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

Gregory Connor and Oliver Linton[*]

Abstract

We introduce an alternative version of the Fama-French three-factor model of stock returns together with a new estimation methodology. We assume that the factor betas in the model are smooth nonlinear functions of observed security characteristics. We develop an estimation procedure that combines nonparametric kernel methods for constructing mimicking portfolios with parametric nonlinear regression to estimate factor returns and factor betas simultaneously. The methodology is applied to US common stocks and the empirical findings compared to those of Fama and French.

JEL codes: G12, C14.

Keywords: characteristic-based factor model, arbitrage pricing theory, kernel estimation, nonparametric estimation.

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 have been large return premia associated with size and value. Size is defined as 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 of the factor 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 factor model estimation methodology. Hence the accuracy and reliability of FF's estimation procedure is a critical issue in this research controversy.

FF (1993) use a simple approach to estimate their factor model. They sort securities according to the securities' 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 book-to-price and low book–to-price fractile portfolios, adjusted for the size characteristic, serves as an estimate of the value factor. They use a capitalization-weighted index as a proxy for the market factor. Although this method is intuitively plausible and computationally simple, there is to our knowledge no rigorous statistical theory to justify it with regard even to consistency. Furthermore, there is no obvious way to generate consistent standard errors for these and subsequent estimates that takes correct account of all sampling error. 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.

In this paper we develop an alternative methodology to describe the same phenomenon as do FF. We introduce a semiparametric characteristic-based factor model in which the factor betas are smooth functions of a small number of characteristics. The model can be viewed as a semiparametric

generalization of Rosenberg (1974, Section 3), where a linear such model is considered. The flexible nonlinearity we allow is important to capture the sort of generality implicit in the FF approach and evident in the data. 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 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. We also give an interpretation of our method as a variant of FF's portfolio construction approach.

The model is applied to US equities using the book-to-price ratio and the market value of equity as characteristics and the results are compared to those of FF. Our results are qualitatively similar to those of FF but with some improvements in model fit. For both characteristics we find that the relationship between the characteristic and associated factor beta is monotonic but not linear.

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. Proofs are given in the appendix.

2 Methodology

2.1 Description of the Factor Model

We assume that there is a 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:

$$r_{it} = f_{ut} + \sum_{j=1}^{J} g_j(C_{ij}) f_{jt} + \varepsilon_{it}, \tag{1}$$

where r_{it} is the return to security i at time t; f_{ut}, f_{jt} are the factor returns; $g_j(C_{ij})$ the factor betas, C_{ij} the security characteristics, which are assumed for simplicity not to vary over time, and ε_{it} are the mean zero asset-specific returns whose properties we discuss further below. The factor returns f_{jt} are linked to the security characteristics by the characteristic-beta functions $g_j(\cdot)$, which map characteristics to the associated factor betas. We assume that each $g_j(\cdot)$ is a smooth time-invariant function of characteristic j, but we do not assume a particular functional form. This a special

case of the unrestricted factor model [Connor and Koracyzk (1993)] $r_{it} = \sum_{j=1}^{J} \beta_{ij} f_{jt} + \varepsilon_{it}$, where β_{ij} are factor loadings, and generalizes the linear model considered in Rosenberg (1974, section 3) where $\beta_{ij} = \sum_k \delta_{jk} C_{ik} + u_{ij}$. We also note that (1) constitutes a weighted additive nonparametric regression model for panel data, where the factors f_{jt} are 'parametric weights' and the functions $g_j(\cdot)$ are univariate nonparametric functions. Some discussion of additive nonparametric models can be found in Linton and Nielsen (1995).

The market factor f_{ut} captures that part of common return not related to the security characteristics; all assets have unit beta to this factor. This factor captures the tendency of all equities to move together, irrespective of their characteristics. It is a common element in panel data models, see Hsiao (2003, section 3.6.2)

There are two indeterminacies in the characteristic-beta functions $g_j(\cdot)$, reflecting the usual rotational and scale indeterminacies of factor models. The first indeterminacy is additive. One can add an arbitrary constant a to any of the functions $g_j(\cdot)$ and subtract $a f_{jt}$ from f_u, 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}. 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_u 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.[1]

2.2   Kernel-based Portfolio Weights for Factor-Mimicking Portfolios

In this subsection we present a new technique for creating factor-mimicking portfolios, based on nonparametric kernel methods. Our purpose in developing this new technique is the estimation of our factor model, 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 new technique is founded on the earlier work of FF (1993) and we very briefly summarize

---

[1] An alternative normalization is to assume that $E[g_j(C_{ij})] = 0$ and $\mathrm{var}[g_j(C_{ij})] = 1$. This normalization has certain advantages from a statistical point of view, but is harder to interpret.

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 beta to the other characteristic, serves as the estimates of the factor returns associated with that characteristic. Specifically they define:

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

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

Our new technique can be viewed as 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 J 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. Let $c_{m,j}; m = 1, \ldots, M, j = 1, \ldots, J$ denote the chosen values, which are assumed to lie in the interior of the support of the random vector C. The grid of target characteristic vectors consists of all $H = M^J$ combinations of the M chosen target values over the J characteristics. Now collect all the target vectors together, and denote a typical member of this set by $c^h = (c_1^h, \ldots, c_J^h)^\top$, where $h = 1, \ldots, H$. Thus each c^h is a J-vector of target characteristics, where a given h corresponds

4

to a unique vector $(m_1, \ldots, m_J)$ and vice versa. Collect the observed characteristics for firm i into J-vectors $C_i = (C_{i1}, \ldots, C_{iJ})^\top$, $i = 1, \ldots, n$. One can also take a different number of target values for each characteristic but we avoid this extra complication for notational reasons.

Let ω_{hi} be 'localizing' weights, depending only on the data through C_i, that concentrate on values close to the vector c^h, and define the local weighted portfolio return as

$$\widehat{r}_{ht} = \widehat{r}_t(c^h) = \sum_{i=1}^{n} \omega_{hi} r_{it}. \tag{4}$$

From the perspective of finance, this can be viewed as the return on a well-diversified portfolio designed to have (approximately) the target characteristics c^h. From the perspective of statistical theory, $\widehat{r}_{ht}$ can be interpreted as a nonparametric estimator of the conditional expectation of r_{it} given $C_i = c^h$. To construct the weights ω_{hi} we use the local linear smoother approach [Fan and Gijbels (1996)]. This method is favoured because of its attractive statistical properties like good boundary behavior and less dependence on the covariate distribution. Let k be a (kernel) density function with finite second moment, and let $K(u_1, \ldots, u_J) = \prod_{j=1}^{J} k(u_j)$ be the product kernel; we take k to be the standard Gaussian density function. Then define the least squares criterion function

$$\sum_{i=1}^{n} \left[r_{it} - a_0 - a^\top (C_i - c^h) \right]^2 K((C_i - c^h)/b), \tag{5}$$

where $b = b(n)$ is a scalar bandwidth, while a_0 and $a = (a_1, \ldots, a_J)$ are local intercept and local slope parameters. Let $\widehat{a}_0, \widehat{a}$ be the minimizing values, which are explicit linear functions of r_{it} of the form (4). We let $\widehat{r}_{ht} = \widehat{a}_0$, and the weights ω_{hi} in (4) are correspondingly defined. There is an explicit formula for these weights given in Fan and Gijbels (1996). They are similar in some respects to the weights for the standard kernel estimator: they sum to one, but they need not be all positive. In practice however most weights are positive for reasonable sample sizes and the magnitude of negative weights when they do arise is small. One could avoid negative weights altogether by fitting instead a local constant procedure.

In our empirical application we vary bandwidth with the location c^h and time period t, typically enlarging bandwidths out in the tails where there is less data. For simplicity, we ignore this in the theoretical derivation and treat the bandwidth as fixed over c^h. It would also be possible to have a multivariate bandwidth that differs across the characteristics.

Now we show that the kernel-based portfolio returns converge to linear combinations of factor returns, with asymptotically normal and independent residuals. To do this, we apply a result from kernel regression theory, see Masry (1996). For each t define the function $r_t(c) = f_{ut} + \sum_{j=1}^{J} g_j(c_j) f_{jt}$ for any J-vector $c = (c_1, \ldots, c_J)^\top$. Using (1) it follows immediately that

$$r_{it} = r_t(C_i) + \varepsilon_{it}. \tag{6}$$

For a given t, equation (6) can be viewed as a multivariate nonparametric regression problem. Our kernel-based portfolio return for characteristic combination h is the local linear estimate of $r_t(c^h)$.

In order to describe the statistical properties of $\widehat{r}_{ht}$ we make some assumptions about the data generating process, although it should be noted that we do not need a full specification. We only rely on large cross-section asymptotics, and so do not need to fully specify the time series dependence. We assume that the observed characteristic J-vectors of the assets $C_i, i = 1, \ldots, n$ are independent and identically distributed across i. Let $p(c)$ denote the marginal density function of C_i evaluated at the point c, and let $\mathcal{C}$ denote the support of C_i. We further suppose that

ASSUMPTION A. *The vector $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})^\top$ is independently distributed across $i = 1, \ldots, n$, and satisfies $E(\varepsilon_i \varepsilon_i^\top | C_i = c) = \mathrm{Diag}\{\sigma_1^2(c), \ldots, \sigma_T^2(c)\}$ with probability one, where each function $\sigma_t^2(\cdot)$ is continuous at all points $c^h \in \mathcal{C}$. Furthermore, for some $\delta > 0$, $E[|\varepsilon_{it}|^{2+\delta}] < \infty$ for all t. The regression functions $r_t(\cdot)$ are twice continuously differentiable at all points $c^h \in \mathcal{C}$, while the density function p is continuous and strictly positive at each $c^h \in \mathcal{C}$. The bandwidth satisfies $b = \lambda n^{-1/(J+4)}$ for some λ with $0 < \lambda < \infty$.*

Define for each t, c^h,

$$\sigma_{ht}^2 = ||K||_2^2 \frac{\sigma_t^2(c^h)}{p(c^h)} \quad ; \quad \tau_t(c^h) = \frac{1}{2}\mu_2(k) \sum_{j=1}^J \frac{\partial^2 r_t}{\partial c_j^2}(c^h), \tag{7}$$

where $||K||_2^2 = \int K(u)^2 du$ and $\mu_2(k) = \int k(t)t^2 dt$. Then let $\widehat{\underline{r}}, \underline{r}$ be the stacked $TH \times 1$ vectors containing all the $\widehat{r}_{ht}$ and $r(c^h)$, and let $\underline{\tau}$ be the vector containing all the $\tau_t(c^h)$ in the same order.

LEMMA 1. *Suppose that Assumption A holds. Then as $n \to \infty$,*

$$(nb^J)^{1/2}(\widehat{\underline{r}} - \underline{r} - b^2 \underline{\tau}) \Longrightarrow N(0, \Omega), \text{ where } \Omega = \mathrm{diag}\{\sigma_{ht}^2\} \in \mathbb{R}^{TH \times TH}.$$

The central limit theorem for $\widehat{r}_t(c^h)$ is coming from the cross-sectional independence of the error terms; this assumption is sufficient but not necessary. Indeed in Connor and Korajczyk (1993) a weaker type of cross-sectional dependence, i.e., a mixing condition, was allowed. Lemma 1 carries over to this case provided the cross-sectional mixing coefficients decline fast enough. To obtain the joint asymptotic distribution over all time points we have assumed that ε_{it} is a martingale difference sequence with respect to time, so uncorrelated over time, but we do not rule out other sorts of temporal dependence in ε_{it} like GARCH effects. This assumption is consistent with the usual efficient market assumptions, and seems like a reasonable assumption to make in this context.[2]

---

[2]In the presence of autocorrelation in ε_{it}, the estimates are still asymptotically normal but the asymptotic variance matrix is no longer diagonal.

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 H combinations of characteristics, there are M^{J-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^{J-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,

$$\widetilde{f}_{jt} = \frac{1}{M^{J-1}} \sum_{h=1}^{H} (\delta_{hj,1} - \delta_{hj,0}) \widehat{r}_{ht}, \tag{8}$$

where the dummy variable $\delta_{hj,1}(\delta_{hj,0})$ equals one if mimicking portfolio h has target characteristic one (zero) for factor j and equals zero otherwise. The factor return estimates are consistent and asymptotically normal, as described in Lemma 2. Define for $j = 1, \ldots, J$ and $t = 1, \ldots, T$:

$$\tau_{fjt} = \frac{1}{M^{J-1}} \sum_{h=1}^{H} (\delta_{hj,1} - \delta_{hj,0}) \tau_t(c^h) \quad ; \quad \sigma_{jt}^2 = ||K||_2^2 \frac{1}{M^{J-1}} \sum_{h=1}^{H} (\delta_{hj1} - \delta_{hj0})^2 \frac{\sigma_t^2(c^h)}{p(c^h)}.$$

LEMMA 2. *As $n \to \infty$, for $j = 1, \ldots, J$ and $t = 1, \ldots, T$:*

$$(nb^J)^{1/2} (\widetilde{f}_{jt} - f_{jt} - b^2 \tau_{fjt}) \Longrightarrow N(0, \sigma_{jt}^2). \tag{9}$$

The joint asymptotic distribution of the estimated factors can also be obtained; typically there is an asymptotic covariance between $\widetilde{f}_{jt}$ and $\widetilde{f}_{ks}$. The FF-type estimates described in Lemma 2 have two weaknesses. First, the estimate of each factor return uses information from only a subset of the kernel-based portfolios and are inefficient as we show below. Second, the estimator only gives estimates of the factor returns, not the factor betas. In the next section we present an alternative estimator that 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. The corresponding estimator $\widehat{f}_{jt}$ is more efficient than $\widetilde{f}_{jt}$.

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

In this subsection we propose an alternative estimate of the factor returns to (8), and provide an estimate of the factor betas. We use the kernel-based portfolio returns described in the last

subsection as dependent 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. The estimator of the factor returns is guaranteed to be more efficient than (8). The estimation method is a version of minimum distance discussed in Rothenberg (1974); we apply these ideas from parametric estimation to our semiparametric problem.

Recall from the last subsection the definition of the kernel portfolios covering all combinations of the M target characteristics for each of the J factors over all T time periods. The returns on all of these kernel-based portfolios can be written as a pooled regression with H 'cross-sectional' observations (in this case 'cross-sectional' means across the kernel portfolios not across individual assets) and T time series observations:

$$\widehat{r}_{ht} = f_{ut} + \sum_{j=1}^{J} g_j(c_j^h) f_{jt} + \widehat{u}_{ht} \tag{10}$$

$$\widehat{u}_{ht} = \sum_{j=1}^{J} \{\sum_{i=1}^{n} \omega_{hi} g_j(C_{ij}) - g_j(c_j^h)\} f_{jt} + \sum_{i=1}^{n} \omega_{hi} \varepsilon_{it}. \tag{11}$$

Note that in (10) the nonparametric functions $g_j(\cdot)$ are each evaluated at M points, corresponding to each target point c^h. 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 $(M-2)J$ components of $\{g_{mj}\}$ as parameters to estimate, along with the $(J+1)T$ factor returns $\{f_{ut}, f_{jt}\}$. Let θ denote the $q = (M-2)J + (J+1)T$-vector of free parameters arranged in some consistent order, and let θ_0 be the true vector. We rewrite (10) as a nonlinear regression equation

$$\widehat{r}_{ht} = f_{ut} + \sum_{j=1}^{J} \sum_{m=1}^{M} g_{mj} \delta_{hj,m} f_{jt} + \widehat{u}_{ht}, \tag{12}$$

where $\delta_{hj,m}$ is a dummy variable equaling one if mimicking portfolio h has target value c_{mj} for factor j, and zero otherwise. Viewed as a regression equation, there are a fixed finite number of observations HT and q unknowns, where we assume that $q < HT$; the error terms in the regression are asymptotically independent across h, and are individually of small order in probability.

For a chosen parameter vector θ, define the predicted values $r_{ht}(\theta) = f_{ut} + \sum_{j=1}^{J} \sum_{m=1}^{M} g_{mj} \delta_{hj,m} f_{jt}$, and let $\underline{r}(\theta)$ be the $HT \times 1$ vectors containing the observations $r_{ht}(\theta)$. Then define $\widehat{\theta}$, as any minimizer of the minimum distance criterion

$$Q_n(\theta) = (\widehat{\underline{r}} - \underline{r}(\theta))^\top \widehat{V} (\widehat{\underline{r}} - \underline{r}(\theta)) \tag{13}$$

8

over $\theta \in \mathbb{R}^q$. The weighting matrix $\widehat{V}$ is a symmetric and positive definite $HT \times HT$ matrix, for example $\widehat{V} = I_{HT}$. The weighting is included to take account of error heteroscedasticity; it is allowed to be estimated from the data. The criterion function $Q_n(\theta)$ is a quartic polynomial in the parameters, and under reasonable conditions will have a global minimum, which will be unique on a suitably chosen compact set, which we denote by Θ. This enables us to use an iterative weighted least squares procedure to find the minimum. The actual algorithm we use exploits the bilinear structure of the regression function (12) and is described in the appendix.[3]

We next show the statistical properties of the estimator $\widehat{\theta}$. Define the $HT \times q$ and $q \times q$ matrices

$$\Gamma(\theta) = \frac{\partial \underline{r}(\theta)}{\partial \theta} \quad ; \quad \Psi(\theta) = \Gamma(\theta)^\top V \Gamma(\theta), \tag{14}$$

and let $\Psi_0 = \Psi(\theta_0)$ and $\Gamma_0 = \Gamma(\theta_0)$. Now we show that the least squares estimator is consistent and asymptotically normal.

THEOREM 1. *Suppose that the weighting matrix $\widehat{V} \to^p V$ as $n \to \infty$, where V is a symmetric positive definite matrix. Then, the least squares estimate defined by (13) exists with probability tending to one and $\widehat{\theta} \to^p \theta_0$. Suppose that Ψ_0 is a nonsingular matrix and that θ_0 is an interior point of Θ. Then, as $n \to \infty$,*

$$(nb^J)^{1/2}(\widehat{\theta} - \theta_0 - b^2 \Psi_0^{-1} \Gamma_0^\top V \tau) \Longrightarrow N(0, \Psi_0^{-1} \Gamma_0^\top V \Omega V \Gamma_0 \Psi_0^{-1}) \equiv N(0, \Sigma).$$

REMARKS

1. The asymptotic covariance matrix Σ in Theorem 1 can be consistently estimated by

$$\widehat{\Sigma} = \widehat{\Psi}^{-1} \widehat{\Gamma} \widehat{V} \widehat{\Omega} \widehat{V} \widehat{\Gamma}^\top \widehat{\Psi}^{-1}, \tag{15}$$

where $\widehat{\Psi} = \Psi(\widehat{\theta})$ and $\widehat{\Gamma} = \Gamma(\widehat{\theta})$, while $\widehat{\Omega} = \mathrm{diag}\{\widehat{\sigma}_{ht}^2\}$ is an estimate of Ω, where

$$\widehat{\sigma}_{ht}^2 = ||K||_2^2 \frac{\widehat{\sigma}_t^2(c^h)}{\widehat{p}(c^h)}$$

with $\widehat{p}(c^h) = n^{-1} b^{-J} \sum_{i=1}^n K((C_i - c^h)/b)$ and $\hat{\sigma}_t^2(c^h) = \sum_{i=1}^n \omega_{hi} r_{it}^2 - \left(\sum_{i=1}^n \omega_{hi} r_{it}\right)^2$. Standard errors for the factors and the betas are then obtained from the square root of the corresponding diagonal element of $\widehat{\Sigma}/nb^J$. The matrix $\widehat{\Psi}$ can be quite large - in our application it is 1422×1422 - and so computing $\widehat{\Psi}^{-1}$ can be time consuming and subject to numerical rounding error. In the appendix we discuss how to compute the inverse $\widehat{\Psi}^{-1}$ exploiting the sparsity structure in the Ψ matrix, thereby avoiding the direct inversion of a very large matrix.

---

[3]We may wish to use only subperiod or even single time period information to estimate θ. In the single period case we would minimize a criterion $(\widehat{r}_t - r_t(\theta))^\top \widehat{V}_t (\widehat{r}_t - r_t(\theta))$ with respect to θ; of course, the degree of overidentification reduces (and hence efficiency worsens) but on the other hand this approach is more robust to time series issues like structural change etc.

2. When $V = \Omega^{-1}$, we have

$$\Sigma = \left(\Gamma_0 \Omega^{-1} \Gamma_0^\top\right)^{-1}. \tag{16}$$

The asymptotic variance in (16) is minimal amongst this class of estimators. The class of estimators includes all those asymptotically linear combinations of the vector $\widehat{r}$ and so $\widetilde{f}_{jt}$ is included in this class of estimators as a very special case. It follows that $\widehat{f}_{jt}$ has a smaller asymptotic variance than $\widetilde{f}_{jt}$. The efficient estimator can be implemented in practice by taking $\widehat{V} = \widehat{\Omega}^{-1}$, where $\widehat{\Omega}$ is the estimator described above. Note that even in this case the matrix Σ is not diagonal, which says that estimation of the factors affects in variance terms estimation of the factor betas and vice versa.

3. We have assumed for the asymptotic normality that the matrix Ψ_0 is non-singular. In general it is difficult to provide primitive conditions to ensure that Ψ_0 is a nonsingular matrix. However, in the special case of homoskedastic errors a sufficient condition is that the vectors $g_1, \ldots, g_J$ are not collinear with themselves or with a vector of ones.

4. We have estimated all the unknown quantities at the rate $(nb^J)^{-1/2}$, which is the standard rate for J-dimensional nonparametric regression. However, the quantities f_{jt} can in principle be estimated at rate $n^{-1/2}$ since they are effectively parametric, and the quantities $g_j(.)$ can in principle be estimated at rate $(nb)^{-1/2}$ since their arguments are only one-dimensional, see Stone (1980) and Bickel, Klaassen, Ritov, and Wellner (1995). The slower rate we have is due to the fact that we have taken a grid set of cardinality H that does not increase with sample size n. The theory can be extended to allow $H = H(n) \to \infty$ and hence yield improvements in rate. We have not done this here because the dataset is so large and so: (a) we are limited in computational time as to how many grid points to average over, (b) the variance is in any case small.

3 Empirical Analysis

3.1 Data

Except for the addition of recent years, our data is essentially identical to that in FF (1993). The monthly returns data covers the period July 1963 to June 2002. To be included in the data set during a given year (July to June) a security must have a complete monthly return record during that year and a recorded book value of equity and market value of equity in the preceding June. All returns are measured in excess of the Treasury Bill rate, i.e., the monthly Treasury Bill rate is subtracted from each security's raw return. The size (log of market value) and value (log of the book to market ratio) of each security is fixed for the July-to-June period and comes from the preceding June. The security returns and equity market values come from the Center for Research In Security Prices monthly database; the equity book values are from Compustat.

Table 1 shows some descriptive statistics for the data: the number of securities in the annual cross-section, and the first four cross-sectional moments of the two characteristics. To save space the table only shows five representative years (years 1, 10, 20, 30 and 39 of the sample) and 39-year averages; the complete table of all individual years is available from the authors. The size characteristic is leptokurtic and slightly negatively skewed relative to the normal distribution, and the opposite holds for the value characteristic. There is fairly strong negative cross-sectional correlation between the two characteristics, large firms tending to have lower book-to-price ratios than small firms. The number of firms in the cross-section increases substantially over the 39 year time period.

3.2 Implementation

To begin estimation of the model we need to choose a set of target characteristics, a kernel function, and a bandwidth-setting procedure.

The choice of a set of target characteristics is analogous to FF's choice of a set of sort portfolios. FF use three different sets of sort portfolios: for factor estimation 3×2=6 portfolios, and for test assets, either 5×5=25 or 10×10=100 portfolios.

For both the size and value characteristics we use target values in the range –2.00 to +3.00 inclusive, spaced at intervals of 0.5, giving eleven target values for each of the two characteristics and therefore 11×11=121 combinations of the two. The asymmetric range of -2.00 to 3.00 was chosen to reflect the importance of very large capitalization stocks and (to a lesser extent) high "value" stocks in the Fama-French theory. FF (1992, 1993) also use asymmetric rules in the construction of their sort portfolios, for the same reason. The grid space between target points needs to be narrow enough to give a rich set of characteristic targets yet wide enough so that there is not excessive overlap between the target portfolios.

We chose a product Gaussian kernel throughout. The advantage of this kernel is that it is very smooth and produces nice regular estimates, whereas, say the Epanechnikov kernel produces estimates with discontinuities in the second derivatives. The product kernel is satisfactory provided the bandwidths are scaled to the units of the different covariates, as they are. 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). A wider bandwidth gives a more diversified portfolio. A narrower bandwidth minimizes the overlap between nearby portfolios, and ensures that the characteristics of each portfolio closely match their target value.

After experimenting with a variety of bandwidth setting methodologies, we decided that a simple rule-of-thumb procedure like Silverman (1986) worked best. For each target vector in each year, we

calculated the sample density of the root-mean-squared differences between all the sample characteristic vectors and the target vector. For each target vector in each year we set the bandwidth equal to the fifth percentile of this sample density. This implies that ninety-five percent of the observations are at least one bandwidth away from the target vector, where distance is measured by root-mean-square. This simple procedure guarantees that the bandwidth is narrow where the data set is locally more densely populated (e.g., near the median values of the two characteristics) and wider where the data set is locally sparse (e.g., near the extreme values of the characteristics). It is rather like a smooth nearest neighbors bandwidth taking 5% of the data in each marginal window. The bandwidths range from 0.237 to 3.32 with a mean of 1.11. Figures 1 and 2 display the chosen bandwidths and relate them to each of the two characteristics.

3.3 The Characteristic-Beta Functions

Table 2 shows the estimates of the characteristic-beta functions at the specified target characteristic values, and standard errors for each estimate. Note that the standard errors of the beta estimates are corrected for the joint estimation error in the factor returns, unlike e.g., FF (1993). The standard errors tend to be larger in the tails, where the data is sparser. The characteristic-beta functions are displayed in Figures 3 and 4. Recall that both characteristic-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. The uniformly positive slope of the functions has implications for analysis of both the size effect and the value effect in equity markets. It implies that the 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.

The characteristic-beta function is relatively flat at the high end of the value characteristic, so the marginal increase in return premia is small over this region. FF(1993,1996) argue that the value factor is related to an economy-wide "financial distress" risk in capital market. Note however that we find that 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.

12

3.4 The Estimated Factors

In this subsection we analyze the estimated factors, and compare them to the factor portfolio returns from the original FF procedure. The FF factors are publicly provided (including updates for recent history) by Ken French.[4] In addition to the value-weighted market index used by FF, we also include the equally-weighted market index for comparison purposes. Table 3 shows the correlation matrix for all the factors. There is a very high positive correlation between the pairs of equivalent factors estimated by the two methods; these are highlighted using bold font. Our unit-beta factor has extremely high correlation with the equally-weighted market index and high, but not extremely high, correlation with the value-weighted market index. Brown (1989) shows analytically that the dominant statistical factor in a large asset market is approximately identical to the equally-weighted index return; Connor and Korajczyk (1988) show empirically that this near-equivalence holds for US equity returns with statistically-derived factors. Given these earlier findings, the extremely high correlation between our unit-beta factor and the equally-weighted index return is not surprising.

Note that our "size" factor has a negative correlation with the SMB factor since "size" in our model is a positive monotonic transformation of capitalization and therefore is defined oppositely from "Small Minus Big" as used by FF. This is merely a sign reversal and has no substantive effect.

An estimated factor return is a linear combination of the sample of asset returns and so it can be expressed as a vector of "portfolio weights," although these weights will not typically sum to one, and will differ each period. It is possible to compare the FF factors and our factors by examining the portfolio weights which underlie the estimated factors. Figures 5-8 compare the "portfolio weights" underlying our size and value factors and the analogous FF factors, for the middle month of the sample (November 1982). Figures 5 and 6 show the two "size" factor portfolios as functions of the size characteristic and Figures 7 and 8, the two "value" factor portfolios as functions of the value characteristic. Other functional representations (each factor portfolio as a function of the other characteristic, and the market and zero-beta portfolios as a function of each characteristic) are available from the authors. Note that our estimation methodology results in much more diversified portfolios than the FF method (in this regard it is important to take note of the differing scales in the figures). Due to the capitalization weighting, the FF portfolios are dominated by the relatively small number of high-capitalization securities.

The remaining analysis in this subsection is based on a simple time-series regression formulation: each time-series of returns in a panel of asset returns is regressed on an intercept and the time-series returns of three factors:

---

[4]See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ for the datasets and details on their construction.

$$r_{it} = \widehat{\alpha}_i + \widehat{\beta}_{i1} f_{1t} + \widehat{\beta}_{i2} f_{2t} + \widehat{\beta}_{i3} f_{3t} + \widehat{\varepsilon}_{it}, \tag{17}$$

whether f_1, f_2, f_3 are either our estimated factors or the three FF factors. For the panel of dependent variables r_{it} we consider individual securities, portfolios sorted by the characteristics, and industry portfolios. The performance of the factor model can be judged either by its ability to explain the time-series of asset returns (small values of $\widehat{\varepsilon}_{it}$), or its ability to explain the cross-section of mean returns ($\widehat{\alpha}_i \approx 0$). We will consider both of these criteria.

We use six sets of dependent variables in the analysis. The first set is the full collection of individual asset returns. The next two sets are 100 portfolios sorted by size and value, provided by Ken French. The first of these uses value-weighting and the second equal-weighting in the portfolio constructions.[5] The fourth and fifth are sets of 30 value-weighted and equally-weighted industry portfolios, again provided by Ken French. The last set is the 121 kernel portfolios which come from the first stage of our estimation procedure.

Table 4 shows average R-squared statistics and mean-square residuals from the time-series regressions (17) using the six sets of dependent variables. For the individual assets the time-series regressions are over the 12-month subperiods used to define the balanced panels of assets returns, and the "averages" are over both assets and years. For the remaining five sets of dependent variables the time-series regression are over the full 39-year period.

The factors estimated by our method outperform the Fama-French factors in terms of explanatory power for four of the six cases, the exceptions being the value-weighted sort portfolios and value-weighted industry portfolios. Using value-weighted portfolios on both sides of (17) induces an errors-in-variables bias, since the idiosyncratic return of the small number of very high-capitalization securities appears nonnegligibly in both the factor return estimates and in the asset returns. It is notable how much more well-diversified are the 121 kernel portfolios compared to the 100 Fama-French value-weighted and equally-weighted sort portfolios. This is demonstrated by the much high average $\overline{R}^2$ values when the kernel portfolios are regressed on the factor returns.

In the first panel of Table 5, we re-estimate (17) for individual securities after dropping the intercept and each factor separately. The difference between the adjusted R-squared statistic with and without a given factor is a simple descriptive measure of the marginal explanatory power of the factor. We show the average of these differences across all assets. The intercept has no explanatory power: due to the adjustment for degrees of freedom it actually lowers average $\overline{R}^2$ and the average residual variance. In both cases (our factors and the FF factors), each of the three factors has nonnegligible explanatory power, with the market factor by far the strongest, then the value factor

---

[5]See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ for details on the construction of these size and value sorted portfolios.

and last the size factor. We use a small sample t-test of the significance of each coefficient, and calculate the proportion significant at 95% confident.

In the next five panels we repeat this regression exercise for the other five sets of assets. We can reach no clear conclusions from the comparisons of the aggregated intercept tests: the estimation and interpretation of the intercepts in this type of factor-return regression is notoriously difficult. The ability to reject the hypothesis that the intercepts are zero in some cases partly reflects the very high power of these tests (note the very high $\overline{R}^2$ as shown in Table 4) rather than the magnitude of the estimated intercepts. On the other hand, we can state definitively that each of the three factors shows a pervasive influence on each set of asset returns, with the same ordering of relative influence as for individual assets: market, value, and size. This holds both for the FF factors and our new estimated factors.

4   Summary

This paper describes a characteristic-based factor model along the lines of the Fama and French (1993) three-factor model, and develops a new estimation methodology that is a mixture of parametric and nonparametric methods. 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 dependent variable. This new methodology allows for a range of approximate (asymptotic) statistical results not available with Fama and French's procedure.

The model is applied to essentially the same dataset as in Fama and French (1993) and the results are compared. In terms of explanatory power the factors estimated by our method and those from Fama and French perform comparably, with some evidence for marginal outperformance by our factors. The mimicking portfolios created by our procedure appear much better diversified than the bivariate size and value sort portfolios provided by Fama and French.

Unlike the original Fama and French model, our model gives explicit estimates of the relationship between security characteristics and the associated factor betas. We find that for both value and size these relationships are monotonic, but not linear.

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.

We have assumed that the characteristic-beta functions are constant through time; it would be interesting and worthwhile to extend the model to allow time-varying betas; both cyclically (possibly related to business cycle indicators) and in terms of secular trends.

A Appendix

A.1 Proofs

PROOF OF LEMMA 1. Following the arguments of Masry (1996), it can be shown that for each t, c^h,

$$\widehat{r}_t(c^h) - r_t(c^h) = \sum_{i=1}^{n} \widetilde{\omega}_{hi}\varepsilon_{it} + b^2\tau_t(c^h) + o_p(n^{-2/(J+4)}),$$

where $\widetilde{\omega}_{hi}$ are the weights

$$\widetilde{\omega}_{hi} = \frac{1}{nb^J}\frac{1}{p(c^h)}K\left(\frac{C_i - c^h}{b}\right). \tag{18}$$

It then follows that for each t, c^h, $(nb^J)^{1/2}\sum_{i=1}^{n}\widetilde{\omega}_{hi}\varepsilon_{it} \Rightarrow N(0, \|K\|_2^2\sigma_t^2(c^h)/p(c^h))$ by Lindeberg's central limit theorem. The estimates $\widehat{r}_t(c^h), \widehat{r}_t(c^{h'})$ are asymptotically independent for $c^h \neq c^{h'}$ because of the localizing property of $\widetilde{\omega}_{hi}$:

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}\widetilde{\omega}_{hi}\varepsilon_{it}\sum_{i=1}^{n}\widetilde{\omega}_{h'i}\varepsilon_{it}\right] &= E\left[\sum_{i=1}^{n}\widetilde{\omega}_{hi}\widetilde{\omega}_{h'i}\varepsilon_{it}^2\right] \\
&= \frac{1}{nb^{2J}}\frac{1}{p(c^h)p(c^{h'})}\int K\left(\frac{c - c^h}{b}\right)K\left(\frac{c - c^{h'}}{b}\right)p(c)dc \\
&= \frac{1}{nb^J}\frac{1}{p(c^h)p(c^{h'})}\int K(u)K\left(u + \frac{c^h - c^{h'}}{b}\right)p(c^h + ub)du \\
&= \frac{1}{nb^J}o(1),
\end{aligned}
$$

because: $\int K(u)K\left(u + \frac{c^h - c^{h'}}{b}\right)du \to 0$ as $n \to \infty$ for any $c^h \neq c^{h'}$ by dominated convergence, and $p(c)$ is bounded away from zero and bounded. The independence across time follows from the fact that ε_{it} are uncorrelated, since for $t \neq s$, $E[\sum_{i=1}^{n}\widetilde{\omega}_{hi}\varepsilon_{it}\sum_{i=1}^{n}\widetilde{\omega}_{hi}\varepsilon_{is}] = E[\sum_{i=1}^{n}\widetilde{\omega}_{hi}^2 E(\varepsilon_{it}\varepsilon_{is}|C_i)] = 0$ using the law of iterated expectation. Therefore we have for any vector $\alpha \in \mathbb{R}^J$, $(nb^J)^{1/2}\alpha^\top(\widehat{r} - r - b^2\tau) \Rightarrow N(0, \alpha^\top\Omega\alpha)$, which by Cramèr's theorem implies the result. ∎

PROOF OF LEMMA 2. Consider two combinations c^h and $c^{h'}$ with j values 1 and 0 respectively and $c_{j'}^h = c_{j'}^{h'}$ for all $j' \neq j$. Using the definition of $r_t(\cdot)$ gives $r_t(c^h) - r_t(c^{h'}) = f_{jt}$. The final estimate of f_{jt} is the average of these differences across all M^{J-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 (9). ∎

PROOF OF THEOREM 1. Note that given $\widehat{r}_{ht}$ and using the definition of $r_{ht}(\theta)$, $Q_n(\theta)$ is a multivariate polynomial in θ. Also note that $Q_n(\theta)$ 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 $Q_n(\theta)$ is a multivariate polynomial it has derivatives to every order, and so when evaluated at any minimum the first-order condition

$$\frac{\partial}{\partial \theta} Q_n(\widehat{\theta}) = 0. \tag{19}$$

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 $\widehat{\theta} \to^p \theta_0$. Since $Q_n(\theta)$ is nonnegative and has a minimum at $\widehat{\theta}$ we have $0 \leq Q_n(\widehat{\theta}) \leq Q_n(\theta_0)$. Note that $Q_n(\theta_0) \to^p 0$ as $n \to \infty$, by virtue of the consistency of the kernel estimator at each point, and therefore $Q_n(\widehat{\theta}) \to^p 0$. We must show that this implies $\widehat{\theta} \to^p \theta$. Recall the definition of the target characteristic vectors c^h and consider the h' such that $c^{h'} = 0^J$. For each t consider the term in $Q_n(\widehat{\theta})$ associated with h', and note that $0 \leq \widehat{v}_t(c^{h'})(r_{h't} - \widehat{r}_{h't})^2 \leq Q_n(\widehat{\theta})$ with probability tending to one, because $\widehat{v}_t(c^{h'})$ has a positive probability limit, and therefore $(r_{h't} - \widehat{r}_{h't})^2 \to^p 0$. Using the definitions of $\widehat{r}_{h't}$ and $r_{h't}$ gives $(\widehat{f}_{ut} - f_{ut} - \widehat{u}_{h't})^2 \to^p 0$, and since $\widehat{u}_{h't} \to^p 0$ this implies $\widehat{f}_{ut} \to^p f_{ut}$. Next consider h' associated with the target characteristic vector such that $c_j^{h'} = 1$ and $c_{j'}^{h'} = 0$ for all $j' \neq j$. Using that quadratic functions of probability limits converge we have $(\widehat{r}_{h't} - r_{h't})^2 \to^p 0$. Using the definitions of $\widehat{r}_{h't}$ and $r_{h't}$ gives $(\widehat{f}_{ut} + \widehat{f}_{jt} - f_{ut} - f_{jt} - \widehat{u}_{h't})^2 \to^p 0$, and since $\widehat{u}_{h't} \to^p 0$ and $\widehat{f}_{ut} \to^p f_{ut}$ this implies $\widehat{f}_{jt} \to^p f_{jt}$. Last, we show that $\widehat{r}_{hj} \to^p r_{hj}$ for $m = 3, \ldots, M, j = 1, \ldots, J$. Consider h' associated with the target characteristic vector such that $c_j^{h'} = r_{hj}$ and $c_{j'}^{h'} = 0$ for all $j' \neq j$. By the same argument as in the last paragraph we have $(\widehat{f}_{ut} + \widehat{r}_{hj}\widehat{f}_{jt} - f_{ut} - r_{hj}f_{jt} - \widehat{u}_{h't})^2 \to^p 0$. By assumption there is at least one t such that $f_{jt} \neq 0$ and using this t we have $(\widehat{f}_{ut} + \widehat{r}_{hj}\widehat{f}_{jt} - f_{ut} - r_{hj}f_{jt} - \widehat{u}_{h't})^2 \to^p 0$ implies $\widehat{r}_{hj} \to^p r_{hj}$.

Rewriting $Q_n(\theta)$ in matrix form and taking the derivative with respect to θ^*, evaluated at $\widehat{\theta}$

$$\begin{aligned} \frac{\partial}{\partial \theta} Q_n(\widehat{\theta}) &= \frac{\partial}{\partial \theta} \left(\widehat{\underline{r}} - \underline{r}(\widehat{\theta}) \right)^\top \widehat{V} \left(\widehat{\underline{r}} - \underline{r}(\widehat{\theta}) \right) \\ &= -2\Gamma(\widehat{\theta})\widehat{V} \left(\widehat{\underline{r}} - \underline{r}(\widehat{\theta}) \right). \end{aligned} \tag{20}$$

Note that this vector of derivatives equals the zero vector by (19) as proven above. Consider a first-order Mean Value expansion of $\underline{r}(\widehat{\theta})$ around θ_0

$$\underline{r}(\widehat{\theta}) = \underline{r}(\theta_0) + \Gamma^\top(\widetilde{\theta})(\widehat{\theta} - \theta_0), \tag{21}$$

17

where $\widetilde{\theta}$ lies between $\widehat{\theta}$ and θ_0. The appropriate value of $\widetilde{\theta}$ may differ for each element of $\widehat{\theta}$ (see Davidson and Mackinnon (1993) p. 154). Note that $\widehat{\underline{r}} - \underline{r}(\theta_0) = \widehat{u}$, where $\widehat{u}$ is the vector with typical element $\widehat{u}_{ht}$. Inserting (21) into (20), setting it equal to zero, then cancelling and rearranging terms, gives $\Gamma(\widetilde{\theta})^\top \widehat{V} \Gamma(\widehat{\theta})(\widehat{\theta} - \theta_0) - \Gamma(\widehat{\theta}) V \widehat{u} = 0$. Because $\Gamma(\theta)$ is a fixed continuous function and $\widetilde{\theta} \to^p \theta_0$ and $\widehat{V} \to^p V$, we obtain

$$\Psi_0 (nb^J)^{1/2} (\widehat{\theta} - \theta_0) - \Gamma_0 V (nb^J)^{1/2} \widehat{u} = o_p(1).$$

By Lemma 1, $(nb^J)^{1/2}(\widehat{u} - b^2 \underline{\tau})$ is asymptotically normal with zero mean vector and covariance matrix Ω. If the difference in the probability limit of two random variables is zero then their distributional limits are the same (White (1984), Lemma 4.7, p. 63). Using that Ψ_0 is invertible completes the proof. ∎

A.2 Estimation Algorithm

Here we describe the estimation algorithm we use to compute $\widehat{\theta} = (\widehat{\underline{g}}^\top, \widehat{f}^\top)^\top$, where $\widehat{f} = (\widehat{f_u}^\top, \widehat{f_1}^\top, \ldots, \widehat{f_J}^\top)^\top$ and $\widehat{\underline{g}} = (\widehat{\underline{g}}_1^\top, \ldots, \widehat{\underline{g}}_J^\top)^\top$ with $\widehat{f}_j, \widehat{\underline{g}}_j$ being $T \times 1$ and $(M-2) \times 1$ vectors respectively. It is an iterative weighted least squares procedure, a variant on partitioned regression. It is designed to exploit the bilinear structure and to thereby reduce computational time.

We first rewrite the estimating equations to give some insight into its algebraic structure. We introduce the quantities of interest: $f = (f_u^\top, f_1^\top, \ldots, f_J^\top)^\top$ and $g = (g_1^\top, \ldots, g_J^\top)^T$ with each f_j being $T \times 1$ and each g_j being $M \times 1$. Define the corresponding unrestricted elements of g by $\underline{g} = (\underline{g}_1^\top, \ldots, \underline{g}_J^\top)^\top$, where each $\underline{g}_j$ is an $(M-2) \times 1$ vector. This removes the zero and one components of g which are fixed for identification purposes and not estimated parameters. We can also represent the factor information as $f_* = (f^{1\top}, \ldots, f^{T\top})^\top$, where $f^t = (f_{ut}, f_{1t}, \ldots, f_{Jt})^\top$ are $(J+1) \times 1$ parameter vectors, so that f_* is just a rearrangement of f.

Suppose that the target values are arranged according to the following order $\{(c_{1,1}, \ldots, c_{1,J}), \ldots, (c_{M,1}, \ldots, c_{1,J}), (c_{1,1}, \ldots, c_{2,J}), \ldots\}$, i.e.,

$$\underline{r}_t(\theta) = \begin{bmatrix} f_{ut} + f_{1t} g_1(c_{1,1}) + f_{2t} g_2(c_{1,2}) + \cdots \\ \vdots \\ f_{ut} + f_{1t} g_1(c_{M,1}) + f_{2t} g_2(c_{1,2}) + \cdots \\ \vdots \end{bmatrix},$$

where $\underline{r}_t(\theta)$ is the $H \times 1$ containing the $r_{ht}(\theta)$ in consistent order. Define the $H \times 1$ vector $P_u = (\otimes^J i_M) = i_M \otimes \cdots \otimes i_M$ and the $H \times M$ matrices of zeros and ones:

$$P_1 = (\otimes^{J-1} i_M) \otimes I_M, \ P_2 = (\otimes^{J-2} i_M) \otimes I_M \otimes i_M, \ldots, P_J = I_M \otimes (\otimes^{J-1} i_M).$$

18

Then

$$\underline{r}(\theta) = f_u \otimes P_u + \sum_{j=1}^{J} f_j \otimes (P_j g_j),$$

where we stack the T vectors $\underline{r}_t(\theta)$ on top of each other to give $\underline{r}(\theta)$. Note that there are the identification restrictions fixing the first two values of each g_j; these can be written as

$$g_j = \delta \underline{g}_j + e_2,$$

where e_2 is an $M - 2 \times 1$ vector with one in its second position and zero else and $\delta = (0, I_{M-2})^\top$ with 0 representing a $(M - 2) \times 2$ vector of zeros.

Combining these equations we have the following conditional linear relationships:

$$
\begin{aligned}
\underline{r}(\theta) &= f_u \otimes P_u + \sum_{j=1}^{J} f_j \otimes (P_j g_j) \\
&= (I_T \otimes P_u) f_u + \sum_{j=1}^{J} (I_T \otimes (P_j g_j)) f_j = X_g f \qquad (22) \\
&= (f_u \otimes I_H) P_u + \sum_{j=1}^{J} (f_j \otimes I_H) P_j \delta \underline{g}_j + \sum_{j=1}^{J} (f_j \otimes I_H) P_j e_2 = X_f \underline{g} + c_f, \qquad (23)
\end{aligned}
$$

where $X_g = I_T \otimes (P_u, P_1 g_1, \ldots, P_J g_J)$ is $HT \times (J+1)T$, while $X_f = ((f_1 \otimes I_H) P_1 \delta, \ldots, (f_J \otimes I_H) P_J \delta)$ is $HT \times (M - 2)J$ and $c_f = (f_u \otimes I_H) P_u + \sum_{j=1}^{J} (f_j \otimes I_H) P_j e_2$ is $HT \times 1$. We exploit this structure in our estimation algorithm. This is:

1. Choose starting values for $f^{[0]}$. We use the consistent estimates described in Lemma 2.

2. Estimate $\underline{g}$ in (23) by weighted least squares using $\widehat{V}$, $X_{f^{[0]}} = ((f_1^{[0]} \otimes I_H) P_1 \delta, \ldots, (f_J^{[0]} \otimes I_H) P_J \delta)$,
$$\underline{g}^{[1]} = (X_{f^{[0]}}^\top \widehat{V} X_{f^{[0]}})^{-1} X_{f^{[0]}}^\top \widehat{V} (\widehat{\underline{r}} - c_{f^{[0]}}) \qquad (24)$$

3. Estimate f in (22) by weighted least squares using $\widehat{V}$, $X_{g^{[0]}} = I_T \otimes (P_u, P_1 g_1^{[0]}, \ldots, P_J g_J^{[0]})$, where $g_j^{[0]} = \delta \underline{g}_j^{[0]} + e_2$,
$$f^{[1]} = (X_{g^{[0]}}^\top \widehat{V} X_{g^{[0]}})^{-1} X_{g^{[0]}}^\top \widehat{V} \widehat{\underline{r}} \qquad (25)$$

4. Continue steps 2 and 3 until convergence criteria is met, e.g., until
$$||\theta^{[r+1]} - \theta^{[r]}|| < \epsilon$$
for some prespecified small $\epsilon > 0$. Call the final value $\widehat{\theta}$.

Note that correct standard errors for $\widehat{f}, \widehat{g}$ cannot be obtained from the above algorithm directly; in the next section we discuss a strategy for obtaining standard errors at minimal computational cost.

19

A.3 Asymptotic Variance and Standard Errors

Here we discuss the form of the asymptotic variance, with a view to computing standard errors. We must find the derivatives of $r(\theta)$ with respect to the components of θ and thence the quadratic forms Ψ_0 and Σ. We work with a rearrangement of θ, given by $\theta = (\underline{g}^\top, f_*^\top)^\top$, where $f_* = ((f^1)^\top, \ldots, (f^T)^\top)^\top$.

Define the generic $TH \times TH$ diagonal weighting matrix V. Then

$$\Psi_0 = \frac{\partial r}{\partial \theta^\top} V \frac{\partial r}{\partial \theta} = \begin{bmatrix} \frac{\partial r}{\partial \underline{g}^\top} V \frac{\partial r}{\partial \underline{g}} & \frac{\partial r}{\partial \underline{g}^\top} V \frac{\partial r}{\partial f_*} \\ \frac{\partial r}{\partial f_*^\top} V \frac{\partial r}{\partial \underline{g}} & \frac{\partial r}{\partial f_*^\top} V \frac{\partial r}{\partial f_*} \end{bmatrix} \equiv \begin{bmatrix} \Psi_{\underline{g}\underline{g}} & \Psi_{\underline{g}f_*} \\ \Psi_{f_*\underline{g}} & \Psi_{f_*f_*} \end{bmatrix}, \tag{26}$$

where $\Psi_{\underline{g}\underline{g}}$ is $(M-2)J \times (M-2)J$, $\Psi_{f_*f_*}$ is $(J+1)T \times (J+1)T$ and $\Psi_{\underline{g}f_*}, \Psi_{f_*\underline{g}}$ have consistent dimensions. The asymptotic variance depends on the inverse of this large matrix, which we now seek to find. In practice, $\Psi_{f_*f_*}$ has larger dimensions than $\Psi_{\underline{g}\underline{g}}$, but happily there is an analytical formula for $\Psi_{f_*f_*}^{-1}$, which we can exploit. We use the partitioned inverse formula

$$\Psi_0^{-1} = \begin{bmatrix} \left(\Psi_{\underline{g}\underline{g}} - \Psi_{\underline{g}f_*} \Psi_{f_*f_*}^{-1} \Psi_{f_*\underline{g}}\right)^{-1} & -\left(\Psi_{\underline{g}\underline{g}} - \Psi_{\underline{g}f_*} \Psi_{f_*f_*}^{-1} \Psi_{f_*\underline{g}}\right)^{-1} \Psi_{\underline{g}f_*} \Psi_{f_*f_*}^{-1} \\ -\Psi_{f_*f_*}^{-1} \Psi_{f_*\underline{g}} \left(\Psi_{\underline{g}\underline{g}} - \Psi_{\underline{g}f_*} \Psi_{f_*f_*}^{-1} \Psi_{f_*\underline{g}}\right)^{-1} & \Psi_{f_*f_*}^{-1} \left(I + \Psi_{f_*\underline{g}} \left(\Psi_{\underline{g}\underline{g}} - \Psi_{\underline{g}f_*} \Psi_{f_*f_*}^{-1} \Psi_{f_*\underline{g}}\right)^{-1} \Psi_{\underline{g}f_*} \Psi_{f_*f_*}^{-1}\right) \end{bmatrix}.$$

The general strategy is to compute $\Psi_{f_*f_*}^{-1}$ analytically, and then let the computer calculate the inverse $(\Psi_{\underline{g}\underline{g}} - \Psi_{\underline{g}f_*} \Psi_{f_*f_*}^{-1} \Psi_{f_*\underline{g}})^{-1}$ and everything else, as these are of smaller dimensions.

We have

$$\frac{\partial r}{\partial f_u} = I_T \otimes P_u \quad ; \quad \frac{\partial r}{\partial f_j} = I_T \otimes (P_j g_j) \quad ; \quad \frac{\partial r_t}{\partial f^s} = \begin{cases} G & \text{if } t = s \\ 0 & \text{else} \end{cases},$$

being $HT \times T$, $HT \times T$, and $H \times (J+1)$ matrices respectively. Here, $G = (P_u, P_1 g_1, \ldots, P_J g_J)$. It follows that:

$$\Psi_{f_*f_*} = \frac{\partial r}{\partial f_*^\top} V \frac{\partial r}{\partial f_*} = \begin{bmatrix} G^\top V_1 G & & 0 \\ 0 & \ddots & \\ 0 & & G^\top V_T G \end{bmatrix}$$

so that

$$\Psi_{f_*f_*}^{-1} = \begin{bmatrix} (G^\top V_1 G)^{-1} & & 0 \\ 0 & \ddots & \\ 0 & & (G^\top V_T G)^{-1} \end{bmatrix}.$$

This just involves computing T inverses of matrices $G^\top V_t G$ each with dimensions $(J+1) \times (J+1)$.

References

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

[2] 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.

[3] Bickel, P.J., Klaassen, C. A. J., Ritov, Y. and J. A. Wellner, 1993, Efficient and adaptive estimation for semiparametric models (The John Hopkins University Press, Baltimore and London).

[4] Brown, S.J., 1989, The number of factors in security returns, Journal of Finance 44, 1247-1262.

[5] Connor, G. and R.A. Korajczyk, 1988, Risk and return in an equilibrium APT: application of a new test methodology, Journal of Financial Economics 21, 255-289.

[6] Connor, G. and R.A. Korajczyk, 1993, A test for the number of factors in an approximate factor model, Journal of Finance 48, 1263-1288.

[7] Daniel, K., M. Grinblatt and S. Titman, 1997, Measuring mutual fund performance with characteristic-based benchmarks, Journal of Finance 52, 1035-1058.

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

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

[10] Davidson, R. and J.G. Mackinnon, 1993, Estimation and Inference in Econometrics (Oxford University Press, New York).

[11] Fama, E.F. and K.R. French, 1992, The cross-section of expected stock returns, Journal of Finance 47, 427-465.

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

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

[14] Fama, E.F. and K.R. French, 1996, Multifactor explanations of asset pricing anomalies, Journal of Finance 51, 55-84.

[15] Fama, E.F. and K.R. French, 1998, Value versus growth: the international evidence, Journal of Finance 53, 1975-2000.

[16] Fan, J., and I. Gijbels, 1996, Local polynomial modelling and applications (Chapman and Hall, London).

[17] Haugen, R., 1995, The new finance: the case against efficient markets (Prentice-Hall, Englewood Cliffs, New Jersey).

[18] Hodrick, R., D. Ng and P. Sengmueller, 1999, An international dynamic asset pricing model, International Taxation and Public Finance 6, 597-620.

[19] Hsiao, C., 2003, Analysis of panel data. Second edition. (Econometric Society Monograph 34, Cambridge).

[20] Lakonishok, J., A. Shleifer and R.W. Vishny, 1994, Contrarian investment, extrapolation and risk, Journal of Finance 49, 1541-1578.

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

[22] Linton, O.B. and J.P. Nielsen, 1995, A kernel method of estimating structured nonparametric regression based on marginal integration, Biometrika 82, 93-100.

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

[24] Masry, E., 1996, Multivariate local polynomial regression for time series: Uniform strong consistency and rates, Journal of Time Series Analysis 17, 571-599.

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

[26] Rosenberg, B., K. Reid and R. Lanstein, 1985, Persuasive evidence of market inefficiency, Journal of Portfolio Management 11, 9-17.

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

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

[29] Silverman, B.W., 1986, Density estimation for Statistics and Data Analysis. Chapman and Hall: New York. Princeton University Press.

[30] Stone, C.J., (1980), Optimal rates of convergence for nonparametric estimators, *Annals of Statistics,* 8, 1348-1360.

[31] White, H., (1984), Asymptotic Theory for Econometricians, Academic Press, New York.

Table 1
Distributions of the Security Characteristics

| Year (five selective years shown)	Number of securities	Log(market value)				Log(book-to-price ratio)				Correlation between the characteristics
		Mean	Variance	Skewness	Excess kurtosis	Mean	Variance	Skewness	Excess kurtosis	
7/63-6/64	963	3.79	3.44	0.314	-0.373	-0.506	0.781	-4.377	64.024	-0.282
7/72-6/73	2163	4.21	2.89	0.372	-0.218	-0.477	0.606	-0.575	0.820	-0.350
7/82-6/83	4002	3.62	3.64	0.342	-0.342	-0.163	0.777	-0.959	2.133	-0.063
7/92-6/93	4661	4.47	4.15	0.366	-0.242	-0.716	1.133	-1.198	4.522	-0.165
7/01-6/02	4738	5.40	4.58	0.330	-0.170	-0.615	1.050	-0.284	0.829	-0.490
Average over all years	3737	4.23	3.63	0.355	-0.221	-0.550	0.823	-1.023	4.691	-.234

For five selected years (the first, last, and three intermediate years at ten-year intervals) the table shows the number of firms, the first four cross-sectional moments of the unstandardized size and value characteristics, and the cross-sectional correlation between the two characteristics. The last row shows the average across all 39 annual cross-sections.

Figure 1
Bandwidths Related to Target Points of the Size Characteristic
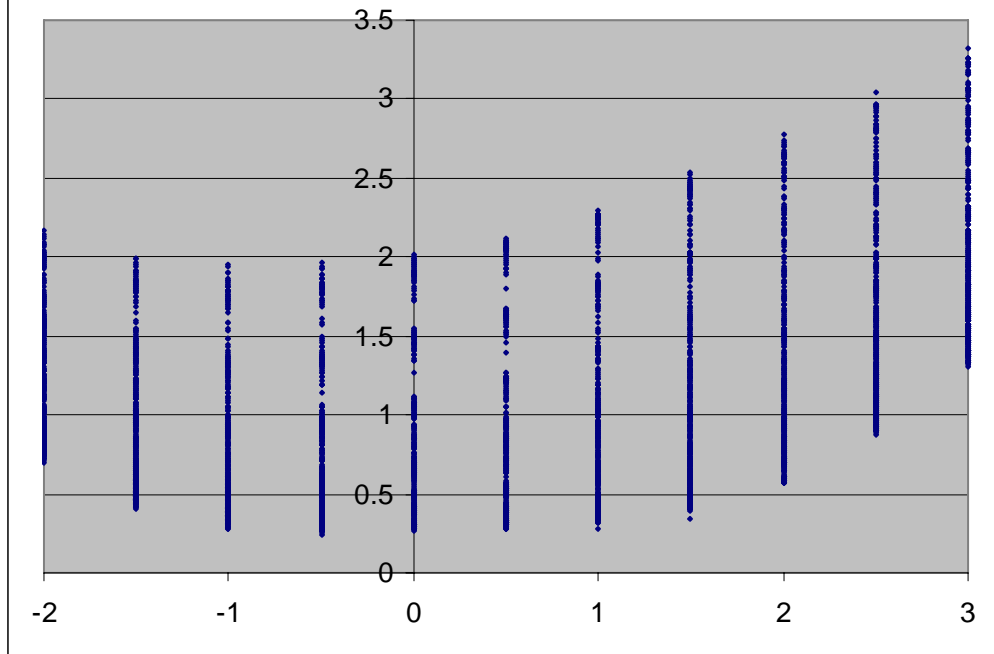
The figure shows the 4719 bandwidths (one for each of the 121 kernel portfolios for each of the 39 years) sorted by the target value of the size characteristic.

Figure 2
Bandwidths Related to Target Points of the Value Characteristic
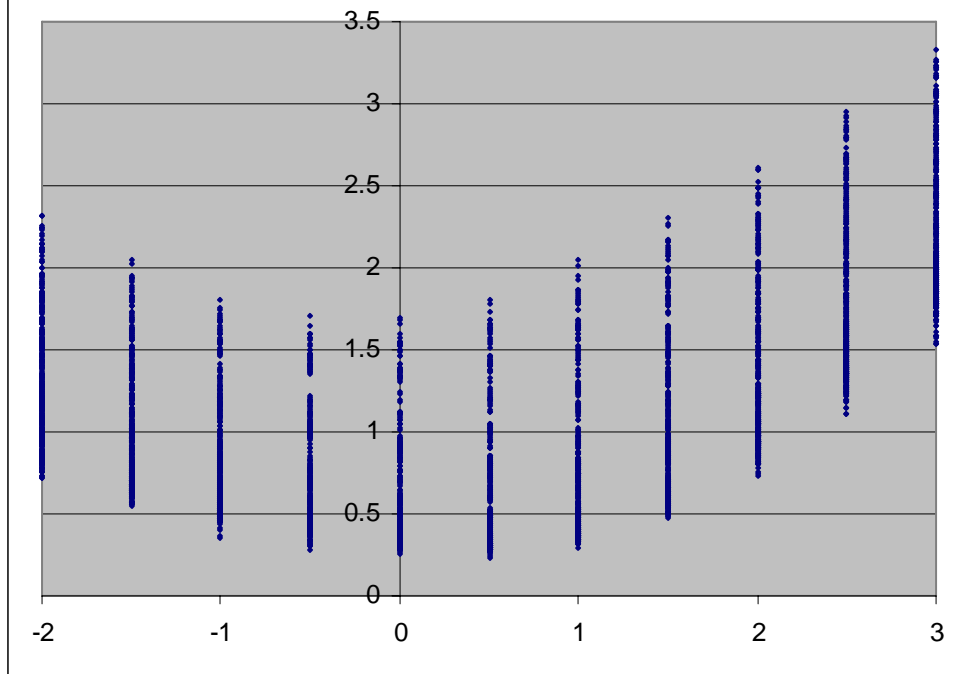
The figure shows the 4719 bandwidths (one for each of the 121 kernel portfolios for each of the 39 years) sorted by the target value of the value characteristic.

Table 2
Estimated Characteristic-Beta Functions

| Standardized Characteristic	Coefficients		Standard Errors of the Coefficients	
	Size factor betas	Value factor betas	Size factor betas	Value factor betas
-2.0	-1.36683	-2.58113	0.344935	0.843173
-1.5	-1.2521	-2.13233	0.327413	0.730035
-.1.0	-0.98441	-1.53518	0.288445	0.583113
-.5	-0.54118	-0.79766	0.227904	0.411341
-0	0	0	0	0
.5	0.542428	0.652333	0.126254	0.223383
1.0	1	1	0	0
1.5	1.326042	1.142241	0.15248	0.282011
2.0	1.524904	1.21038	0.171697	0.298094
2.5	1.63813	1.247786	0.183848	0.309015
3.0	1.705015	1.270598	0.191337	0.316954

The table shows the estimated factor betas for each point on the selected grid of characteristic values. The model is estimated by weighted nonlinear regression using a three-factor model that is based on two characteristics (value and size). The factor betas are set to zero and one for standardized characteristic values zero and one (respectively) as an identification condition of the nonlinear regression model.

Figure 3
Characteristic-Beta Function for the Size Characteristic
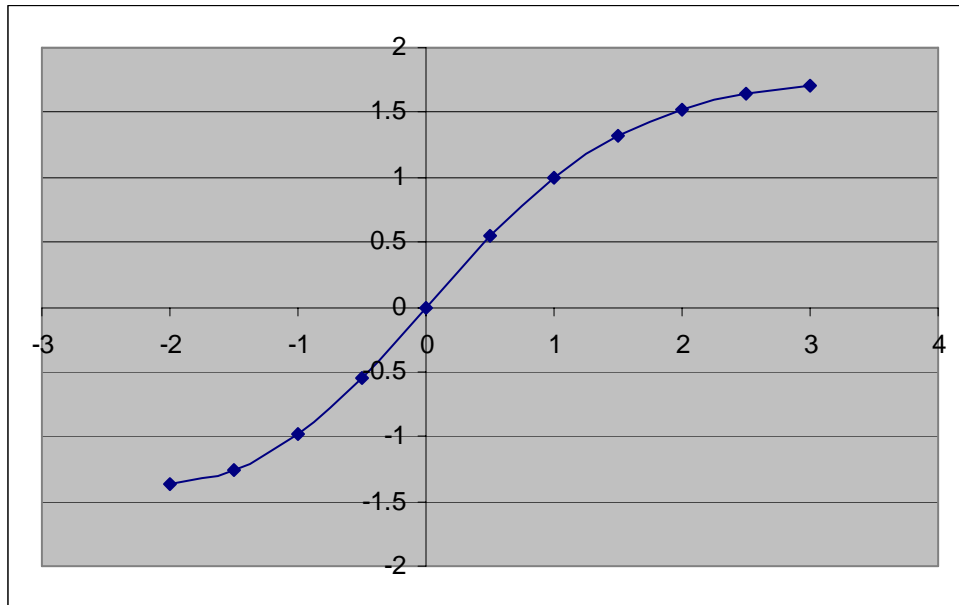
The figure displays the relationship between the size factor betas and the standardised size characteristic; see Table 2 columns one and two.

Figure 4
Characteristic-Beta Function for the Value Characteristic
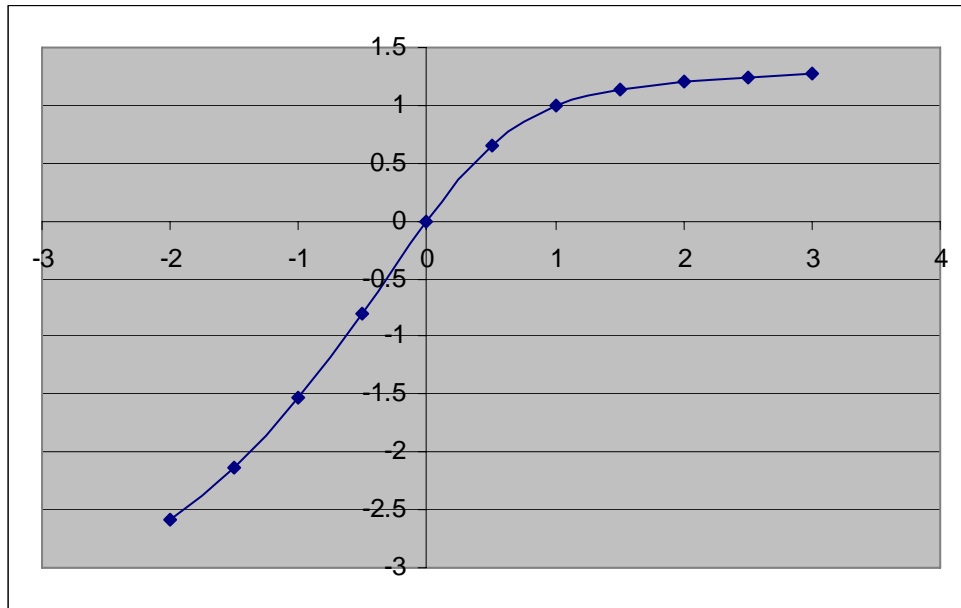
The figure displays the relationship between the value factor betas and the standardised value characteristic; see Table 2 columns one and three.

Table 3
Correlations between the factor returns

	f_u	f_s	f_v	EWMKT	VWMKT	SMB	HML
f_u	1	-0.539	-0.254	**0.998**	**0.840**	0.698	-0.220
f_s		1	0.014	-0.523	-0.069	**-0.781**	0.093
f_v			1	-0.288	-0.430	-0.140	**0.789**
EWMKT				1	0.849	0.700	-0.255
VWMKT					1	0.304	-0.371
SMB						1	-0.252
HML							1

The table shows the time-series contemporaneous correlation coefficients between our three factors, f_u, f_s, f_v (unit-beta factor, size factor, and value factor), the equally-weighted market index, EWMKT, and the three factors provided by Ken French, VWMKT, SMB and HML (capitalization-weighted market index, small-minus-big size factor, and high-minus-low value factor). The correlations are calculated over the 468 month sample period and each has an asymptotic standard error of 0.046.
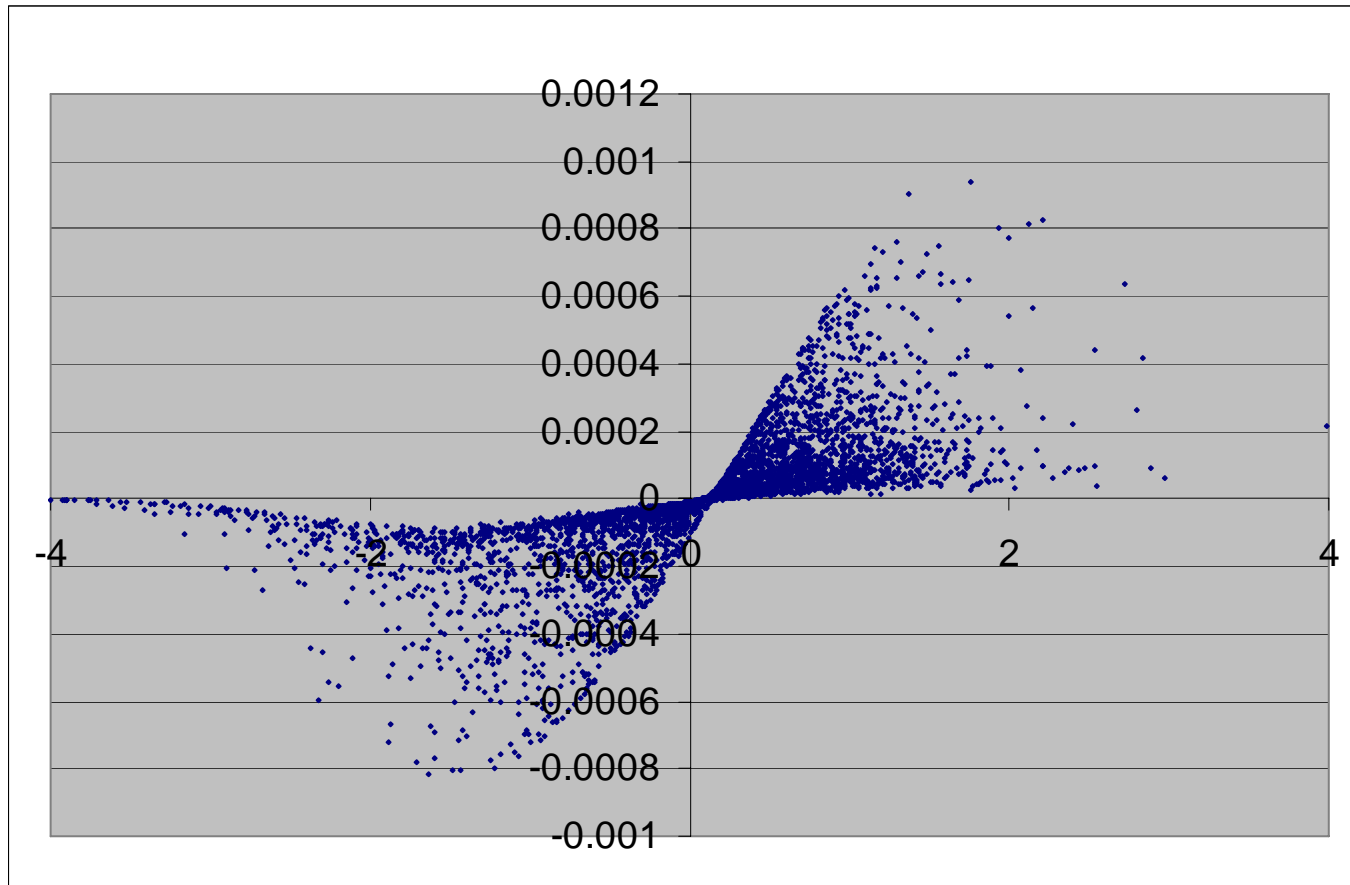
Figure 5: Size factor portfolio weights related to size characteristic
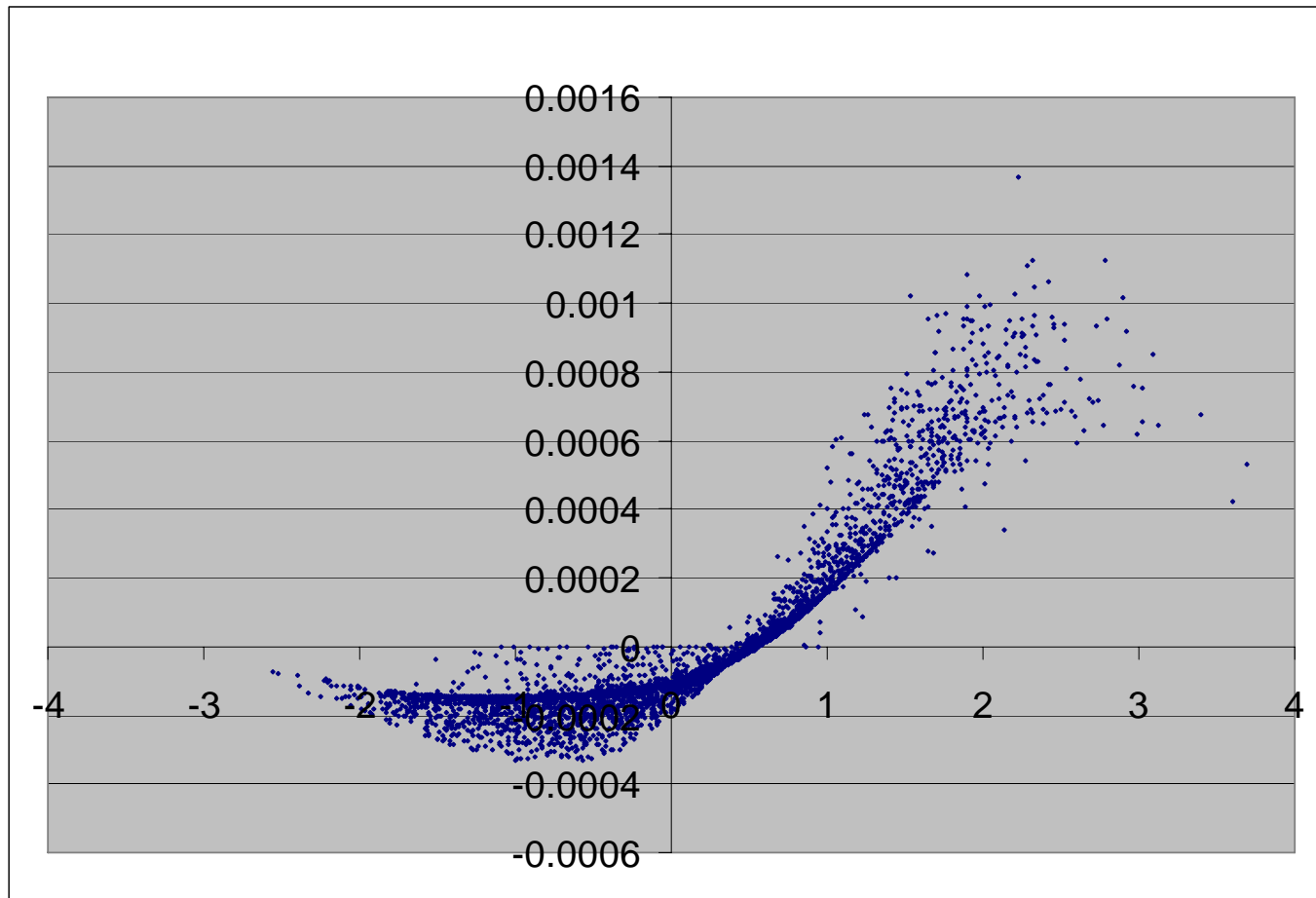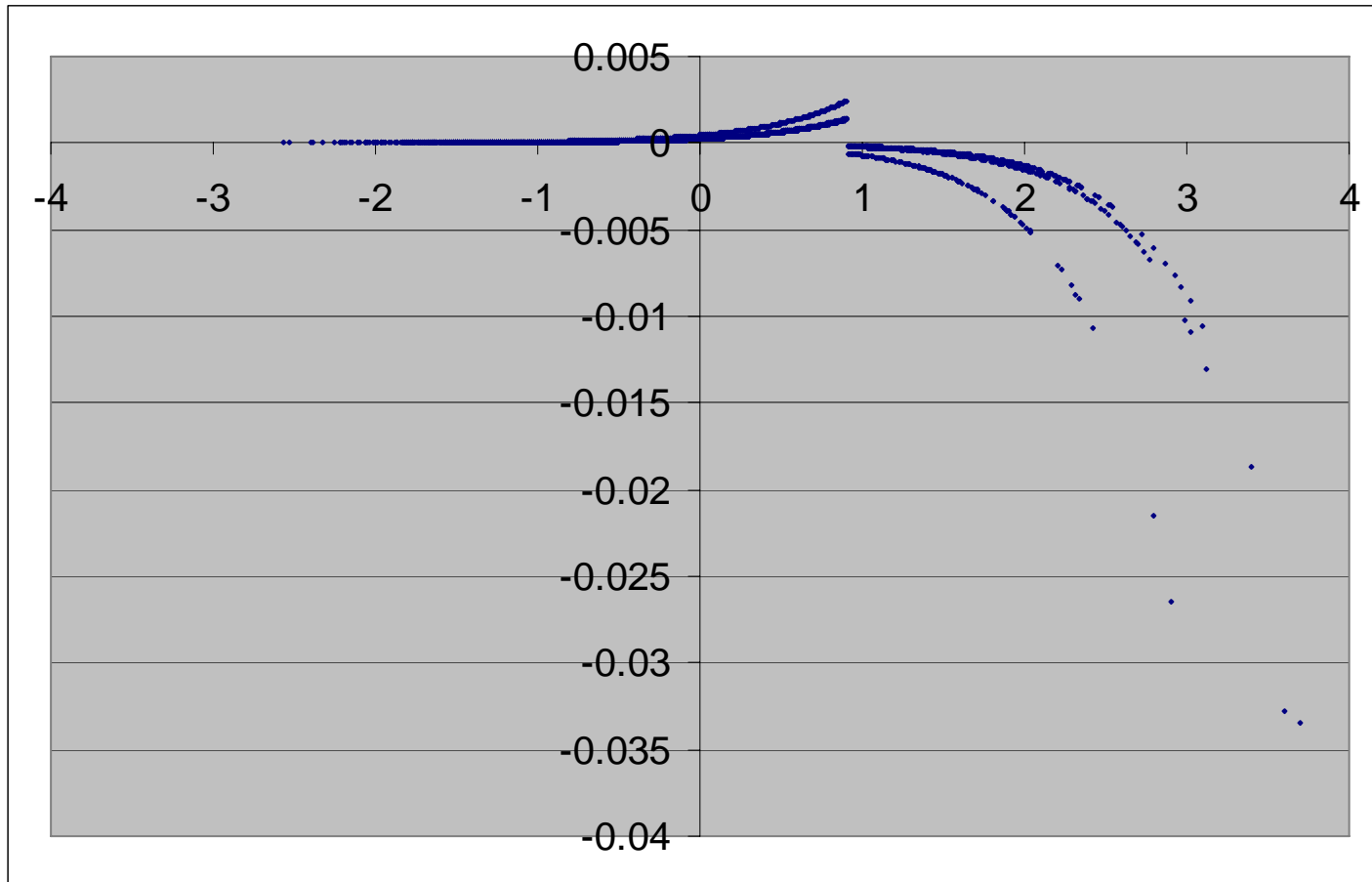

The figure shows the portfolio weights of the size factor plotted against the size characteristic, for the middle month of the sample (November 1982).

Figure 6: Value factor portfolio weights related to value characteristic

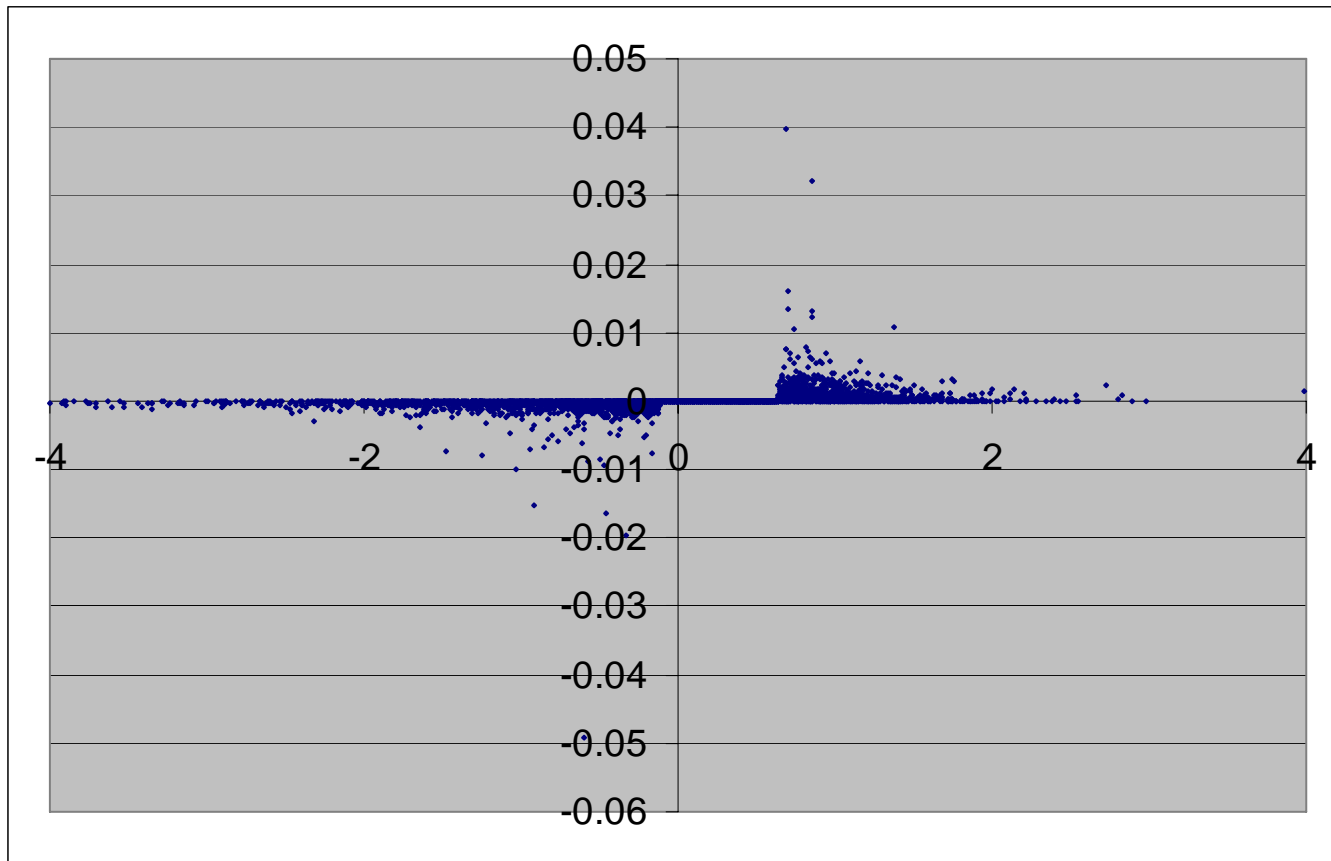The figure shows the portfolio weights of the value factor plotted against the value characteristic, for the middle month of the sample (November 1982).

Figure 7: Fama-French SMB portfolio weights related to size characteristic

This figure shows the Fama-French SMB (Small Minus Big) portfolio weights plotted against the size characteristic, for the middle month of the sample (November 1982).

Figure 8: Fama-French HML portfolio weights related to value characteristic

The figure shows the Fama-French HML (High Minus Low) portfolio weights plotted against the value characteristic for the middle month of the sample (November 1982).

Table 4
Factor Model Fit Using Time-Series Regressions

	Average Adjusted R^2		Average Residual Variance	
	CL	FF	CL	FF
Individual Assets	.2030	.1935	.02471	.02557
100 Value-weighted Sort Portfolios	.7629	.7639	.00279	.00275
100 Equally-weighted Sort Portfolios	.7943	.7683	.00269	.00278
30 Value-Weighted Industry Portfolios	.5135	.5212	.00119	.00117
30 Equally-Weighted Industry Portfolios	.6446	.6133	.00088	.00102
121 Kernel Portfolios	.9817	0.9035	7.742e-05	4.147e-04

The table reports the average fit from sets of time-series regressions with asset returns as dependent variables and three factors plus intercept as independent variables. We use two alternative sets of factors in the regressions. The columns labelled CL use the three factors f_u, f_s, f_v (unit-beta factor, size factor, and value factor) derived by our model. The columns labelled FF use the three factors provided by Ken French, VWMKT, SMB and HML (capitalization-weighted market index, small-minus-big size factor, and high-minus-low value factor). The first set of dependent variables are all the individual asset returns. The next two sets of dependent variables are 100 value-weighted and equally-weighted sort portfolios (doubly sorted by capitalization and book-to-price) provided by Ken French. The next two are 30 value-weighted and equally-weighted industry portfolios also provided by Ken French. The six and final set of dependent variables are the 121 kernel portfolios derived in our model. Both R^2 and residual variance are degrees-of-freedom adjusted.

Table 5
Model Fit After Deleting Each Explanatory Variable

	Variable Deleted	Decrease in Average Adjusted R^2		Increase in Average Residual Variance		Proportion of assets rejecting the restriction at 95% confidence	
		CL	FF	CL	FF	CL	FF
Individual Assets:	Intercept	.0068	.0126	-2.813e-04	-6.379e-05	.0484	.0528
	Market	.1109	.0783	2.114e-03	1.528e-03	.2057	.1601
	Value	.0262	.0303	7.792e-04	1.106e-03	.0885	.0909
	Size	.0130	.0136	3.838e-04	6.179e-04	.0696	.0715
100 Value-Weighted Sort Portfolios:	Intercept	.0025	.0101	1.028e-04	1.342e-04	.2500	.8600
	Market	.5928	.4647	2.074e-03	1.630e-03	.9800	1.000
	Value	.0967	.1148	3.007e-04	5.005e-04	.9300	.8600
	Size	.0307	.0427	1.613e-04	1.660e-04	.8600	.8800
100 Equally-Weighted Sort Portfolios:	Intercept	.0024	.0094	1.113e-04	1.316e-04	.2800	.8600
	Market	.6182	.4835	2.240e-03	1.760e-03	.9800	1.000
	Value	.1028	.1167	3.351e-04	5.341e-04	.9100	.8600
	Size	.0299	.0495	1.587e-04	1.888e-04	.8700	.8500
30 Value-Weighted Industry Portfolios:	Intercept	.0017	.0066	5.517e-06	2.171e-05	.1389	.5833
	Market	.4448	.4168	1.540e-03	1.410e-03	.8333	.8333
	Value	.1105	.0179	3.412e-04	7.091e-05	.8333	.6944
	Size	.0127	.0193	4.939e-05	6.519e-05	.4167	.6667
30 Equally-Weighted Industry Portfolios:	Intercept	.0010	.0036	4.339e-06	1.414e-05	.3333	.5278
	Market	.4250	.3001	1.700e-03	1.190e-03	.8333	.8333
	Value	.0146	.1551	4.474e-05	6.803e-04	.5556	.8056
	Size	.0122	.0345	5.186e-05	1.243e-04	.5278	.7222
121 Kernel Portfolios:	Intercept	.0013	.0092	5.623e-05	3.629e-04	.6860	.8843
	Market	.6575	.5043	2.183e-03	1.635e-03	1.000	1.000
	Value	.0946	.1859	2.852e-04	8.153e-04	.9835	.9422
	Size	.0347	.0496	1.831e-04	1.557e-4	.9669	.7107

The table shows the change in the results for sets of time-series regressions described in Table 4 when one of the independent variables is deleted. Both R^2 and residual variance are degrees-of-freedom adjusted. The last two columns summarize the results from the set of t-tests of the hypothesis that the true coefficient on the associated independent variable is zero.