization-weighted portfolios of the securities within each fractile. For each characteristic, the average difference between the returns on a collection of high and low fractile portfolios, screened to preserve a common exposure to the other characteristic, serves as the estimates of the factor returns associated with that characteristic. Specifically they define:

$$\begin{aligned}\text{Size factor return} &= 1/3[(\text{large size/high BTP portfolio return} \\ &\quad \text{-small size/high BTP portfolio return}) \\ &\quad +(\text{large size/medium BTP portfolio return} \\ &\quad \text{-small size/medium BTP portfolio return}) \\ &\quad +(\text{large size/low BTP portfolio return} \\ &\quad \text{-small size/low BTP portfolio return})] \qquad (2)\end{aligned}$$

$$\begin{aligned}\text{Book-to-price factor return} &= 1/2[(\text{large size/high BTP portfolio return} \\ &\quad \text{-large size/low BTP portfolio return}) \\ &\quad +(\text{small size/high BTP portfolio return} \\ &\quad \text{-small size/low BTP portfolio return })] \qquad (3)\end{aligned}$$

The first step in our new methodology is a kernel-based variant of FF's portfolio construction technique. Instead of target ranges for the characteristics (such as high, medium and low), we create a set of portfolios, each one designed to capture one from a grid of target characteristic vectors. Instead of capitalization-weighting for the portfolios, we use kernel-weighting, where the kernel weights are constructed to trade-off portfolio diversification against the distance of each asset's characteristic vector from the target vector.

We choose m distinct target values for each of the k characteristics, where the values must include the two values used to set the scale of the factors, zero and one, and these are listed first and second. The other $m-2$ values need not be the same for the k different characteristics. Let $c_{ej}; e = 1, \ldots, m$, $j = 1, \ldots, k$ denote the chosen values. The grid of target characteristic vectors consists of all m^k

combinations of the m chosen target values over the k characteristics. Define $m^* = m^k$. In our empirical implementation, m is not the same across all characteristics and we do not utilize all grid points; hence we use the more general notation m^* rather than m^k. For simplicity of notation, we continue to assume $m^* = m^k$ in the statistical derivation. Let c_h, $h = 1, \ldots, m^*$, denote the k-vectors of target characteristics. Define the Gaussian product kernel associated with the h^{th} target vector as $K_{hi} = (2\pi b^2)^{-1/2} \exp(-\sum_{j=1}^{k}(c_{ij} - c_{hj})^2/2b^2)$, where b is the bandwidth. For bandwidth b define the kernel-based portfolio weights for the h^{th} combination as $\omega_{hi} = K_{hi}/\sum_{i=1}^{n} K_{hi}$ and the kernel-based portfolio return as $\widehat{y}_{ht} = \sum_{i=1}^{n} \omega_{hi} r_{it}$. From the perspective of statistical theory, $\widehat{y}_{ht}$ can be interpreted as a nonparametric kernel estimator of the conditional expectation of r_{it} given $c_i = c_h$. From the perspective of financial econometrics, it can be viewed as the return on a well-diversified portfolio designed to have (approximately) the target characteristics c_h. The approach in this paper combines these two perspectives.

Now we show that the kernel-based portfolio returns converge to linear combination of factor returns, with asymptotically normal and independent residuals. To do this, we apply a result from kernel regression theory. For each t define the functions $g_t^*(c) = f_{zt} + \sum_{j=1}^{k} g_j(c) f_{jt}$. Using (1) it follows immediately that

$$r_{it} = g_t^*(c_i) + \varepsilon_{it}. \tag{4}$$

For a given t, equation (4) can be viewed as a multivariate nonparametric regression problem. Our kernel-based portfolio return for characteristic combination h is the Watson-Nadaraya estimate of $g_t^*(c_h)$ using a multivariate Gaussian kernel. We use a result from Bierens (1994) to guarantee the consistency and asymptotic normality of the estimates. We assume that the observed characteristic vectors of the assets $c_i, i = 1, \ldots, n$ are an n-sample realization of a k-vector random variable. Let $p(c_i)$ denote the marginal density function of this k-vector random variable evaluated at the point c_i. Let $\sigma_{th}^2(c) = E[\varepsilon_{it}^2 | c_i = c]$ and $\sigma_{th}^3(c) = E[|\varepsilon_{it}|^3 | c_i = c]$. We impose the following conditions:

Assumption A. The vector $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ is independently distributed across $i = 1, \ldots, n$, and satisfies $E(\varepsilon_{it}\varepsilon_{is} | c_i) = 0$ with probability one for all $t \neq s$ and for all i. The functions $g_t^*(c_i)$ and $p(c_i) g_t^*(c_i)$ and their first and second partial derivatives are continuous and uniformly bounded. The density $p(c_h) > 0$ for each h. The moments $\sigma_{th}^2(c)$ and $\sigma_{th}^3(c)$ are uniformly bounded. The bandwidth satisfies $b_n = o(n^{-1/(k+5)})$.

Theorem 1 *Given Assumption A and the other conditions mentioned in the text, then as* $n \to \infty$,

$$(n b_n^k)^{1/2} (\widehat{y}_{ht} - g_t^*(c_h)) \Longrightarrow N(0, \sigma_{th}^2),$$

where

$$\sigma_{th}^2 = (\sigma_t^2(c_h)/(p(c_h)\sqrt{2})). \tag{5}$$

Furthermore, $(nb_n^k)^{1/2}(\widehat{y}_{ht} - g_t^*(c_h))$ *are asymptotically independent across* h *and* t*.*

PROOF. See the Appendix.

In order to apply Lemma 1 we need to approximate the asymptotic variance in (5). We use a standard procedure from nonparametric estimation of residual variances (see Pagan and Ullah (1999, p 95) for discussion). Each marginal density $p(c_h)$ is consistently estimated by $\hat{p}(c_h) = \frac{1}{n}\sum_{i=1}^{n} K_{hi}$ and, given observation of the asset-specific returns in (1), each variance $\sigma_t^2(c_h)$ is consistently estimated by $\hat{\sigma}_t^2(c_h) = \frac{1}{n}\sum_{i=1}^{n} K_{hi} r_{it}^2 - \left(\frac{1}{n}\sum_{i=1}^{n} K_{hi} r_{it}\right)^2$.

Using Lemma 1, it is easy to create a parallel to FF's factor return estimates shown in equations (2) and (3). Consider two target characteristic vectors c_h and $c_{h'}$ which are equal in all components except that c_h has characteristic j value 1 whereas $c_{h'}$ has characteristic j value 0. Using the scaling assumptions $g_j(1) = 1$ and $g_j(0) = 0$, it is easy to see that the difference in the target factor betas of the two associated kernel-based portfolios equals one for factor j and zero for all other factors. Applying Lemma 1, the return difference between these two portfolios provides a consistent, asymptotically normal estimate of factor return j. From among the $m^* = m^k$ combinations of characteristics, there are m^{k-1} pairs that differ only in characteristic j and have values 0 and 1 respectively for this characteristic. Hence, for each factor j, we have m^{k-1} asymptotically independent estimates of the time t factor return. In parallel with FF, we could use the average across these pairs of matched portfolio returns as the factor return estimate, that is,

$$\hat{f}_{jt} = \frac{1}{m^{k-1}} \sum_{h=1}^{m^*} (\delta_{h1} - \delta_{h0}) y_{ht}, \tag{6}$$

where the dummy variable $\delta_{h1}(\delta_{h0})$ equals one if mimicking portfolio h has target characteristic one (zero) for factor j and equals zero otherwise. Using this framework the factor return estimates have a well-defined asymptotic distribution, as described in Lemma 2.

Lemma 2 *As* $n \to \infty$,

$$(nb_n^k)^{1/2}(\hat{f}_{jt} - f_{jt}) \Longrightarrow N(0, \sigma_j^2), \tag{7}$$

where $\sigma_j^2 = \frac{1}{m^{k-1}} \sum_{h=1}^{m^*} (\delta_{h1} - \delta_{h2})^2 (\sigma_t^2(c_h)/(p(c_h)\sqrt{2}))$, $j = 1, \ldots, k$.

PROOF. See the Appendix.

The FF-type estimates described in Lemma 2 are consistent and asymptotically normal, but have two weaknesses. One, the estimate of each factor return uses information from only a subset of the kernel-based portfolios. Two, the estimator only gives estimates of the factor returns, not the factor betas. In the next section we present an alternative estimator which uses information from all the kernel-based portfolios simultaneously and produces joint estimates of all the factor returns and of all the factor beta functions evaluated at the target characteristics.

2.3 Joint Estimation of the Factor Beta Functions and Factor Returns Using Nonlinear Regression

In this subsection we use the kernel-based portfolio returns described in the last subsection as independent variables in a nonlinear regression system. The unknown parameters in this parameterized system are the realized factor returns and the beta functions evaluated at the target characteristics. The regression is nonlinear because it includes products of factor returns and factor betas.

Recall from the last subsection the definition of the m^* kernel portfolios, covering all combinations of the m target characteristics for each of the k factors. The returns on all of these kernel-based portfolios can be written as a pooled regression with m^* 'cross-sectional' observations and T time series observations:

$$\widehat{y}_{ht} = f_{zt} + \sum_{j=1}^{k} g_j(c_{hj}) f_{jt} + \widehat{u}_{ht} \tag{8}$$

$$\widehat{u}_{ht} = [\sum_{j=1}^{k} \{\sum_{i=1}^{n} \omega_{hi} g_j(c_{ij}) - g_j(c_{hj})\} f_{jt}] + \sum_{i=1}^{n} \omega_{hi} \varepsilon_{it}. \tag{9}$$

Note that in (8) the nonparametric functions $g_j(\cdot)$ are evaluated at mk points, hence we define the mk-vector $g_{ej} = g_j(c_{ej}), e = 1, \ldots, m, j = 1, \ldots, k$. The factor model scaling assumptions $g_j(0) = 0$ and $g_j(1) = 1$ imply that $g_{1j} = 0$ and $g_{2j} = 1$ for each j. We treat the remaining $(m-2)k$ components of g_{ej} as parameters to estimate, along with the $(k+1)T$ factor returns $f_{zt}, f_{jt}; t = 1, \ldots, T; j = 1, \ldots, k$. Let $\beta = \{g_{ej}, f_{zt}, f_{jt}; e = 1, \ldots, m; j = 1, \ldots, k; t = 1, \ldots, T\}$ denote the $(m-2)k + (k+1)T$ vector of parameters and let β_0 be the true vector. We rewrite (8) as a nonlinear regression equation with m^*T observations and $p = (m-2)k + (k+1)T$ unknowns

$$\widehat{y}_{ht} = f_{zt} + \sum_{j=1}^{k} g_{ej} \delta_{ej,h} f_{jt} + \widehat{u}_{ht},$$

where the parameters β are subject to the zero/one restrictions described above. The dummy variable $\delta_{ej,h}$ equals one if mimicking portfolio h has target value c_{ej} for factor j, and zero otherwise. Note that there are a fixed finite number of observations $h = 1, \ldots, m^*$ and $t = 1, \ldots, T$, but that the error terms are asymptotically independent across h, and are individually of small order.

For a chosen parameter vector β, define the fitted values in the usual way

$$y_{ht}(\beta) = f_{zt} + \sum_{j=1}^{k} g_{ej} \delta_{ej,h} f_{jt}. \tag{10}$$

The nonlinear weighted least squares estimator, $\hat{\beta}$, is any minimizer of the criterion function

$$Q_n(\beta) = \sum_{h=1}^{m^*} \sum_{t=1}^{T} \widehat{v}_t(c_h)(\widehat{y}_{ht} - y_{ht}(\beta))^2 \tag{11}$$

over $\beta \in \mathcal{B} \subseteq \mathbb{R}^p$, where the parameter set $\mathcal{B}$ reflects the restrictions on β. We introduce the weighting function $\widehat{v}_t(c_h)$ to account for error heteroscedasticity, and it is allowed to be estimated from the data. The criterion function $Q_n(\beta)$ is a quartic polynomial in the parameters, and under reasonable conditions will have a global minimum, which will be locally unique. We can also show that any solution obeys the following first-order condition:

$$\frac{\partial}{\partial \beta} Q_n(\widehat{\beta}) = 0. \tag{12}$$

This enables to use an iterative weighted least squares procedure to find the minimum. The actual algorithm we use to solve (12) exploits the bilinear structure of the regression function and is described in the appendix.

Theorem 3 *Suppose that $m^* > mk^2$ and that for each t and h, the weighting function $\widehat{v}_t(c_h) \to^p v_t(c_h)$ as $n \to \infty$, where $v_t(c_h) > 0$. Then, the least squares estimate defined by (11) exists with probability tending to one and obeys the first-order condition (12). Furthermore, $\hat{\beta} \xrightarrow{p} \beta_0$.*

Proof. See the Appendix.

Let $\Gamma(\beta) = \partial y / \partial \beta'$ be the $m^*T \times p$ matrix of derivatives of $y_{ht}(\beta)$ with respect to each element of the parameter vector β. Define the $p \times p$ matrix

$$H(\beta) = \Gamma(\beta) V \Gamma(\beta)',$$

where V is the $m^*T \times m^*T$ diagonal matrix whose elements are $v_t(c_h)$. We shall suppose that $H(\beta)$ is a nonsingular matrix for $\beta = \beta_0$ and that β_0 is an interior point of the parameter space $\mathcal{B}$. Now we show that the least squares estimate has an asymptotic normal distribution with known covariance matrix. Let D be the $m^*T \times m^*T$ asymptotic covariance matrix of $(nb_n^k)^{1/2}\widehat{u}$, where $\widehat{u}$ is the m^*T vector with typical element $\widehat{u}_{ht}$. D is a diagonal matrix under our assumptions.

Theorem 4 *As $n \to \infty$,*

$$(nb_n^k)^{1/2}(\hat{\beta} - \beta_0) \Longrightarrow N(0, \Sigma), \quad \text{where } \Sigma = H_0^{-1}\Gamma_0 V D V \Gamma_0' H_0^{-1},$$

where $H_0 = H(\beta_0)$ and $\Gamma_0 = \Gamma(\beta_0)$. When $V = D^{-1}$, we have $\Sigma = \left(\Gamma_0 D^{-1} \Gamma_0'\right)^{-1}$.

Proof. See the Appendix.

Note that the covariance matrix Σ described in Theorem 4 covers both the factor return estimates and the estimates of the factor betas since we are estimating both sets of parameters simultaneously. The asymptotic variance matrix can be consistently estimated by

$$\widehat{\Sigma}_\beta = \widehat{H}^{-1}\widehat{\Gamma}\widehat{V}\widehat{D}\widehat{V}\widehat{\Gamma}'\widehat{H}^{-1},$$

where $\widehat{H} = H(\widehat{\beta})$ and $\widehat{\Gamma} = \Gamma(\widehat{\beta})$ and $\widehat{D} = \text{diag}(\widehat{\sigma}^2_{th})$ is an estimate of D, while $\widehat{V}$ is the diagonal matrix with elements $\widehat{v}_t(c_h)$.

3 Empirical Analysis

3.1 Data

The data set consists of monthly returns and beginning-of-month security characteristics for UK and US equities over the 150-month period January 1986 – June 1998. All the data come from the proprietary database of BARRA Inc. For the UK (US) sample, the number of securities in the monthly samples varies from a low of 489 (1685) to a high of 811 (1998) with an average of 582 (1817) per month. The data set includes three security characteristics: book-to-price ratio, log of market value, and dividend yield. The book-to-price ratio is the ratio of the current book value of common equity to current market capitalization. If book value is negative, the book-to-price ratio is set equal to zero. Market value is defined as the number of common shares outstanding times the current share price. Dividend yield is defined as the most recent twelve months of dividend payments divided by the current stock price.

The "current" figures (share prices, shares outstanding) are from the last trading day of the previous calendar month. The book value of common equity is from the most recent annual report that is available on or before the last day of the previous month. There is no look-ahead bias in the annual report data since it is stored using when-reported dating rather than fiscal-period dating.

Table 1 shows some distribution statistics for the three security characteristics. Each month, we truncate the three characteristics at -3 and +3 standard deviations from the mean. This seems justified since values beyond these extremes are unlikely to provide additional information about factor betas. The standardized characteristics (hereafter the size, value and yield characteristics) equal the truncated characteristics minus the cross-sectional mean at time t, divided by the cross-sectional standard deviation at time t. After this transformation, the standard deviations and means all equal respectively one and zero and are not shown (note that the means and standard deviations are calculated *after* the truncation). All three characteristics are leptokurtic and positively skewed, a feature of these characteristics pointed out earlier by Brennan, Chordia and Subrahmanyan (1997).

Table 2 shows the correlation between the three standardized characteristics. Although these correlations are based on the stacked sample with both cross-sectional and time-series components, they only capture cross-sectional correlation due to the monthly cross-sectional standardization of the characteristics. There is fairly strong positive correlation between the value and yield characteristics.

3.2 The Two-Characteristic Case

In this subsection we estimate the model with three factors, based on the size and value characteristics (plus the zero-characteristic factor). This gives a concise model and allows for comparison to FF-type factors. In a later subsection we extend the analysis to four factors based on three characteristics, by adding the yield characteristic.

To begin estimation of the model we need to choose a set of target characteristics, and a bandwidth, b. For both the size and value characteristics we use target values in the range –1.25 to +1.75 spaced at intervals of 0.25, giving thirteen potential target values for each of the two characteristics and therefore 169 (13x13) potential combinations of the two. Kernel-based estimates have poor finite-sample performance at the extremes of the multivariate distribution of the explanatory variables. Therefore we limit the target vectors to those which fall within an ellipse of zero, with the shape of the ellipse determined by the covariance matrix of the factors. In particular we use the target vectors c_h such that $c_h' \Sigma_c^{-1} c_h < 4.10$, which would correspond to an optimally chosen 95% range under multivariate normality of the characteristics. We know from Table 2 that the distribution is not normal, but still this seems a reasonable metric for restricting the set of values. After deleting the values outside the ellipse, we have a total of 159 target characteristic vectors in both countries.

The bandwidth choice involves a trade-off between having kernel portfolios whose constituent asset characteristics more closely match the target values (smaller bandwidth) versus having portfolios with lower asset-specific variance (larger bandwidth). We use the rule-of-thumb methodology explained in Silverman (1986) to select bandwidth. This involves specifying a parametric model for the data generating process, computing the (integrated) mean squared error of the estimator under this hypothesis, and then finding the bandwidth that would minimize that mean squared error. For the specification of $g_j(c_{ij})$ we take a fourth order polynomial in c_{ij}, while the marginal distributions of c_i is taken as normal, and the error term is assumed homoskedastic. The parameters of this specification are estimated by least squares and sample moments, and these are then plugged into the formula for the optimal bandwidth. In the UK (US), the optimal bandwidths vary from a minimum of .267 (.263) to a maximum of .449 (.441) with an average of .333 (.344) across the 159 target characteristic vectors.

Figures 1-4 examine a representative kernel portfolio to illustrate the nature of the portfolio weights. We use the "middle" month of the sample (month 75, which is March 1995) and pick, for illustrative purposes, target characteristics of .5 for value and .5 for size. The portfolio weights are shown in relationship to the value and size characteristics of the assets. One can think of the graphs as two-dimensional projections of a three-dimensional graph relating the portfolio weights to both characteristics simultaneously. The kernel portfolio puts highest weights on assets which are close to the target values in both characteristic. For this particular case the UK portfolio has a maximum weight of .02841 which applies to an asset with value characteristic .40858 and size characteristic .43864; note that both values are near the target value of .5. The largest US portfolio weight in this particular case is .00603 for an asset with value characteristic .45292 and size characteristic .48551. In both the US and UK case, there are assets which have characteristics closer to .5 for either size or value, but not for both.

Note that these portfolio weights come from one of the 159 kernel portfolios used in the estimation of the factor returns in this month. As we discuss later, our factor return estimate for a given month is basically the return on a "portfolio of portfolios", consisting of a particular linear combination of all 159 kernel portfolios. Graphs of the resulting factor return portfolios are described in the next subsection.

Table 3 shows equally-weighted R-squared statistics and regression-weighted residual variances. For each factor, we also re-estimate the model and calculate R-squared and residual variance after dropping the factor. The difference between the R-squared statistics with and without a given factor is a simple descriptive measure of the marginal explanatory power of the factor. The difference between the regression-weighted error variance with and without a factor times the number of regression observations m^*T is a likelihood ratio test of the null hypothesis that the factor return equals zero for all t. This statistic has an approximate $\chi^2(T)$ distribution under the null hypothesis. The p-values from this test are shown in the table. For all three factors we can strongly reject this null hypothesis.

The last three columns of Table 3 are descriptive statistics: the time-series means, variances and Sharpe ratios (annualized mean divided by annualized standard deviation) of the estimated factor returns. The time-series mean can be viewed as a measure of the risk premium associated with each factor. The Sharpe ratio provides a natural metric for comparing the reward/risk ratios of the factors. The 150-month sample period is not long enough to draw reliable conclusions about expected returns. This paper concentrates on the estimation of the factor model, not on the implications for tests of expected return models.

Table 4 shows the estimates of the factor beta functions at the specified target characteristic values. Recall that both beta functions are set to zero at zero and to one at one, as identification conditions. The pointwise functions from target characteristics to factor betas are monotonically increasing at all points in both markets. It is interesting that the beta functions look reasonably close to linear, which lends some empirical support for Rosenberg's (1974) simplifying assumption of linearity. The uniformly positive slope of the functions has implications for analysis of both of the size effect and the value effect in equity markets. If the return premia associated with these characteristics are due to factor-based risk premia, then a marginal return premia should apply across the whole spectrum of firms, not just to low-capitalization firms or to firms with very low book-to-price ratios. This is because, under a standard factor beta pricing model, the difference in return premia between two firms is proportional to the difference in factor beta.

Note that in the US case the value factor beta function has a steeper slope below zero ("low-value" firms) than above zero ("high-value" firms). This seems to imply that the value factor betas capture something other than just sensitivity to financial distress. The marginal increase in sensitivity to financial distress for a marginal change in the book-to-price ratio should be fairly small for "low-value" firms.

3.3 Comparison with Fama and French Factors

An obvious comparison for our three estimated factors are market, size and value factors estimated by the method of FF. We modify their procedures slightly to suit our data set and our standardization of the characteristics. We define a "big" stock as one with size characteristic greater than or equal to one and a "small" stock as any stock with size characteristic less than that. We define a "medium" value stock as one with a value characteristic between -.5 and +.5 and a "high" ("low") value stock as any falling on the positive (negative) side of this closed interval. We calculate capitalization-weighted portfolios and estimate value and size factors using the portfolio return differences described in equations (2) and (3). We define the market factor as the return on a capitalization-weighted portfolio of all stocks. FF use the market factor in parallel to our use of the zero-characteristic factor.

Table 5 shows the correlation matrix for all the factors. There is obviously a positive correlation between the pairs of equivalent factors estimated by the two methods; these are highlighted using bold italic font. These correlations are only modestly high given that the two estimation procedures are applied to identically the same underlying dataset, and rely on nearly the same factor definition. The lowest of these correlations is between the two estimated size factors in the UK, with a correlation of 0.6265.

The cross-national correlations of particular factors are shown in bold. There is a very strong correlation between the two "market" factors and between the two "zero-characteristic" factors across the two countries. Previous research (Capaul, Rowley and Sharpe (1993), FF (1998)) has found that value and size factors are only weakly correlated across national markets and we find the same. We attribute the somewhat higher cross-market correlation for the FF-type factors than for the kernel-based factors to the use of capitalization-weighting in the FF procedure. The FF approach puts most of the weight on the larger stocks in each market. The larger-capitalization stocks tend to be more internationally visible and have stronger cross-border links, raising the cross-border correlations between the factor estimates.

Recall that Figures 1 - 4 showed the portfolio weights for a particular kernel portfolio in a particular month. Figures 5-20 take the analysis of portfolio weights a step further. As noted originally by Fama and MacBeth (1973), factor return estimates can be represented as the returns on portfolios of assets. For month 75, we calculate the portfolio of assets (where the assets are kernel portfolios) implicit in our nonlinear regression estimate of each of the three factors. The portfolio implicit in the zero-characteristic factor return estimate has weights summing to one, but the other two implicit portfolios have weights summing to zero. We rescale these weights so that the positive weights sum to one (and obviously the negative weights sum to minus one). This is a natural scaling for zero-cost portfolios for the purpose of measuring their diversification. Next, we decompose each kernel portfolio into its constituent asset weights, and so derive the implicit portfolio in terms of the primitive assets. It is also easy to calculate the portfolio weights implicit in the FF-type factor returns in the same way.

To save space, we show these portfolio weights only for the value and size factors (those for the zero-characteristic and market factors are available from the authors). As in Figures 1 - 4 above, the portfolio weights are graphed against both the value and size characteristics of the constituent assets. Consider Figures 5 - 12, which give the portfolio weights for our factor return estimates. To understand the graphs intuitively, suppose that the estimation routine has found factor betas which are exactly linear in each characteristic. Then the estimation routine will try to construct a portfolio which is well-diversified, has near-unit average slope in terms of the relevant characteristic, and near-zero average slope in the other characteristic. This is only an approximate guide, of course, since the estimated factor betas shown in Table 4 are not exactly linear. However one can discern the tendency toward a roughly linear average slope in the size factor/size characteristic graphs (Figures 5 and 9) and value factor/value characteristic graphs (Figure 7 and 11). Figures 6, 8, 10 and 12, which graph each factor portfolio against the other characteristic, have average slopes near zero. All the portfolios are very well-diversified: in the UK (US) the largest magnitude weight for the value factor portfolio is .0201 (.00649) and for the size it is .0160 (.0056).

Figures 13-20 show the portfolio weights for the FF-type factor return estimates in the same month. The portfolios are fairly well-diversified: in the UK (US) the largest magnitude weight in the HML portfolio is .071 (.081) and in the BMS portfolio it is .106 (.054). Comparison of the factor weights shows that our approach and the FF approach rely on quite different portfolio weighting schemes. It is notable that the UK HML portfolio has 27% of its positive weights in four high-capitalization assets. These same four assets have extremely small weights (effectively, zero weights) in the corresponding UK value factor portfolio. The FF approach has the advantage that the factor return estimates are based on investable portfolios, whereas our implicit portfolios are too diverse to be investable. Also, their use of capitalization-weighting means that their portfolio weights are related to the economic importance of the constituent assets. Our weighting scheme is driven solely by statistical error minimization criteria.

FF (1993, Table 6) find that the estimated betas for the market factor are very close to unity for a wide variety of portfolios, in a factor model including the value and size factors. Our model has this unit-market-beta feature imposed as an assumption, since the betas of all assets against the zero-characteristic factor are set equal to one in the definition of the factor model. Having estimated the factors, we can test this assumption by re-estimating the factor betas without imposing the prior restriction on them. We do this by unconstrained time-series regression of each kernel portfolio's returns on the factor returns. We also regress each kernel portfolio'ss returns on the FF-type factor returns. The results are shown in Table 6. We find in both cases that the betas to the "first factor" (market factor or zero-characteristic factor) are very close to a unit vector across the kernel portfolios.

This finding is important for three reasons. First, as noted previously by FF, it means that the risk premium associated with the "first factor" is poorly identified if the betas are all close to one, or unidentified if they all equal one, using cross-sectional regression methods. A Fama-MacBeth cross-sectional regression of sample means on a cross-sectional intercept and factor betas has weak power if all the market factor betas are all near unity; it is not identified if they all equal unity. Similarly, the Hansen and Jagannathan (1997) test comparing beta pricing to risk-neutral pricing will not be able to reject risk-neutral pricing if all the betas equal one. This is because "risk-equivalent" pricing holds – the assets have the same level of market factor risk and so the associated risk premium cannot be identified cross-sectionally. Other types of tests such as Gibbons, Ross and Shanken (1989) that do not depend on cross-sectional variation in the factor betas are still valid.

Second, this finding has implications for the best factor estimation strategy. It is inefficient to use estimated betas when estimating this "first factor" if the betas can be proxied by a constant term with little or no loss of accuracy. Third, it means that the zero-beta return is poorly or not identified. With all unit betas for a factor, there does not exist a unit-cost portfolio of equities with zero exposure to the factors.

3.4 Adding a Dividend Yield Factor

Next we add yield as a third characteristic in the model. Yield is an obvious candidate for a pervasive-risk-related characteristic since it captures the part of each asset's return due to cash flow directly received by the investor. If investors' preferences for yield versus capital gains vary through time, or the relative performance of high-payout versus low-payout firms varies, then these pervasive influences on returns will be related to yield.

For dividend yield we use a range from -1 to +1.75 with 0.25 intervals between target values. Combined with the ranges for size and value this gives a total of 13x13x12 = 2028 potential target combinations for the three characteristics. As in the two-characteristic model, we limit the target vectors to those which fall within an ellipse of zero, with the shape of the ellipse determined by the covariance matrix of the factors. In particular we use the target vectors c_h such that $c_h' \Sigma_c^{-1} c_h < 6.25$, which would correspond to an optimally chosen 95% range under multivariate normality of the characteristics. This gives a number of target vectors of 1829 in the US and 1926 in the UK (the numbers differ due to the differences between the characteristic covariance matrices in the two countries). We choose an optimal bandwidth by identically the same procedure as in the two-characteristic case. In the UK (US), the optimal bandwidths vary from a minimum of .264 (.252) to a maximum of .517 (.537) with an average of .3647 (.3651) across the 1926 (1829) target characteristic vectors.

Tables 7 is analogous to Table 3 for the three-characteristic case. Adding dividend yield increases the R-square in the UK (US) from .8873 to .9083 (from .9030 to .9418). In both countries the decrease in residual variance is highly significant for each factor, based on the likelihood ratio test described earlier. The R-squares in Tables 3 and 7 are not directly comparable since the model estimation described in the two tables uses a different set of underlying kernel portfolios.

Table 8 shows the estimated beta functions and their t-statistics. As in the two-characteristic estimates, the beta functions are monotonically increasing between all pairs of points. In the US, adding the dividend yield factor noticeably changes the shape of the value factor beta function. In particular, the value factor betas of "low-value" firms are diminished in magnitude. The same effect is discernible (although more muted) in the UK. Note from Table 2 that value and yield are quite strongly correlated. One might think of the value factor in the two-characteristic model as capturing both value and yield-related factor shocks. When the yield factor is added explicitly, the shape of the value factor beta function changes.

4 Summary

This paper describes a new estimation methodology for characteristic-based factor models. The methodology combines elements of Rosenberg's (1974) linear specification and the unrestricted portfolio grouping procedure of Fama and French (1993). This new methodology allows for a more general specification than Rosenberg's, and facilitates a range of approximate (asymptotic) statistical results not available with Fama and French's procedure. The methodology has two steps. The first step uses nonparametric kernel methods to construct mimicking portfolios for a chosen grid of values of the characteristics. The second step uses parametric nonlinear regression to estimate factor betas and factor returns simultaneously, using the collection of first-step mimicking portfolio returns as the independent variable.

The model is applied to UK and US equities using three security characteristics: book-to-price ratio, log of market capitalization, and dividend yield. We confirm some of Fama and French's findings using a factor model based on the book-to-price ratio and log of market capitalization. We find an independent role for dividend yield as a source of factor risk.

There are a number of possible extensions and applications of our findings. Daniel, Grinblatt and Titman (1997) provide a framework for using characteristic-based benchmarks in performance measurement. Our new methodology for the construction of characteristic-based mimicking portfolios has obvious applications there. Constructing normal performance benchmarks in event studies is a closely related problem, and our new methodology might prove useful. Fama and French (1993,1995) stress that an important research problem is explaining why these security characteristics provide so much information about return co-movements and about sample mean returns. Our findings do not answer this directly, but may provide useful input. In particular, one might attempt to explain our estimated factor model in terms of more primitive variables such as business cycles and interest rates and their relative effects on stock price for securities with particular characteristics.

A Appendix

Proof of Lemma 1. Using assumption A it follows immediately from Bierens (1994, Theorem 10.2.4) that for each c_h,

$$(nb_n^k)^{1/2}(\widehat{y}_{ht} - g_t^*(c_h)) \Longrightarrow N(0, \sigma_t^2(c_h)/p(c_h)\sqrt{2}),$$

and the estimates are asymptotically independent across h by Bierens *ibid*, Theorem 10.2.7. The independence across time is a consequence of our assumption that the error terms are uncorrelated. ■

Proof of Lemma 2. Consider two combinations c_h and $c_{h'}$ with j values 1 and 0 respectively and $c_{hj'} = c_{h'j'}$ for all $j' \neq j$. Using the definition of $g_t^*(\cdot)$ gives $g_t^*(c_h) - g_t^*(c_{h'}) = f_{jt}$. The final estimate of f_{jt} is the average of these differences across all m^{k-1} such h, h' pairs. The distribution limit of a fixed finite linear combination of sequences of random variables is the linear combination of the distribution limits. By Lemma 1 each sequence has a normal distribution limit and they are asymptotically independent. Using the formula for the variance of a linear combination of independent random variables gives (7). ■

Proof of Theorem 1. Let $Q_n(\beta)$ denote the sum of squared residual function which is to be minimized for least squares estimation:

$$Q_n(\beta) = \sum_{h=1}^{m^*} \sum_{t=1}^{T} \widehat{v}_t(c_h)(\widehat{y}_{ht} - y_{ht}(\beta))^2.$$

Note that given $\widehat{y}_{ht}$ and using the definition of $y_{ht}(\beta)$, $Q_n(\beta)$ is a multivariate polynomial in β. Also note that $Q_n(\beta)$ is a sum of squared terms times some positive weights and therefore is nonnegative everywhere. Hence it has a well-defined minimum (which need not be unique). Since $Q_n(\beta)$ is a multivariate polynomial it has derivatives to every order, and so when evaluated at any minimum the first-order condition must hold. The local uniqueness of the minimizers follows from the fact, discussed below, that the variables δ are not collinear, and are of dimensions less than or equal to the number of observations.

Now we show that $\widehat{\beta} \to^p \beta_0$. For any n let c^n denote $Q_n(\beta_0)$, that is, the sum of squared residual function evaluated at the true parameter vector β. Since $Q_n(\beta)$ is nonnegative and has a minimum at $\widehat{\beta}$ we have $0 \leq Q_n(\widehat{\beta}) \leq c^n$. Note that $c^n \to^p 0$ as $n \to \infty$, by virtue of the consistency of the kernel estimator at each point, and therefore $Q_n(\widehat{\beta}) \to^p 0$. We must show that this implies $\widehat{\beta} \to^p \beta$. Recall the definition of the target characteristic vectors c_h and consider the h' such that $c_h = 0^k$. For each t consider the term in $Q_n(\widehat{\beta})$ associated with h', and note that $0 \leq \widehat{v}_t(c_{h'})(y_{h't} - \widehat{y}_{h't})^2 \leq Q_n(\widehat{\beta})$ with probability tending to one, because $\widehat{v}_t(c_{h'})$ has a positive probability limit, and therefore $(y_{h't} - \widehat{y}_{h't})^2 \to^p 0$. Using the definitions of $\widehat{y}_{h't}$ and $y_{h't}$ gives $(\widehat{f}_{zt} - f_{zt} - \widehat{u}_{h't})^2 \to^p 0$, and since $\widehat{u}_{h't} \to^p 0$ this implies $\widehat{f}_{zt} \to^p f_{zt}$. Next consider h' associated with the target characteristic vector such that $c_{h'j} = 1$ and $c_{h'j'} = 0$ for all $j' \neq j$. By the same argument as in the last paragraph we have $(\widehat{y}_{h't} - y_{h't})^2 \to^p 0$. Using the definitions of $\widehat{y}_{h't}$ and $y_{h't}$ gives $(\widehat{f}_{zt} + \widehat{f}_{jt} - f_{zt} - f_{jt} - \widehat{u}_{h't})^2 \to^p 0$, and since $\widehat{u}_{h't} \to^p 0$ and $\widehat{f}_{zt} \to^p f_{zt}$ this implies $\widehat{f}_{jt} \to^p f_{jt}$. Last, we show that $\widehat{g}_{ej} \to^p g_{ej}$ for $e = 3, \ldots, m, j = 1, \ldots, k$. Consider h' associated with the target characteristic vector such that $c_{h'j} = g_{ej}$ and $c_{h'j'} = 0$ for all $j' \neq j$. By the same argument as in the last paragraph we have $(\widehat{f}_{zt} + \widehat{g}_{ej}\widehat{f}_{jt} - f_{zt} - g_{ej}f_{jt} - \widehat{u}_{h't})^2 \to^p 0$. By assumption there is at least one t such that $f_{jt} \neq 0$ and using this t we have $(\widehat{f}_{zt} + \widehat{g}_{ej}\widehat{f}_{jt} - f_{zt} - g_{ej}f_{jt} - \widehat{u}_{h't})^2 \to^p 0$ implies $\widehat{g}_{ej} \to^p g_{ej}$.
■

Proof of Theorem 2. Let $\widehat{y}$ and $y(\beta)$ denote the m^*T-vectors defined by y_{ht} and $y_{ht}(\beta)$ over all h and t. Rewriting $Q_n(\beta)$ in matrix form and taking the derivative with respect to β^*, evaluated at $\widehat{\beta}$

$$\begin{aligned} \frac{\partial}{\partial \beta} Q_n(\widehat{\beta}) &= \frac{\partial}{\partial \beta} \left(\widehat{y} - y(\widehat{\beta})\right)' \widehat{V} \left(\widehat{y} - y(\widehat{\beta})\right) \\ &= -2 \left(\widehat{y} - y(\widehat{\beta})\right)' \widehat{V}\Gamma(\widehat{\beta}). \end{aligned} \tag{13}$$

Note that this vector of derivatives equals the zero vector by (12) as proven in Theorem 1. Consider a first-order Mean Value expansion of $y(\widehat{\beta})$ around β_0

$$y(\widehat{\beta}) = y(\beta_0) + \Gamma(\widetilde{\beta})(\widehat{\beta} - \beta_0), \tag{14}$$

where $\widetilde{\beta}$ lies between $\widehat{\beta}$ and β_0. The appropriate value of $\widetilde{\beta}$ may differ for each element of $\widehat{\beta}$ (see Davidson and Mackinnon (1993) p. 154). Note that $\widehat{y} - y(\beta_0) = \widehat{u}$, where $\widehat{u}$ is the vector with typical element $\widehat{u}_{ht}$. Inserting (14) into (13), setting it equal to zero, then cancelling and rearranging terms, gives

$$(\widehat{\beta} - \beta_0)'\Gamma(\widetilde{\beta})'\widehat{V}\Gamma(\widehat{\beta}) - \widehat{u}'V\Gamma(\widehat{\beta}) = 0.$$

Because $\Gamma(\beta)$ is a fixed continuous function and $\widetilde{\beta} \to^p \beta_0$ and $\widehat{V} \to^p V$, we obtain

$$(nb_n^k)^{1/2}(\widehat{\beta} - \beta_0)H - (nb_n^k)^{1/2}\widehat{u}'V\Gamma(\beta_0) = o_p(1).$$

By Lemma 1, $(nb_n^k)^{1/2}\widehat{u}$ is asymptotically normal with a zero mean vector and covariance matrix D. If the difference in the plim of two random variables is zero then their dlims are the same (White (1984), Lemma 4.7, p. 63). Using that V, H are invertible completes the proof. ■

Bilinear Regression

We can rewrite the model as

$$y_\ell = d_\ell'\alpha + \gamma' x_\ell \delta + \varepsilon_\ell,$$

where the observations ℓ run from $\ell = 1, \ldots, m^*T$, and the matrix x_ℓ is of dimensions $mk \times kT$, while d_ℓ is $T \times 1$ vector of time dummies. The parameters γ satisfy the zero/one restrictions, which can be represented by $\gamma = A\eta + s$, where A, s are matrices consisting of zeros and ones. The free parameters η are of dimension $(m-2)k \times 1$. We now have a linear model

$$y_\ell = c_\ell'\theta + \varepsilon_\ell \tag{15}$$

where $c_\ell' = (d_\ell', a_\ell', b_\ell')$, and $\theta = (\alpha', \delta', \phi')'$, where $a_\ell' = s'x_\ell$ and $b_\ell = vec(A'x_\ell)$, while $\phi = \eta \quad \delta$. A sufficient condition for the parameters θ to be uniquely determined is that the matrix $\sum_{\ell=1}^{m^*T} c_\ell c_\ell'$ be of full rank. The regressors are all zeros and one, so the only question is about the dimensionality of the parameter vector relative to the number of observations, that is, we requires $m^*T > T + mk^2T$ or $m^* > mk^2 + 1$. This is likely to be satisfied in practice. In any case, this condition is not necessary, it refers to the overparameterized model that does not impose the restrictions on θ.

We now discuss our estimation algorithm, which exploits the bilinear structure of the problem. Estimation of the linear regression (15) is not practical, because of the very large dimensions involved. Instead we work with the bilinear regression where the restrictions have been imposed

$$y_\ell = d_\ell'\alpha + s'x_\ell\delta + \eta' A' x_\ell \delta + \varepsilon_\ell.$$

If we linearize this regression about the true values, we get that the regression function is approximately

$$y_\ell = d_\ell'\alpha + s'x_\ell\delta + \eta_0'A'x_\ell\delta + \eta'A'x_\ell\delta_0 + \varepsilon_\ell,$$

which suggests the following algorithm. We begin with starting values for the parameters α, δ, η, which we denote by $\alpha^{[0]}, \delta^{[0]}, \eta^{[0]}$. We then regress y_ℓ on $d_\ell, s'x_\ell + \eta^{[0]\prime}A'x_\ell$, and $A'x_\ell\delta^{[0]}$ to give us $\alpha^{[1]}, \delta^{[1]}, \eta^{[1]}$ respectively. This process is to be repeated until convergence. It can be seen that this is the so-called Gauss-Newton regression discussed in Davidson and MacKinnon (1993), and should converge to a zero of the first order condition. The weighting can easily be accomodated.

References

[1] Bansal, Ravi, David A. Hsieh and S. Viswanathan, 1993, A new approach to international arbitrage pricing, Journal of Finance, 48, 1719-1747.

[2] Banz, Rolf W., 1981, The relationship between return and market value of common stocks, Journal of Financial Economics, 9, 3-18.

[3] Basu, S., 1977, The investment performance of common stocks in relation to their price to earnings ratio: a test of the efficient markets hypothesis, Journal of Finance, 32, 663-682.

[4] Bierens, Herman J., 1994, Topics in Advanced Econometrics, Cambridge University Press, Cambridge, UK.

[5] Brennan, Michael J., Tarun Chordia and A. Subrahmanyan, 1998, "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns," Journal of Financial Economics, 49, 345-373.

[6] Capaul, Carlo, Ian Rowley and William F. Sharpe, 1993, International value and growth stock returns, Financial Analysts Journal, January-February, 27-36.

[7] Chan, Louis K.C., Jason Karceski and Josef Lakonishok, 1998, The risk and return from factors, Journal of Financial and Quantitative Analysis, 33, 159-188.

[8] Connor, Gregory and Robert Korajczyk, 1988, Risk and return in an equilibrium APT: Application of a new test methodology, Journal of Financial Economics 21, 255-289.

[9] Chapman, David A., 1997, Approximating the asset pricing kernel, Journal of Finance, 52, 1383-1410.

[10] Daniel, Kent, Mark Grinblatt and Sheridan Titman, Measuring Mutual Fund Performance with Characteristic-based Benchmarks, Journal of Finance 52, July 1997.

[11] Daniel, Kent and Sheridan Titman, 1997, Evidence on the characteristics of cross-sectional variation in stock returns, Journal of Finance 52, 1-34.

[12] Davis, James, 1994, The cross-section of realized stock returns: the pre-Compustat evidence, Journal of Finance, 49, 1579-1593.

[13] Davidson, Russell and James G. Mackinnon, 1993, Estimation and Inference in Econometrics, Oxford University Press, New York.

[14] Fama, Eugene F. and Kenneth R. French, 1992, The cross-section of expected stock returns, Journal of Finance 47, 427-465.

[15] Fama, Eugene F. and Kenneth R. French, 1993, Common risk factors in the returns to stocks and bonds, Journal of Financial Economics 33, 3-56.

[16] Fama, Eugene F. and Kenneth R. French, 1995, Size and book to market factors in earnings and returns, Journal of Finance 50, 131-156.

[17] Fama, Eugene F. and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, Journal of Finance 51, 55-84.

[18] Fama, Eugene F. and Kenneth R. French, 1998, Value versus Growth: The International Evidence, Journal of Finance 53, 1975-2000.

[19] Fama, Eugene F. and James MacBeth, 1973, Risk, return and equilibrium: empirical tests, Journal of Political Economy 71, 607-636.

[20] Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, Econometrica, 57, 1121-1152.

[21] Haugen, Robert, 1995, The New Finance: The Case against Efficient Markets, Prentice-Hall, Englewood Cliffs, New Jersey.

[22] Hansen, Lars P. and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, Journal of Finance, 52, 557-590.

[23] Härdle, Wolfgang, 1990, Applied Nonparametric Regression, Econometric Society Monographs no. 19, Cambridge University Press, Cambridge, United Kingdom.

[24] Jagannathan, Ravi and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, Journal of Finance, 51, 3-53.

[25] Knez, Peter J. and Mark J. Ready 1997, On the Robustness of Size and Book-to-Market in Cross-Sectional Regressions, Journal of Finance 52, 1355-1382.

[26] Lakonishok, Josef, Andrei Shleifer and Robert W. Vishny, 1994, Contrarian investment, extrapolation and risk, Journal of Finance, 49, 1541-1578.

[27] Lewellen, Jonathan, 1999, The time-series relations among expected return, risk, and book to market value, Journal of Financial Economics, 54, 5-44.

[28] Litzenberger, Robert and Krishna Ramaswamy, 1979, The Effect of Personal Taxes and Dividends on Capital Asset Prices: Theory and Evidence," Journal of Financial Economics, 7, 163-196.

[29] MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the CAPM, Journal of Finance, 38, 3-23.

[30] Mammen, E., O. Linton, and J.P. Nielsen, 1999. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. The Annals of Statistics, October.

[31] Pagan, A.R., and A. Ullah (1999). Nonparametric Econometrics. Cambridge University Press: Cambridge.

[32] Rosenberg, Barr, Kenneth Reid and Ron Lanstein, 1985, Persuasive evidence of market inefficiency, Journal of Portfolio Management, 11, 9-17.

[33] Rosenberg, Barr, 1974, Extra-market components of covariance among security prices, Journal of Financial and Quantitative Analysis, 9, 263-274.

[34] Rosenberg, Barr and Vinay Marathe, 1976, Prediction of investment risk: systematic and residual risk, Proceedings of the Seminar on the Analysis of Security Prices, University of Chicago, 85-224.

[35] Rothenberg, T.J. (1973). Efficient estimation with a priori information. Cowles Foundation Monograph.

[36] Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall: New York.