Efficient Semiparametric Estimation of the Fama-French Model and Extensions

Gregory Connor
The London School of Economics

Matthias Hagmann
Concordia Advisors and Swiss Finance Institute

Oliver Linton*
The London School of Economics

August 15, 2007

Abstract

This paper develops a new estimation procedure for characteristic-based factor models of stock returns. We treat the factor model as a weighted additive nonparametric regression model, with the factor returns serving as time-varying weights, and a set of univariate nonparametric functions relating security characteristic to the associated factor betas. We use a time-series and cross-sectional pooled weighted additive nonparametric regression methodology to simultaneously estimate the factor returns and characteristic-beta functions. By avoiding the curse of dimensionality our methodology allows for a larger number of factors than existing semiparametric methods. We apply the technique to the three-factor Fama-French model, Carhart's four-factor extension of it adding a momentum factor, and a five-factor extension adding an own-volatility factor. We find that momentum and own-volatility factors are at least as important if not more important than size and value in explaining equity return comovements. We test the multifactor beta pricing theory against the Capital Asset Pricing model using a standard test, and against a general alternative using a new nonparametric test.

*Connor: Department of Finance, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom, E-mail g.connor@lse.ac.uk.; Hagmann: Concordia Advisors, Unit 112 Harbour Yard, Chelsea Harbour, London SW10 0XD, United Kingdom. E-mail: mhagmann@concordiafunds.com. Linton: Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail: o.linton@lse.ac.uk.

KEYWORDS: Additive Models; Arbitrage pricing theory; Characteristic-based factor model; Kernel estimation; Nonparametric regression.

JEL CODES: G12, C14.

1 Introduction

Individual stock returns have strong common movements, and these common movements can be related to individual security characteristics such as market capitalization and book-to-price ratios. Rosenberg (1974) develops a factor model of stock returns in which the factor betas of stocks are linear functions of observable security characteristics. Rosenberg's approach requires a strong assumption of linearity. Fama and French (1993) use portfolio grouping to estimate a characteristic-based factor model without assuming linearity. They estimate a three-factor model, with a market factor, size factor and value factor. The market factor return is proxied by the excess return to a value-weighted market index. The size factor return is proxied by the difference in return between a portfolio of low-capitalization stocks and a portfolio of high-capitalization stocks, adjusted to have roughly equal book-to-price ratios. The value factor is proxied by the difference in return between a portfolio of high book-to-price stocks and a portfolio of low book-to-price stocks, adjusted to have roughly equal capitalization. Using these factor returns, the factor betas are estimated via time-series regression.

Connor and Linton (2007) use a semiparametric method which combines elements of the Rosenberg and Fama-French approaches. They describe a characteristic-based factor model like Rosenberg's but replacing Rosenberg's assumption that factor betas are linear in the characteristics with an assumption that factor betas are smooth nonlinear functions of the characteristics. In a model with two characteristics, size and value, plus a market factor, they form a grid of equally-spaced characteristic-pairs. They use multivariate kernel methods to form factor-mimicking portfolios for the characteristic-pairs from each point on the grid. Then they estimate factor returns and factor betas simultaneously using bilinear regression applied to the set of factor mimicking portfolio returns.

A weakness of the Connor-Linton methodology is the reliance on multivariate kernel methods to create factor-mimicking portfolios. These multivariate kernel methods severely restrict the number of factors which can be estimated well using their technique due to the curse of dimensionality, see Stone (1980). The same problem appears in a different guise with the Fama-French methodology. To create their size and value factor returns, Fama and French double-sort assets into size and value categories. Adding a third characteristic with this method requires triple-sorting, adding a fourth requires quadruple-sorting; like Connor-Linton, the method quickly becomes unreliable for typical sample sizes and more than two characteristic-based factors.

In this paper we develop a new estimation methodology that does not require any portfolio grouping or multivariate kernels. Instead, we estimate the factor returns and characteristic-beta functions using weighted additive nonparametric regression. This relies on the fact that in each time period the characteristic-based factor model proposed in Connor-Linton is a weighted additive sum of univariate characteristic-based functions. The nonparametric part of the estimation problem is made univariate by decomposing the full problem into an iterative set of sub-problems in each characteristic singly, a standard trick in weighted additive nonparametric regression. We modify the weighted additive nonparametric regression methodology to account for our model's feature that the weights vary each time period while the characteristic-beta functions stay constant. The theoretical basis for our estimation method has been developed in a series of papers: Mammen, Linton, and Nielsen (1999), Linton, Nielsen, and van de Geer (2004), Linton and Mammen (2005), and Linton and Mammen (2006). See also Carrasco, Florens and Renault (2006) for an intuitive discussion and application to other areas in economics.

Our model falls in the class of semiparametric panel data models for large cross-section and long time series. There has been some work on semiparametric models for panel data, see for example Kyriazidou (1997), and nonparametric additive models, see for example Porter (1996) and more recently Mammen, Støve, and Tjøstheim (2006). Most of this work is in the context of short time series. More recently, there has been work on panel data with large cross-section and time series dimension, especially in finance where the datasets can be large along both dimensions and in macroeconomics where there are cross-sectional panels of many related series (such as business conditions survey data) with quite long time series length. Some recent papers include Phillips and Moon (1999), Bai and Ng (2002), Bai (2004,2005), and Pesaran (2006). These authors have addressed a variety of issues including nonstationarity, estimation of unobserved factors, and model selection. They all work with essentially parametric models. Our semiparametric model allows more flexibility with regard to the functional form. We establish pointwise asymptotic normality of the functional components of our model at what appears to be an optimal rate. We also establish the asymptotic normality of our estimated factors.

Our model allows for any number of factors with no theoretical loss of efficiency, and we exploit this in our application. In addition to the market, size and value factors of the standard Fama-French model, we add a momentum factor as suggested by Jagadeesh and Titman (1993) and Carhart (1997), and an own-volatility factor, a choice influenced by the recent work of Goyal and Santa Clara (2003) and Ang, Hodrick, Xing and Zhang (2006a, 2006b). This reflects the feature that our methodology allows us to estimate a model with more factors. We find that the two added factors, momentum and volatility, are as important or more important than size and value in explaining equity return comovements. Hence, the improved data-efficiency of our new method has real empirical value.

We test the single-factor Capital Asset Pricing Model against a five-factor asset pricing model using the standard Gibbons, Ross and Shanken (1989) test; we find that the CAPM is rejected. We also develop a new nonparametric test for multifactor pricing models as part of our estimation methodology. To implement the test we assume that mispricing is a smooth multivariate function of observable security characteristics. We estimate this mis-pricing function simultaneously with the factor model of returns, and test whether the mispricing function is the null function. We find that the five-factor model does a good job of explaining asset return premia; the alpha function differs only negligibly from a null function, at least for the four security characteristics that we consider.

We evaluate various time series models for the risk factors. We establish the asymptotic properties of two-stage estimators of the parameters of this model and estimate vector autoregressions both for the levels of the factor returns and the factor returns squared to explore factor return and volatility dynamics.

We proceed as follows: Section 2 presents the model. Section 3 describes the estimation algorithm in the balanced and unbalanced panel case. Section 4 develops the distribution theory. Section 5 presents an empirical application to the cross-section of monthly U.S. stock returns. Section 6 summarizes the findings and concludes. The Appendix contains all the proofs.

2 The Model

We assume that there is a large number of securities, indexed by $i = 1, \ldots, n$. Asset excess returns (returns minus the risk free rate) are observed for a number of time periods $t = 1, \ldots, T$, where $n/T \to \infty$ as $n, T \to \infty$. We assume that the following characteristic-based factor model generates excess returns:

$$y_{it} = f_{ut} + \sum_{j=1}^{J} g_j(X_{ji}) f_{jt} + \varepsilon_{it}, \tag{1}$$

where y_{it} is the excess return to security i at time t; f_{ut}, f_{jt} are the factor returns; $g_j(X_{ji})$ the factor betas, X_{ji} are observable security characteristics, and ε_{it} are the mean zero asset-specific returns. 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 is the same type of factor model used by Connor and Linton (2007). To simplify the exposition we are assuming that the characteristics X_{ji} are time invariant. We will later on discuss the case where characteristics are allowed to vary over time.

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). In applications to returns data it is convenient to exclude own-effect intercept terms from (1) since they provide little benefit in terms of explanatory power and necessitate an additional time-series estimation step; see Connor and Korajczyk (1988, 1993), Connor and Linton

(2007).

We also the allow for case where one or more of the factors are directly observed and need not be estimated. So for example in our empirical analysis we investigate the special case where the market factor f_{ut} is exogenously observable, and in particular is equal to a capitalization-weighted index. Other observable factors, such as the time-series innovations to unemployment or inflation, or fixed income portfolio returns, are possible within our econometric framework.

Note that for fixed t, equation (1) constitutes a weighted additive nonparametric regression model for panel data, where the factor returns f_{jt} are 'parametric weights' and the characteristic-beta functions $g_j(\cdot)$ are univariate nonparametric functions. Some discussion of additive nonparametric models can be found in Linton and Nielsen (1995). The situation here is somewhat nonstandard, since the same regression equation (1) holds each time period, with parametric weights varying each time period and the characteristic-beta functions time-invariant. We extend the weighted nonparametric regression methodology to account for this feature of time-varying weights in a pooled time-series, cross-sectional model.

Our model can be thought of as a special case of the usual statistical factor model

$$y_{it} = \sum_{j=1}^{J} \beta_{ij} f_{jt} + \varepsilon_{it}, \tag{2}$$

where the factor loadings β_{ij} are unrestricted, Ross (1976). Connor and Koracyzk (1993) developed the asymptotic principal component method for estimation of the factors in the case where the cross-section is large but the time series is fixed. Recent work of Bai and Ng (2002) and Bai (2004,2005) have provided analysis for this method for the case where both n and T are large. Bai (2004) establishes pointwise asymptotic normality for estimates of the factors (at rate $\sqrt{n}$) and the loadings (at rate $\sqrt{T}$) under weak assumptions regarding cross-sectional and temporal dependence.[1] The nesting of our model within (2) could be used for specification testing. Note however that in the case where the covariates in (1) are time varying, this nesting no longer holds. Rosenberg (1974) considers the case where $\beta_{ij} = X_{ji}$ for some observed characteristic for which estimation is just simple panel linear regression.

2.1 Factor Scale Identification Conditions

In the case in which both characteristic-beta functions and factor returns are estimated from the data there is an obvious scale indeterminacy in the model. We assume that the observed characteristic $J-$vectors of the assets $X_i, i = 1, \ldots, n$ are independent and identically distributed across i. We

[1] Bai assumes that the loadings β_{ij} are fixed in repeated samples but treats f_{jt} as random.

impose the identifying restrictions that for each factor the cross-sectional average beta equals zero and the cross-sectional variance of beta equals one, that is, $E^*[g_j(X_{ji})] = 0$ and $\text{var}^*[g_j(X_{ji})] = 1$, where E^* and var^* denote moments with respect to some distribution. Note that this does not restrict the return model since the additive semiparametric model (1) is invariant to this rescaling. The choice of distribution to use in the normalization affects the interpretation of the factors. The condition $\text{var}^*[g_j(X_{ji})] = 1$ sets the magnitude of factor return j; the conditions $E^*[g_j(X_{ji})] = 0$ affects the interpretation of the intercept. If we use the population distribution, then $E^*[g_j(X_{ji})] = 0$ means that the intercept can be interpreted as the return to the average asset in the infinite population of assets; if we use a capitalization-weighted population distribution, then $E^*[g_j(X_{ji})] = 0$ means that the intercept can be interpreted as the return to the capitalization-weighted average asset. We will consider both of these in our empirical implementation; for simplicity our econometric theory focuses on the case in which we use the unweighted population distribution.

2.2 Estimating Additive Nonparametric Mispricing Functions

A central concern in the asset pricing literature is the determination of the expected returns on assets and their relationship to the risk exposures of the assets. Note that the information set of investors includes the characteristics X. Taking investors' expectation of excess returns y_{it} (conditional on the information set)using (1):

$$E[y_{it}] = E[f_{ut}] + \sum_{j=1}^{J} g_j(X_{ji}) E[f_{jt}], \tag{3}$$

which is the standard multi-factor asset pricing model: expected excess returns are linear in factor betas. Hence our model as developed so far imposes the standard multi-factor pricing condition on expected excess returns.

Our methodology provides a new asset pricing test against a general nonparametric pricing alternative. Fama and French (1993) create factor mimicking portfolios from size and value-sorted portfolios and then estimate characteristic-related mispricing based on a finer grid of value and size-sorted portfolios. This two-stage procedure leaves open the question whether there is a hidden "identification condition" when using the same characteristics to create mimicking portfolios and to test for mispricing. We show that this is in fact the case.

Adapting the Fama-French mispricing test to our additive nonparametric framework generates an explicit identification condition. The characteristic-mispricing functions are only identified up to an orthogonality condition relative to the characteristic-beta functions. This is because the same characteristics are used to identify the factor risk premia and factor model mispricing.

We assume that there are mispricing inefficiencies given by a smooth additive univariate nonparametric functions $\alpha_j(X_{ij})$ using the same characteristics X_{ij} as in the factor model.[2] The return generating process becomes:

$$y_{it} = f_{ut} + \sum_{j=1}^{J} \alpha_j(X_{ij}) + g_j(X_{ji}) f_{jt} + \varepsilon_{it}. \tag{4}$$

In order for the functions $\alpha_j(Z_{ij})$ to be identified we must impose:

$$E[\alpha_j(X_{ij})] = 0 \tag{5}$$

$$E[\alpha_j(X_{ij}) g_j(X_{ji})] = 0. \tag{6}$$

The mean-zero condition (5) is standard in additive nonparametric models, so that the intercept f_{ut} can be identified. The condition (6) is necessary in order for the risk premia of each factor return to be identified. To see why this is so, suppose that we relax the identification condition, then for any constant a we can replace $\alpha_j(X_{ij})$ with $\alpha_j^*(X_{ij}) = \alpha_j(X_{ij}) + a g_j(X_{ji})$ and $f_{jt}^* = f_{jt} - a$ and the fit of the model is exactly the same. This indeterminacy is only eliminated by imposing (6). The intuition for the condition is clear: mean return which is in the linear span of the characteristic-beta function must be treated as "factor risk premia" rather than "mispricing."

2.3 An Important Special Case: An Observed Market Factor

As mentioned above it is possible to avoid estimation of one or more of the factor returns by replacing them with observed proxies. An important case is replacing the estimated intercept f_{ut} with a market index return f_{mt}. Let returns obey a $J + 1$ statistical factor model:

$$y_{it} = \sum_{j=0}^{J} \beta_{ij} f_{jt} + \varepsilon_{it} \tag{7}$$

and let the market index return f_{mt} be some linear combination of the $J + 1$ factor returns. By a suitable factor rotation we can rewrite (7) as:

$$y_{it} = \beta_{0i}^* f_{mt} + \sum_{j=1}^{J} \beta_{ij}^* f_{jt}^* + \varepsilon_{it},$$

where $\{f_{jt}^*, j = 1, \ldots, J\}$ is the set of the other J factors after the rotation. In order to implement our approach we impose $\beta_{0i}^* = 1$ and that $\sum_{j=1}^{J} \beta_{ij}^* f_{jt}^*$ can be written in the additive nonparametric form shown in (1). This produces the same factor model formulation as (1) but with $f_{ut} = f_{mt}$ observed

[2] It is possible to include additional nonparametric functions based on other observed variates strictly exogenous relative to y_{it}; this does not require any additional identification conditions beyond the mean-zero condition.

rather than estimated. Note that the condition $\beta_{0i}^* = 1$ is well supported empirically (see Fama and French (1993) and Connor and Linton (2007)) and that this does not require that the market betas of all assets equal one. Rather it requires that the market betas of assets can be written as one plus a linear combination of their non-market factor exposures (see (9) in the next subsection).

In the case of an identified market factor (using a capitalization-weighted index) the identification conditions $E^*[g_j(X_{ji})] = 0$ must use the capitalization-weighted population distribution. This ensures that if we take capitalization-weighted averages of both sides of (1) we get $f_{mt} = f_{mt}$.

2.4 Embedding the CAPM as a Testable Restriction

Consider the case described in the previous subsection in which $f_{ut} = f_{mt}$ an observed market index. In this subsection we will describe how to restrict that version of the model to encompass the CAPM. Assume (admittedly, with a loss of generality, and against some of the empirical evidence) that the factor returns are i.i.d. through time. In this case we can easily embed the Capital Asset Pricing Model inside (3) and test the restriction. The CAPM imposes the pricing restriction:

$$E[y_{it}] = \frac{\mathrm{cov}(y_{it}, f_{mt})}{\mathrm{var}(f_{mt})} E[f_{mt}]. \tag{8}$$

Using (1) to derive market covariances, and then dividing by $\mathrm{var}(f_{mt})$:

$$\frac{\mathrm{cov}(y_{it}, f_{mt})}{\mathrm{var}(f_{mt})} = 1 + \sum_{j=1}^{J} g_j(X_{ij}) \frac{\mathrm{cov}(f_{jt}, f_{mt})}{\mathrm{var}(f_{mt})}. \tag{9}$$

Since the functions $g_j(X_{ij})$ can take arbitrary values, the CAPM (8) will hold if and only if:

$$E[f_{jt}] = \frac{\mathrm{cov}(f_{jt}, f_{mt})}{\mathrm{var}(f_{mt})} E[f_{mt}] \tag{10}$$

which is an easily testable condition. We present empirical evidence in Section Five below.

3 Estimation Strategy

For simplicity of exposition we focus on the case in which all the factors are estimated and there are no mispricing functions $\alpha_j(Z_{ij})$ included in the factor model. Connor and Linton (2007) propose to estimate the period by period conditional expectation of y_{it} given the characteristics $X_{1i}, \ldots, X_{Ji}$ at a grid of points and then to estimate the factors and beta functions at the same grid of points using an iterative algorithm based on bilinear regression. This approach works well enough when the cross-section is very large and when J is small, like two in their case. However, it is inefficient in general and works poorly in practice when J is larger than two. For this reason we develop an alternative estimation strategy that makes efficient use of the restrictions embodied in (1).

In order to describe the statistical properties of our estimators we make some assumptions about the data generating process. For notational convenience, we treat in detail the case of a *fully balanced panel*, where the set of assets and the characteristics of each asset do not vary through time. (In subsection 3.6 below we describe the modifications necessary for the case of an unbalanced panel.) We assume that each ε_{it} is a martingale difference sequence with finite conditional and unconditional variance.

3.1 Population Characterization

To motivate our estimation methodology we first define the parameters of interest through a population least squares criterion. This is one way of defining the quantities f, g consistent with (1); it has the advantage of usually implying an efficient procedure under i.i.d. normal error terms. The solution to this population problem is characterized by first order conditions; to derive estimators we mimic this population first order condition by a sample equivalent. For clarity, we just treat the case where all the factors are unknown - if some factors are known then they do not need to be chosen in the optimization below.

Consider the population criterion

$$Q_T(f, g) = \frac{1}{T} \sum_{t=1}^{T} E\left[\left\{y_{it} - f_{ut} - \sum_{j=1}^{J} g_j(X_{ji}) f_{jt}\right\}^2\right]. \tag{11}$$

In this criterion, the expectation is taken over the distribution of returns and characteristics, treating the factors as fixed parameters that are to be chosen - we are thinking of the factors as an exogenous stochastic process. Under some conditions there may exist a limiting (as $T \to \infty$) criterion function $Q(f, g)$ but we do not require this. We minimize $Q_T(f, g)$ with respect to the factors f (which contains f_{ut}, f_{jt} for all j, t) and the functions $g = (g_1, \ldots, g_J)$ subject to the identifying restrictions $E^*[g_j(X_{ji})] = 0$ and $\mathrm{var}^*[g_j(X_{ji})] = 1$.

This minimization problem can be characterized by a set of first order conditions for f, g. For expositional purposes we shall divide the problem in two: an equation characterizing f given known g, and an equation characterizing g given f.

3.1.1 Characterization of the Factor Returns

First we solve for the minimization of (11) over f_{ut}, f_{jt} for all j, t given $g(.)$ is known. Note that if the population of assets is treated as fixed rather than random, then (11) simply amounts to a collection of unrelated cross-sectional regression problems, one per time period. In this case the solution to the minimization problem is obviously period-by-period least squares regression. We now show that this intuition extends to our environment with a random population of assets rather than a fixed cross-section.

Taking the first derivatives of (11) with respect to f_{ut}, f_{jt} and setting to zero, the first order conditions are (for each $t = 1, \ldots, T$):

$$E\left[\left\{y_{it} - f_{ut} - \sum_{j=1}^{J} g_j(X_{ji}) f_{jt}\right\}\right] = 0, \tag{12}$$

$$E\left[\left\{y_{it} - f_{ut} - \sum_{k=1}^{J} g_k(X_{ki}) f_{kt}\right\} g_j(X_{ji})\right] = 0,\ j = 1, \ldots, J. \tag{13}$$

These equations are linear in f given g. This delivers a linear system of $J+1$ equations in $J+1$ unknowns for each time period t. Letting $f_t = [f_{ut}, f_{1t}, \ldots, f_{Jt}]^\top$, $y_t = [y_{1t}, \ldots, y_{nt}]^\top$, and $G(X_i) = [1, g_1(X_{1i}), \ldots, g_J(X_{Ji})]^\top$ we have $E[G(X_i)G(X_i)^\top] f_t = E[G(X_i) y_t]$. It follows that

$$f_t = E[G(X_i)G(X_i)^\top]^{-1} E[G(X_i) y_t],$$

provided $E[G(X_i)G(X_i)^\top]$ is non-singular, which we assume to be the case.

3.1.2 The Characteristic-beta Functions

Next we turn to the characterization of g given f. Consider the point-wise derivative of (11) with respect to $g_j(X_j)$ conditional on a fixed value of X_j :

$$\lim_{\delta \to 0} \frac{1}{T} \sum_{t=1}^{T} E\left[\left\{y_{it} - f_{ut} - \{g_j(X_{ji}) + \delta\} f_{jt} - \sum_{k \neq j} g_k(X_{ki}) f_{kt}\right\}^2 | X_{ji} = x_j\right] / \delta$$

Setting this derivative to zero gives a first-order condition defining the criterion-minimizing functions $g_j(x)$ at value x_j.

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} f_{jt} E\left[y_{it} | X_{ji} = x_j\right] &= \frac{1}{T}\sum_{t=1}^{T} f_{jt} f_{ut} + g_j(x_j) \frac{1}{T}\sum_{t=1}^{T} f_{jt}^2 \\ &\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{k \neq j} f_{jt} f_{kt} E\left[g_k(X_{ki}) | X_{ji} = x_j\right] \end{aligned} \tag{14}$$

for $j = 1, \ldots, J$. These equations are linear in g given f but they are only implicit, that is, they constitute a system of integral equations (of type 2) in the functional parameter g, see Mammen, Linton, and Nielsen (1999) and Linton and Mammen (2005). In the appendix we give some discussion of the properties of this integral equation. To simplify the notation define the conditional expectations:

$$\lambda_{1t}(j,x) = E\left[y_{it}|X_{ji}=x\right] \quad ; \quad \lambda_2(j,k,x) = E\left[g_k(X_{ki})|X_{ji}=x\right].$$

Substituting these functions into (14) and rearranging we obtain

$$g_j(x_j) = \frac{\sum_{t=1}^T f_{jt}\left[\lambda_{1t}(j,x_j) - f_{ut} - \sum_{k\neq j} f_{kt}\lambda_2(j,k,x_j)\right]}{\sum_{t=1}^T f_{jt}^2} \tag{15}$$

for $j = 1, \ldots, J$. The solution to (15) does not impose the identification conditions on $g_j(.)$. One can impose these restrictions by instead considering the constrained optimization problem and manipulating the first order condition of the associated Lagrangean. An equivalent approach is to take any unrestricted solution $g_j(x_j)$ and replace it by

$$\overline{g}_j(x_j) = \frac{g_j(x_j) - \int g_j(x_j)dP_j^*(x_j)}{\sqrt{\int g_j^2(x_j)dP_j^*(x_j)}}, \tag{16}$$

where P_j^* is the probability distribution associated with the identifying assumption on $g_j(.)$.

3.2 The Kernel Estimates

The first set of kernel estimates measure the period-specific expected return of an asset given its j^{th} characteristic is x_j and conditional on the observed factor returns in period t. We use the following (boundary corrected) kernel estimate:

$$\widehat{\lambda}_{1t}(j,x_j) = \frac{\sum_{i=1}^n K_h(X_{ji},x_j)y_{it}}{\sum_{i=1}^n K_h(X_{ji},x_j)},$$

where for each x in the support of X_t, $K_h(x,y) = K_h^x(x-y)$ for some kernel K^x such that $K_h^x(u) = h^{-1}K^x(h^{-1}u)$ and $K_h^x(u) = K_h(u)$ for all x in the interior of the support of X_{ji}. Here, $K_h(.) = K(./h)/h$ and K is a kernel while h is a bandwidth. We shall assume that each covariate is supported on $[\underline{x}, \overline{x}]$ for some known $\underline{x}, \overline{x}$ and that the covariate density is bounded away from zero on this support.

The second set of kernel estimates give the expected factor beta of asset i for factor j based on the asset's characteristic X_{ji} and the estimated characteristic-beta function $g_k^{[i]}(\cdot)$:

$$\widehat{\lambda}_2^{[i]}(j,k,x_j) = \frac{\sum_{i=1}^n K_h(X_{ji},x_j)g_k^{[i]}(X_{ki})}{\sum_{i=1}^n K_h(x_j,x_j)}.$$

Both of these kernel estimates are familiar features of weighted additive nonparametric regression. Note that $\widehat{\lambda}_2^{[i]}$ does not require a time subscript since under our assumption of a fully-balanced panel, all assets have constant characteristics over time. We will weaken this assumption later.

3.3 Estimation of Factor Returns and Characteristic-Beta Functions

We replace the unknown quantities in A, b_t and equation (15) by estimated values, denoted by hats, and iterate between the factor return f and characteristic-beta function $g(.)$ estimation problems. The solution for f depends upon $g(.)$, and the solution for $g_j(.)$ depends both upon f and $\{g_k(.), k \neq j\}$. We use the Gauss-Seidel iteration to reconcile these component solutions.

Let $f^{[0]}, g^{[0]}(.)$ be initial estimates.

Then let for all x

$$\begin{aligned} \widehat{g}_j^{[i+1]}(x) &= \frac{\sum_{t=1}^T \widehat{f}_{jt}^{[i]}(\widehat{\lambda}_{1t}(j,x) - \widehat{f}_{ut}^{[i]})}{\sum_{t=1}^T \widehat{f}_{jt}^{[i]2}} - \frac{\sum_{t=1}^T \sum_{k>j} \widehat{f}_{jt}^{[i]} \widehat{f}_{kt}^{[i]} \widehat{\lambda}_2^{[i]}(j,k,x)}{\sum_{t=1}^T \widehat{f}_{jt}^{[i]2}} \\ &\quad - \frac{\sum_{t=1}^T \sum_{k<j} \widehat{f}_{jt}^{[i]} \widehat{f}_{kt}^{[i]} \widehat{\lambda}_2^{[i+1]}(j,k,x)}{\sum_{t=1}^T \widehat{f}_{jt}^{[i]2}} \end{aligned} \tag{17}$$

$$\widehat{g}_j^{[i+1]}(x) = \frac{\widehat{g}_j^{[i+1]}(x) - \int \widehat{g}_j^{[i+1]}(x) dP_j^*(x)}{\sqrt{\int \widehat{g}_j^{[i+1]}(x)^2 dP_j^*(x)}}, \tag{18}$$

where hats denote estimated quantities defined above.

Then given estimates $\widehat{g}_j^{[i]}(X_{ji})$ from the previous iteration on g given f we compute for each t

$$\widehat{f}_t^{[i+1]} = \left[\frac{1}{n} \sum_{i=1}^n \widehat{G}^{[i]}(X_i) \widehat{G}^{[i]}(X_i)^\top \right]^{-1} \frac{1}{n} \sum_{i=1}^n \widehat{G}^{[i]}(X_i) y_{it}, \tag{19}$$

where $\widehat{G}^{[i]}(X_i) = [1, \widehat{g}_1^{[i]}(X_{1i}), \ldots, \widehat{g}_J^{[i]}(X_{Ji})]^\top$.

Iterating (17),(18), and (19) repeatedly to convergence gives the final simultaneous solutions for $\widehat{f}$ and $\widehat{g}(.)$. The convergence properties of this algorithm are not studied here, we refer the reader to Mammen, Linton, and Nielsen (1999) for a fuller discussion of this issue in a special case of our model.

3.4 Initial Estimators

The procedure we propose works iteratively, estimating the factor returns given the characteristic-beta functions, and the characteristic-beta functions given the factor returns. We describe two approaches to finding starting values.

The first approach produces consistent starting values under the model assumptions, and is based on using time averaged data. This approach has similarities to the averaging method proposed in Pesaran (2006) except that our averaging is over time rather than cross-sectional. In particular, let $\{w_{Tt}\}$ be some sequence with $\sum_{t=1}^{T} w_{Tt} \leq \overline{w} < \infty$, and let $\overline{y}_i = \sum_{t=1}^{T} w_{Tt} y_{it}$, $\overline{\varepsilon}_i = \sum_{t=1}^{T} w_{Tt} \varepsilon_{it}$, $\overline{f}_u = \sum_{t=1}^{T} w_{Tt} f_{ut}$, $\overline{f}_j = \sum_{t=1}^{T} w_{Tt} f_{jt}$. For example, $w_{Tt} = 1/T$. Then we have

$$\overline{y}_i = \overline{f}_u + \sum_{j=1}^{J} g_j(X_{ji}) \overline{f}_j + \overline{\varepsilon}_i = \overline{f}_u + \sum_{j=1}^{J} \overline{g}_j(X_{ji}) + \overline{\varepsilon}_i, \tag{20}$$

where $\overline{g}_j(.) = g_j(.)\overline{f}_j$. This constitutes an additive nonparametric regression with components $\overline{g}_j(.)$ that are mean zero, i.e., $E[\overline{g}_j(X_{ji})] = 0$, $j = 1, \ldots, J$. This means that we can estimate the functions $\overline{g}_j(.)$ by the smooth backfitting method of Mammen, Linton, and Nielsen (1999). Then, since $g_j(.)$ is proportional to $\overline{g}_j(.)$ we see that in fact $g_j(.) = \overline{g}_j(.)/\int \overline{g}_j(x_j)^2 dP_j^*(x_j)$, i.e., a corresponding renormalization of the estimated $\overline{g}_j$ yields an estimate of g_j. The quantity $\overline{f}_u$ can be estimated by the grand mean $\overline{y} = \sum_{i=1}^{n} \overline{y}_i/n$. The theory of Mammen, Linton, and Nielsen (1999) can be directly applied to the estimator of $\overline{g}_j$, except that the error term in (20) is $O_p(T^{-1/2})$, which makes the convergence rate of the nonparametric estimators faster by this magnitude. To estimate the time series factors f_t rather than their means we must go back and cross-sectionally regress y_{it} on a constant and $\widehat{g}_1(X_{1i}), \ldots, \widehat{g}_J(X_{Ji})$ for each time period t.[3] The procedure described above provides consistent initial estimates for the case of the fully balanced panel.

The second approach to starting values uses a variant of Rosenberg's (1974) linear model:

$$g_j(X_{ji}) = X_{ji}. \tag{21}$$

In this linear case it is simple to rescale the characteristics so that the identification constraints hold using (21). We scale the mean and variance of the characteristics so that $E^*[X_{ji}] = 0$ and $\text{var}^*[X_{ji}] = 1$ for each j; for each characteristic, this just requires subtracting the weighted cross-sectional mean and dividing by the weighted cross-sectional standard deviation each time period. The simple linear model for $g(.)$ gives rise to a linear cross-sectional regression model to estimate f_{jt} :

$$y_{it} = f_{ut} + \sum_{j=1}^{J} X_{ji} f_{jt} + \varepsilon_{it} \tag{22}$$

We begin with ordinary least squares estimation of (22). These estimates of f_{ut} and f_{jt} serve merely as starting values and have no consistency properties. Connor and Linton (2007) find that this linear model provides quite a reasonable first approximation. As long as these initial estimates are in a convergent neighbourhood of the maximizing values, the final estimates after repeated iteration will be unaffected. We use this second approach in our empirical implementation below.

[3] In our case, the cross-sectionally averaged returns are not informative except about f_{ut}, since $\overline{y}_t = f_{ut} + \sum_{j=1}^{J} \overline{g}_j f_{jt} + \overline{\varepsilon}_t = f_{ut} + O_p(n^{-1/2})$ under the assumption that $E[g_j(X_{ji})] = 0$, where $\overline{y}_t = \sum_{i=1}^{n} y_{it}/n$, $\overline{\varepsilon}_t = \sum_{i=1}^{n} \varepsilon_{it}/n$, and $\overline{g}_j = \sum_{i=1}^{n} g_j(X_{ji})/n$. Like Pesaran (2006) we require that not all $\overline{f}_j = 0$.

3.5 The Iterative Algorithm

To summarise, the algorithm has the following steps

1. Initialize $f_{ut}^{[0]}$, $f_{jt}^{[0]}$ by cross-sectionally estimating a linear characteristic-based factor model via period-by-period cross-sectional weighted least squares regression of (22).

2. Iteratively solve for $g_j(.)$ in (15) using the most recent iterative estimates of f and $g_k(.)$, $k \neq j$.

3. Re-estimate f_t by cross-sectional regression of y_{it} on an intercept and $g(X_t)$.

4. Repeat steps (2) and (3) until a convergence criteria is met. Let $\widehat{f}_{jt}, \widehat{g}_j(.)$ be the estimators on convergence.

3.6 Unbalanced, Time-varying Panel Data

We have so far mostly assumed a fully balanced panel dataset. The set of observed assets is assumed constant over time, with each asset having a fixed vector of characteristic betas. The only time variation in this fully balanced panel comes through the random factor realizations and random asset-specific returns. In applications, the set of assets must be allowed to vary over the time sample, since the set of equities with full records over a reasonably long sample period is a small subset of the full dataset. Also, the characteristics of the assets must be allowed to vary through time.

We assume that the observations are unbalanced in the sense that in time period t we only observe n_t firms (for simplicity labelled $i = 1, \ldots, n_t$). Also, we assume that the characteristics are time varying but stationary over time for each i and i.i.d. over i. This yields first order conditions for f, g that are similar to the balanced case. Now the matrix A depends on time, while the expression for g_j becomes

$$g_j(x_j) = \frac{\frac{1}{T}\sum_{t=1}^{T} \frac{1}{n_t}\sum_{i=1}^{n_t} f_{jt}\left(E\left[y_{it}|X_{jit} = x_j\right] - f_{ut} - \sum_{k\neq j} f_{kt} E\left[g_k(X_{kit})|X_{jit} = x_j\right]\right)}{\frac{1}{T}\sum_{t=1}^{T} f_{jt}^2}. \tag{23}$$

The estimation algorithm is essentially the same as outlined above. The two unknown expectations in (23) are replaced by kernel estimates given by

$$\begin{aligned} \widehat{\lambda}_{1t}(j, x_j) &= \frac{\sum_{i=1}^{n_t} K_h(X_{jit}, x_j) y_{it}}{\sum_{i=1}^{n_t} K_h(X_{jit}, x_j)} \\ \widehat{\lambda}_{2t}^{[i]}(j, k, x_j) &= \frac{\sum_{i=1}^{n_t} K_h(X_{jit}, x_j) g_k^{[i]}(X_{kit})}{\sum_{i=1}^{n_t} K_h(X_{jit}, x_j)}. \end{aligned} \tag{24}$$

Note that $\widehat{\lambda}_{1t}$ and $\widehat{\lambda}_{2t}^{[i]}(j, k, x_j)$ now both have to be estimated in each sample period separately to allow for time variation.

4 Distribution Theory

In this section we provide the distribution theory for our estimates of the factors and of the characteristic functions in the balanced case. The general approach uses the methods developed in Mammen, Linton, and Nielsen (1999) and Linton and Mammen (2005) for treating estimators defined as the solutions of type 2 linear integral equations. The novelty here is due to the weighting by the factors and the fact that we wish to allow both the cross-section and the time dimension to grow. Regarding the asymptotics, we take joint limits as $n, T \to \infty$ under the restriction that $n/T \to \infty$ as described in Phillips and Moon (1999, Definition 2(b)).

Let $\mathcal{F}_a^b$ be the σ-algebra of events generated by the vector random variable $\{U_t; \quad a \leq t \leq b\}$. The processes $\{U_t\}$ is called strongly mixing [Rosenblatt (1956)] if

$$\sup_{1\leq t} \sup_{A\in\mathcal{F}_{-\infty}^t, B\in\mathcal{F}_{t+k}^\infty} \left|\Pr\left(A\cap B\right) - \Pr(A)\Pr(B)\right| \equiv \alpha(k) \to 0 \quad \text{as } k \to \infty.$$

We make the following assumptions.

Assumptions A.

A1. *The double array $\{X_i, \varepsilon_{it}\}_{i,t=1}^{n,T}$ are defined on a probability space $(\Omega, \mathcal{F}, P)$. For each t, (X_i, ε_{it}) are i.i.d. across i. The processes $\{\varepsilon_{it}\}$ are strongly mixing with a common mixing coefficient, $\alpha(k)$ such that $\alpha(k) \leq C\overline{\alpha}^k$ for some $C \geq 0$ and some $\overline{\alpha} < 1$. Furthermore, $E(\varepsilon_{it}|\mathcal{F}_{t-1}) = 0$ and $E(\varepsilon_{it}^2|\mathcal{F}_{t-1}) = \sigma_t^2(X_i)$ a.s., where $\mathcal{F}_{t-1}$ is the sigma field generated by X_i and the past of ε_{it}. Furthermore, for some $\kappa > 4$, $\sup_t E[|\varepsilon_{it}|^\kappa] < \infty$.*

A2. *The covariate $X_i = (X_{1i}, \ldots, X_{Ji})^\top$ has absolutely continuous density p supported on $\mathcal{X} = [\underline{x}, \overline{x}]^J$ for some $-\infty < \underline{x} < \overline{x} < \infty$. The functions $g_j(\cdot)$ together with the density $p(\cdot)$ are twice continuously differentiable over the interior of $\mathcal{X}$ and are bounded on $\mathcal{X}$. The density function $p(x)$ is strictly positive at each $x \in \mathcal{X}$. Denote by $p_j(x)$ the marginal probability density for characteristic j with support $\mathcal{X}_j = [\underline{x}, \overline{x}]$. The matrix $E[G(X_i)G(X_i)^\top]$ is strictly positive definite.*

A3. *For each $x \in [\underline{x}, \overline{x}]$ the kernel function K^x has support $[-1, 1]$ and satisfies $\int K^x(u)du = 1$ and $\int K^x(u)udu = 0$, such that for some constant C, $\sup_{x\in[\underline{x},\overline{x}]} |K^x(u) - K^x(v)| \leq C|u - v|$ for all $u, v \in [-1, 1]$. Define $\mu_j(K) = \int u^j K(u)du$ and $||K||_2^2 = \int K^2(u)du$. The kernel K is bounded, has compact support ($[-c_1, c_1]$, say), is symmetric about zero, and is Lipschitz continuous, i.e., there exists a positive finite constant C_2 such that $|K(u) - K(v)| \leq C_2 \left|u - v\right|$.*

A4. *$n, T \to \infty$ in such a way that $n/T \to \infty$.*

A5. *For some* $a \geq 2$, *the quantities* $\sup_{T\geq 1}\sum_{t=1}^{T} f_{jt}^{a}/T < \infty$. *The quantities* $\Phi_j = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} f_{jt}^2$ *and* $\Psi_j(x_j) = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} f_{jt}^2\sigma_{jt}^2(x_j)$ *exist and* $\Phi_j > 0$. *Here,* $\sigma_{jt}^2(x_j) = E[\varepsilon_{it}^2|X_{ji} = x_j]$.

A6. *The bandwidth sequence* $h(n,T)$ *satisfies* $nh^4(\log T)^{\rho} \to 0$ *and* $nTh \to \infty$ as $n, T \to \infty$.

A7. *For* $j = u, 1, \ldots, J$, *there exists* $\rho' > 0$ *such that* $\max_{1\leq t\leq T}|f_{jt}| = O((\log T)^{\rho'})$.

In A1 we allow ε_{it} to have certain types of nonstationarity: this is possible since the CLT is coming from the cross-sectional independence. In particular, we can allow general form of time series and cross-sectional conditional heteroskedasticity. We could also allow ε_{it} to be weakly autocorrelated, with no change in the limiting distribution, but the data do not seem to require this, so for simplicity we maintain mean unpredictability of the errors. We can also allow ε_{it} to be cross-sectionally correlated without affecting the result for $\widehat{g}_j$. However, this will affect the asymptotic distribution of $\widehat{f}_t$ in such a way that it makes inference difficult without some assumptions on the nature of this cross-sectional dependence. Assumption A5 embodies an assumption about the magnitudes of the factors; here we assume that they behave like the outcome of a stationary process with finite moments of order a. This could be relaxed to either faster growth in $\sum_{t=1}^{T} f_{jt}^2$, reflecting nonstationary factors, or slower growth reflecting perhaps many zero values in the factors, but we do not do this here as the data seem to support this assumption. Assumption A7 is needed for the uniform convergence rates below. This assumption is consistent with the factors being realizations from a stationary Gaussian process. Again, this condition could be weakened to allow faster growth in $\max_{1\leq t\leq T}|f_{jt}|$ at the expense of further restrictions elsewhere.

Define:

$$\Omega_j(x_j) = \frac{\Psi_j(x_j)}{p_j(x_j)\Phi_j^2}||K||_2^2. \tag{25}$$

$$V_{t,t} = E[G(X_i)G(X_i)^{\top}]^{-1}E[\varepsilon_{it}^2 G(X_i)G(X_i)^{\top}]E[G(X_i)G(X_i)^{\top}]^{-1}. \tag{26}$$

THEOREM 1. *Suppose that Assumptions A1-A6 hold. Then, for each t there exists a stochastically bounded sequence* $\delta_{n,t}$ *such that*

$$\sqrt{n}(\widehat{f}_t - f_t - h^2\delta_{n,t}) \Longrightarrow N(0, V_{t,t}), \tag{27}$$

while $\widehat{f}_t, \widehat{f}_s$ *with* $t \neq s$ *are asymptotically independent. Suppose also that A7 holds. Then, for some* $\rho > 0$

$$\max_{1\leq t\leq T}\left|\widehat{f}_t - f_t\right| = O_p((n^{-1/2} + h^2)(\log T)^{\rho}). \tag{28}$$

Also, given Assumptions A1-A7, there exists a bounded continuous function $\beta_j(.)$ *such that for each* $x_j \in (\underline{x}, \overline{x})$,

$$\sqrt{nTh}\left(\widehat{g}_j(x_j) - g_j(x_j) - h^2\beta_j(x_j)\right) \Longrightarrow N(0, \Omega_j(x_j)). \tag{29}$$

Furthermore, $\widehat{g}_j(x_j), \widehat{g}_k(x_k)$ are asymptotically independent for any x_j, x_k for $j \neq k$.

REMARKS.

1. Provided the bandwidth is chosen so that $nh^4 \to 0$, the estimator $\widehat{f}_t$ is consistent at rate $n^{-1/2}$ and the asymptotic distribution is as if the characteristic functions were known and least squares were applied. The estimators $\widehat{g}_j(x_j)$ are consistent at rate $(nT)^{-2/5}$ provided a bandwidth of order $(nT)^{-1/5}$ is chosen and under some restrictions on the rates at which T, n increase. This should be the optimal rate for this problem, Stone (1980).[4] It can be that $\widehat{g}_j(x_j)$ converges to $g_j(x_j)$ faster than $\widehat{f}_t$ converges to f_t; this happens when $T^4/n \to \infty$. This is because of the extra pooling over time in the specification of g_j. Note that the asymptotic variance of the characteristic function estimates is as if the factors were known.

2. Suppose that ε_{it} is homoskedastic, i.e., $\sigma^2_{jt}(x_j) = \sigma^2_\varepsilon$ for all j, t, the asymptotic variance of $\widehat{g}_j(x_j)$ simplifies to $\Omega_j(x_j) = (\sigma^2_\varepsilon/(p_j(x_j)\Phi_j))||K||^2_2$. We argue that this is a natural 'oracle' bound for the performance of estimators in this case, see Linton (1997). Suppose that we could observe the partial residuals $U_{jit} = y_{it} - f_{ut} - \sum_{k\neq j} f_{kt} g_k(X_{ki})$, then we can compute the pooled regression smoother

$$\widehat{g}_j^{oracle}(x) = \frac{\sum_{t=1}^T \sum_{i=1}^n K_h(X_{ji}, x_j) f_{jt} U_{jit}}{\sum_{t=1}^T \sum_{i=1}^n K_h(X_{ji}, x_j) f^2_{jt}}. \tag{30}$$

This shares the asymptotic variance of our estimator, and, since it uses more information than we have available, it is comforting that our estimator performs as well as it.

2. Standard errors (for the general case) can be obtained in an obvious way by plugging in estimated quantities. In particular, valid standard errors for the factors can be obtained from the final stage least squares regression of returns on the characteristic functions. Standard errors for $\widehat{g}_j(x_j)$ can be computed from

$$\frac{\widehat{\Omega_j(x_j)}}{nTh} = \frac{\sum_{t=1}^T \sum_{i=1}^n K_h^2(X_{ji}, x_j) \widehat{f}^2_{jt} \widehat{\varepsilon}^2_{it}}{\left(\sum_{t=1}^T \sum_{i=1}^n K_h(X_{ji}, x_j) \widehat{f}^2_{jt}\right)^2}, \tag{31}$$

where $\widehat{\varepsilon}_{it} = y_{it} - \widehat{f}_{ut} - \sum_{j=1}^J \widehat{f}_{jt} \widehat{g}_j(X_{ji})$ are residuals computed from the estimated factors and characteristic functions, see Fan and Yao (1998) for discussion of nonparametric standard errors.

3. Bandwidth and factor selection can both be handled in the framework of penalized least squares, see Mammen and Park (2005). Adapted to our problem, this involves choosing bandwidth h to minimize

$$PLS(h) = \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n \widehat{\varepsilon}^2_{it} \left[1 + \frac{2JK(0)}{nTh}\right]$$

[4] The rate is partly determined by our assumption that $\sum_{t=1}^T f^2_{jt}/T$ stays bounded away from zero and infinity, as is appropriate for a stationary time series. A more general theory can be written for the case where $\sum_{t=1}^T f^2_{jt} = O(T^\alpha)$ for any $\alpha \geq 0$. In our application the case $\alpha = 1$ seems the most relevant.

over a set H of bandwidths. Mammen and Park showed the consistency of this method for additive nonparametric regression in a single cross-section. Bai and Ng (2002) propose a test for the inclusion of an additional factor based on residual sum of squares. Specifically, let

$$PC(J) = \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \widehat{\varepsilon}_{it}^{2}(J) + J\varkappa(n,T),$$

where $\varkappa(n,T) \to 0$ and $\min\{n,T\}\varkappa(n,T) \to \infty$. Then we expect that $\Pr[\widehat{J} = J] \to 1$ as $n, T \to \infty$, where $\widehat{J} = \arg\min_{J \leq J_{\max}} PC(J)$ and $J_{\max}$ is some predetermined upper bound.

4. The results (27)-(29) follow also for the unbalanced case with suitable generalizations. The case where the covariate process is stationary is particularly simple because then one only needs to replace n by n_t in (27), n by $\min_{1\leq t\leq T} n_t$ in (28), and nT by $\sum_{t=1}^{T} n_t$ in (29).

4.1 Specification Testing

One can test the underlying specification in a number of ways. Given our sampling scheme there are two general restrictions on the conditional expectation $E[y_{it}|X_i] = m_t(X_i)$: poolability, and additivity. Baltagi, Hidalgo, and Li (1996) propose a test of poolability that can be adapted to our framework. Gozalo and Linton (2001) have proposed tests of additivity in a cross-sectional setting based on comparing restricted with unrestricted estimators that work with marginal integration estimators, Linton and Nielsen (1995).

We discuss further the issue of testing whether a given parametric shape on the characteristic-beta functions g_j is plausible, i.e., $\mathbf{H}_0 : g_j(\cdot) = g_j(\cdot;\theta_{j0})$ for some parameter vector θ_{j0}. Consider the test statistic

$$\widehat{\tau} = \sum_{i=1}^{n} \left[\widehat{g}_j(X_{ji}) - g_j(X_{ji};\widehat{\theta}_j)\right]^2 \pi(X_{ji}),$$

where $\widehat{\theta}_j$ are parametric estimates of θ_{j0} with the property that under the null hypothesis $\sqrt{nT}(\widehat{\theta}_j - \theta_{j0})$ is asymptotically normal, and $\pi(\cdot)$ is a bounded continuous weighting function. Under some conditions, it can be shown that for deterministic (and estimable) sequences $\{\mu_{nT}, V_{nT}\}$

$$\frac{\widehat{\tau} - \mu_{nT}}{V_{nT}^{1/2}} \Longrightarrow N(0,1) \tag{32}$$

under the null hypothesis, while $(\widehat{\tau} - \mu_{nT})/V_{nT}^{1/2} \to^P \infty$ under fixed alternatives.[5] Let $w_{i,i'} =$

[5] A number of different nonparametric test statistics based on standard kernel estimators are reviewed in Li and Racine (2007). Haag (2007) has analyzed a testing problem involving backfitting estimators.

$K_h(X_{ji}, X_{ji'})/np_j(X_{ji})$ and $\widetilde{\varepsilon}_i = \sum_{t=1}^T f_{jt}\varepsilon_{it}/\sum_{t=1}^T f_{jt}^2$, and let $\sigma_i^2 = \mathrm{var}[\widetilde{\varepsilon}_i|X_i]$. Then

$$\mu_{nT} = h^{1/2}\sum_{i=1}^n\sum_{i'=1}^n w_{i',i}^2\sigma_{i'}^2\pi_{ji} \text{ and } V_{nT} = 2h\sum_{\substack{i=1 \\ i\neq i'}}^n\sum_{i'=1}^n \rho_{i',i}^2\sigma_{i'}^2\sigma_i^2\pi_{ji}\pi_{ji'},$$

where $\rho_{i',i} = \sum_{l=1}^n w_{i',l}w_{il}$ and $\pi_{ji} = \pi(X_{ji})$. The quantities μ_{nT} and V_{nT} do not depend on the properties of $\widehat{\theta}_j$. Furthermore, they can be estimated consistently by the plug-in method. Specifically, it suffices to replace σ_i^2 in μ_{nT}, V_{nT} by $\widehat{\widetilde{\varepsilon}}_i^2$, where $\widehat{\widetilde{\varepsilon}}_i = \sum_{t=1}^T \widehat{f}_{jt}\widehat{\varepsilon}_{it}/\sum_{t=1}^T \widehat{f}_{jt}^2$. In our empirical application below we consider the special case of Rosenberg's linear model, in which the parameter set θ_{j0} is the null set and $g_j(X_{ji}) = X_{ji}$.

4.2 Time Series Analysis

In order to test the embedded CAPM pricing restriction (10) and to examine the dynamic behaviour of the factors, we need to apply the above theory to modelling the time series behaviour of the estimated factors. A large literature has considered this problem for specific models, including Stock and Watson (1998). Suppose that the true factors obey some time series model (this is consistent with our earlier treatment of the factors as fixed under the assumption of strong exogeneity). In particular, suppose that

$$E[\psi(F_t, Z_t; \theta_0)] = 0 \tag{33}$$

for some instruments Z_t for some true value θ_0 of a vector of parameters $\theta \in \mathbb{R}^p$, where ψ is a q-vector of moment conditions with $q \geq p$ and $F_t = \mathrm{vec}(f_t, f_{t-1}, \ldots, f_{t-k})$. This includes a number of cases of interest. In our empirical application we consider the case in which f_t follows a vector autoregression $A(L)f_t = \eta_t$, where η_t is i.i.d. and $A(L) = A_0 - A_1L - \cdots - A_rL^r$ for parameter matrices $A_0, A_1, \ldots, A_r$, see Borak, Härdle, Mammen, and Park (2007). It follows that $E[A(L)f_t \otimes Z_t] = 0$ for any Z_t in the past of f_t.

To estimate the parameters we use the estimated factors and minimize the quadratic form

$$\widehat{M}_T(\theta)^\top W_T \widehat{M}_T(\theta) \tag{34}$$

$$\widehat{M}_T(\theta) = \frac{1}{T-k}\sum_{t=k+1}^T \psi(\widehat{F}_t, Z_t; \theta)$$

with respect to θ, where W_T is a symmetric positive definite weighting matrix. Let $\widehat{\theta}$ be any minimizer of (34). For simplicity we assume that the factors are stationary and mixing; for most finance applications the assumptions of stationarity (or at least local stationarity) and mixing seem reasonable.

Hansen, Nielsen, and Nielsen (2004) consider the problem of using estimated values in linear time series models. They prove a general result that provided $\sum_{t=1}^{T}(\widehat{f}_t - f_t)^2 \xrightarrow{P} 0$ as $T \to \infty$, then we may use the predicted time series as if it was the true unobserved time series for instance in estimation and unit root testing in the sense that using the estimated values leads to the same asymptotic distribution (for $T \to \infty$) as if the true values were used. Their result applies to stationary and nonstationary factors. However, in the case of stationary factors where $\sqrt{T}$ consistent estimation of θ_0 is possible, this condition is too strong. Specifically, it suffices that there is an expansion for $\widehat{f}_t - f_t$ and the uniform rate $\max_{1\leq t\leq T}|\widehat{f}_t - f_t| = o_p(T^{-1/4})$ holds, both of which are obtainable from Theorem 1.

For the factor modelling result we treat the factors as random and we need some additional assumptions. We suppose that $\{f_t, Z_t\}_{t=1}^{\infty}$ is a jointly stationary process satisfying strong mixing. In this case one should interpret $\{f_t, Z_t\}_{t=1}^{\infty}$ as being independent of X, ε and A5 and A7 as holding with probability one. Assumption A7 holds with probability one when f_t is a stationary mixing Gaussian process.

Assumptions B.

B1. *We suppose that the process (F_t, Z_t) is strictly stationary and strong mixing with $\alpha(k)$ satisfying$\alpha(k) \leq C\overline{\alpha}^k$ for some $C \geq 0$ and some $\overline{\alpha} < 1$.*

B2. *The limiting moment condition $M(\theta) = E[\psi(F_t, Z_t; \theta)]$ has a unique zero at $\theta = \theta_0$, where θ_0 is an interior point of the compact parameter set $\Theta \subseteq \mathbb{R}^p$. Furthermore, $\psi(F_t, Z_t; \theta_0)$ is a martingale difference sequence.*

B3. *The function ψ is twice continuously differentiable in both F and θ with for some $\delta > 0$*

$$E\left[\sup_{||\theta-\theta_0||\leq\delta, ||F'-F||\leq\delta}\left\|\frac{\partial^j\psi}{\partial F^{j_1}\partial^{j_2}\theta}(F', Z; \theta)\right\|^r\right] < \infty$$

for $j = 0, 1, 2$ with $j_1 + j_2 = j$ and some $r > 2$.

B4. *$W_T \xrightarrow{P} W$, where W is a symmetric positive definite matrix. The matrix Γ_0 is of full rank.*

For convenience we assume that $\psi(F_t, Z_t; \theta_0)$ is a martingale difference sequence as is plausible in the finance applications we have in mind. Define W to be the probability limit of W_T and let:

$$\Gamma_0 = E\left[\frac{\partial}{\partial\theta}\psi(F_t, Z_t; \theta_0)\right] \quad ; \quad V_0 = \mathrm{var}\left[\psi(F_t, Z_t; \theta_0)\right]$$

$$\Psi = \left(\Gamma_0^\top W\Gamma_0\right)^{-1}\left(\Gamma_0^\top WV_0W\Gamma_0\right)\left(\Gamma_0^\top W\Gamma_0\right)^{-1}.$$

THEOREM 2. *Suppose that Assumptions A1-A7 and B1-B4 given in the appendix hold. Then,*

$$\sqrt{T}(\widehat{\theta} - \theta_0) \Longrightarrow N(0, \Psi).$$

This shows that the estimation of factors does not affect the limiting distribution of the parameters of the factor process. This means that standard errors for $\widehat{\theta}$ can be constructed as if the factors were observed. To explore volatility dynamics, we also estimate a vector autoregression using squared factor returns, which is covered by Theorem 2.

5 Empirical Analysis

5.1 Data

We follow Fama French (1993) in the construction of the size and value characteristics. For each separate twelve-month period July-June from 1963 - 2005 we find all securities which have complete CRSP return records over this twelve-month period and the previous twelve month period, and both market capitalization (from CRSP) and book value (from Compustat) records for the previous June. The raw size characteristic each month equals the logarithm of the previous June's market value of equity. The raw value characteristic equals the ratio of the market value of equity to the book value of equity in the previous June. In addition to the Fama-French size and value characteristics we derive from the same return dataset a momentum characteristic as in Carhart (1997). This variable is measured as the cumulative twelve month return up to and including the previous month. Finally we add an own-volatility characteristic, a choice inspired by the recent work of Goyal and Santa Clara (2003) and Ang et al. (2006a,2006b). We define raw volatility as the standard deviation of the individual security return over twelve months up to and including the previous month. The characteristics equal the raw characteristics except standardized each month to have zero mean and unit variance. The size and value characteristics are held constant from July to June whereas the momentum and own-volatility characteristics change each month. Table 1 reports some descriptive statistics for the data: the number of securities in the annual cross section, and the first four cross-sectional moments of the four characteristics. To save space the table just shows nine representative dates (July at five-year intervals), as well as time series medians over the full 42 year period, using July data.

Three notes on the interpretation of these characteristics in terms of our econometric theory.

1. We treat all four characteristics as observed without error. Informally, we think of momentum and own-volatility as behaviourally-generated sources of return comovement. Investors observe momentum and own-volatility over the previous twelve months (along with the most recent observations of size and value), adjust their portfolio and pricing behaviour to account for the observed

values, and this in turn accounts (for some unspecified reasons) for the subsequent return comovements associated with these characteristics. Understanding more fundamentally the sources of the characteristic-related comovements is an important topic which we do not address here.

2. The cited references Ang et al. (2006a,b) and Goyal and Santa Clara (2003) use idiosyncratic volatility rather than total volatility as a characteristic. From our perspective, total volatility is preferable since it does not require a previous estimation step to remove market-related return from each asset's total return.

3. In our econometric theory we allow all the characteristics to vary freely over time. Since size and value change annually whereas momentum and own-volatility change monthly, another approach would be to modify the econometric theory to allow some characteristics to change only at a lower frequency. We do not pursue this alternative approach here.

A useful descriptive statistic is the correlation matrix of the explanatory variables. This is complicated in our model by the time-varying nature of the characteristics which serve as our explanatory variables. Figure 1 shows for each pair of characteristics the time series evolution of the cross-sectional correlation between them, using the cross-section each July. It is clear that these correlations are not constant over time. The correlation between size and value exhibits slow and persistent swings, with a negative average. Size and momentum on the other hand are on average uncorrelated. Most interesting is the relationship between own-volatility and momentum, taking large swings from high positive correlation of 0.7 to negative correlation of -0.35. None of the correlations are large enough in magnitude to be worrisome in terms of accurate identification of the model.

5.2 Implementation

In the case of a fully balanced panel it would be straightforward to estimate the characteristic-beta function at each data point in the sample. However in the presence of time-varying characteristics this is not feasible since the number of asset returns (each with a unique vector of characteristics) equals 1,886,172 in our sample. In order to make the algorithm described in Section 3 computationally feasible we concentrate estimation of the characteristic functions on 61 equally-spaced grid points between -3 and 3, which corresponds to a distance of 0.1 between contiguous grid points. We use linear interpolation between the values at these grid points to compute the characteristic-beta function at all 1,886,172 sample points. Then we use the full sample of 1,886,172 asset returns and associated factor betas to estimate the factor returns. This procedure greatly improves the speed of our algorithm while sacrificing little accuracy, since the characteristic-beta functions are reasonably linear between these closely-spaced grid points.

We chose a Gaussian kernel throughout to nonparametrically estimate the conditional expectations summarized in (24). 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 bandwidth choice is done separately for each characteristic function. We follow Connor and Linton (2007) and use their variable bandwidth tied to local data density. For each characteristic value and each year, we calculate the sample density of the root-mean-squared differences between all the sample characteristic and the individual grid point. We then set the bandwidth for this grid point 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 grid point, 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 characteristic) and wider where the data set is locally sparse (e.g., near the extreme values of the characteristic). It is rather like a smooth nearest neighbors bandwidth taking 5% of the data in each marginal window.

5.3 The Characteristic Beta Functions

Table 2 shows the estimates of the characteristic-beta functions at a small selected set of characteristic values and the heteroskedasticity-consistent standard errors from (31) for each of these estimates. To avoid any spurious non-linearity results due to smoothing in regions where there are no data, we report results for each characteristic only over a support ranging from the empirical 2.5% to the 97.5% quantile. The standard errors tend to be somewhat larger in the tails, where the data is sparser. Given that our procedure is able to use all 1.8 million return observations to estimate the characteristic-beta functions, the standard errors are small.

The characteristic-beta functions over all grid points are displayed in Figure 2. Note that these characteristic-beta functions satisfy the equally-weighted zero mean/unit variance identification conditions described in Section 3. For comparative purposes we overlay the linear Rosenberg-type model with the same identification conditions imposed. The characteristic-beta functions are mostly monotonically increasing for all four characteristics. Size and value show strongly non-linear characteristic-beta functions, both with concave shapes. The observed shapes for momentum and own-volatility are closer to linear.

5.4 Explanatory Power of Each Factor

Note that at each step of the iterative estimation, the factor returns are the coefficients from period-by-period unconstrained cross-sectional regression of returns on the previous iteration's factor betas. To measure the explanatory power of the factors we take the final-step estimates of factor betas and perform the set of cross-sectional regressions with all the factors, each factor singly, and all the factors except each one. Table 3 shows the time-series averages of uncentered R^2 (UR2) statistic in all these cases: all five factors, each single factor, and each subset of four factors. The market factor is dominant in terms of explanatory power; a well-known result. The own-volatility factor is the strongest of the characteristic-based factors, followed by size, momentum, and value. The ordering of relative importance is the same whether we consider the factors singly or their marginal contribution given the other four.

We test for the statistical significance of each factor by calculating, for each cross-sectional regression, the t-statistic for each estimated coefficient, based on Hansen-White heteroskedasticity-consistent standard errors. Then for each factor we find the average number of cross-sectional regression t-statistics that are significant at a 95% confidence level across the 504 time periods. The resulting count statistic has an exact binomial distribution under the null hypothesis that the factor return is zero each period. Table 4a shows the annualized means and standard deviations of the factor returns, the percentage of significant t-statistics for each factor, and the aggregate p-value. All five factors are highly significant.

Table 4b displays the correlations of the estimated factors, along with the three Fama-French factors, RMRF, SMB, and HML. RMRF is the Fama-French market factor, it is the return to the value-weighted market index minus the risk free return; SMB is the return to a small capitalization portfolio minus the return to a large-capitalization portfolio; HML is the return to a high book-to-price portfolio minus the return to a low book-to-price portfolio. See Fama and French (1993) for detailed discussion of their portfolio formation rules. We also include a momentum factor created by Ken French; this is the return to a portfolio with high cumulative returns over the past twelve months minus the return to a portfolio with low cumulative returns over the past twelve months, adjusted to have roughly equal average capitalization; see Ken French's website[6] for details, where all the Fama-French data is freely available. Our factors and the analogous Fama-French factors are highly correlated (note that the size characteristic is defined inversely in the two models, hence the negative correlation). The Fama-French factors are based on capitalization-weighted portfolios whereas our factors are statistically generated, treating all assets equally. Since the cross-section of securities is dominated, in terms of the number of securities, by low-capitalization firms, this induces a strong positive correlation between our market factor and the Fama-French SMB factor. Our volatility factor has strong positive correlation with the market factor. This corroborates the finding in Ang et al. (2006b) Table 10, which shows high covariance between their idiosyncratic-volatility factor returns and the Fama-French market factor returns. It also seems theoretically consistent with the finding in Ang et al. (2006a) that the market factor return is negatively correlated with changes in

[6]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_mom_factor.html

VIX, a forward-looking index of market volatility. Essentially, the positive correlation between the own-volatility factor and market factor means that high own-volatility stocks outperform when the overall market rises and underperform when the overall market falls. There is also a strong negative correlation between the own-volatility and momentum factor returns, for which we have no ready explanation.

Figure 3 shows the characteristic-beta functions re-estimated on the four non-overlapping 126-month subintervals in the data set. The functions seem stable over time although we do not attempt a formal test.

5.5 Nonparametric Mispricing Functions

Next we add four additive nonparametric mispricing functions to the model, one for each of the characteristics. The estimation algorithm is essentially unchanged, except that after each iteration we impose the orthogonality and mean-zero identification conditions (6)-(5) on each mispricing function $\widehat{\alpha}_j$. We do this by regressing $\widehat{\alpha}_j$ on an intercept and $\widehat{g}_j$ across the set of grid points, and then replacing $\widehat{\alpha}_j$ with the residuals from this regression. The estimated component $\alpha-$functions can be interpreted as a nonparametric version of the tables of estimated alpha coefficients for characteristic-sorted portfolios shown in Fama and French (1993). Figure 4 shows the results. There is little evidence against the five-factor asset pricing model: all four mispricing functions differ only negligibly from zero, mostly lying in the interval from -20 to $+20$ basis points (in units of monthly return). An obvious feature of all four graphs is the absence of any monotonic pattern upward or downward. On more careful analysis it is clear that this is a consequence of the identification condition (6) rather than an empirical finding. Since the estimated characteristic-beta functions are monotonically increasing, the estimated characteristic-mispricing functions (which are forced to be orthogonal to them) have average slopes close to zero. The same lack-of-monotonicity is apparent in the empirical results of Fama and French(1993); see in particular Table 9a of that paper.

5.6 Time Series Dynamics and Trends

Table 5 shows the results from a first-order vector autoregression of the five factors returns on their lagged values. The size factor has the strongest autocorrelation and cross-correlation, as measured by its R-squared in the vector autoregression. This is to be expected, since this factor has the heaviest concentration in low-capitalization, illiquid securities where autocorrelation and cross-correlation is strongest; see, e.g., Lo and Mackinlay (1990). Table 6 shows the first-order vector autoregression using squared factor returns; this formulation is useful for identifying multivariate volatility dynamics in the factor returns. The squared value factor has the highest R^2 followed by size and own-volatility.

It is interesting that the squared market factor shows the lowest R^2. These results are only intended to be suggestive of the general pattern of dynamic volatility linkages between the factors; to complete the specification would require building a full multivariate volatility model which we do not attempt here, see, e.g., Laurent, Bauwens and Rombouts (2006) and references therein.

Campbell, Lettau, Malkiel and Xu (2001), in a model with industry and market factors, and Jones (2001) and Connor, Korajczyk and Linton (2006) in models with statistical factors, show that cross-sectional mean-square asset-specific return, $\gamma_t = n_t^{-1}\sum_{i=1}^{n_t}\varepsilon_{it}^2$, varies through time, with a strong upward trend over sample periods covering the 1950s - 1990's. Figure 5 replicates this finding in our model with characteristic-based factors. It is notable that the long upward trend in mean-square asset-specific return seems to reverse in the post-2000 sample period included here. We also examined each of the squared factor returns but found no evidence of similar trends (figures not shown, but available from the authors).

5.7 A Weighted Least-Squares Objective Function

Jones (2001) describes how to modify statistical factor estimation methods for the presence of time-varying mean-square asset-specific return as shown in Figure 5. Jones' adjustment is not strictly necessary with our method: although we use a least-squares-type objective function to motivate the estimators, we do not need to assume time-series or cross-sectional homogeneity of asset-specific variances to derive their asymptotic properties. Nonetheless, the evidence in Figure 5 points toward an improvement in the efficiency of the estimators by modifying the objective function (11) to account for the average time-series heteroskedasticity of asset-specific returns. The altered objective function is:

$$Q_T^*(f,g) = \frac{1}{T}\sum_{t=1}^{T}\gamma_t^{-1}E\left[\left\{y_{it} - f_{ut} - \sum_{j=1}^{J} g_j(X_{ji})f_{jt}\right\}^2\right]$$

If γ_t is the only source of heteroskedasticity in returns, then this adjustment allows us to attain Linton's (1997) oracle bound. Repeating the analysis with this altered objective function the only change to the estimation algorithm is in (15) which becomes:

$$g_j(x_j) = \frac{\frac{1}{T}\sum_{t=1}^{T}\gamma_t^{-1}\frac{1}{n_t}\sum_{i=1}^{n_t} f_{jt}\left(E\left[y_{it}|X_{jit}=x_j\right] - f_{ut} - \sum_{k\neq j} f_{kt}E\left[g_k(X_{kit})|X_{jit}=x_j\right]\right)}{\frac{1}{T}\sum_{t=1}^{T}\gamma_t^{-1}f_{jt}^2}$$

The re-estimated characteristic-beta functions (not shown, available from the authors) differ only very marginally from the original homogenously-time-weighted estimates.

5.8 The Case of an Observed Market Factor

Next we re-estimate the model using the market index from Fama and French as f_{ut}, and normalizing the characteristic-beta functions using capitalization-weighted means and variances. Figure 6 shows the characteristic-beta functions in this case. The only substantial change from the equally-weighted case is in the location of the characteristic-beta function for size. Now, most firms in the sample have a negative size betas rather than, as previously, a roughly equal split between positive and negative size betas. This reflects the fact that approximately 80% of the securities have size characteristics below the capitalization-weighted mean characteristic. The own-volatility characteristic-beta function has a steeper positive slope than in the equally-weighted case. Value and Momentum show less of a change from the equally-weighted case.

Table 6 shows tests of the embedded CAPM pricing (8). The well-known size and value premia relative to the CAPM are significant in the overall period, but inconsistent across subperiods. There is a significant negative premium relative to the CAPM for the own-volatility factor. In this five-factor model, the momentum factor on its own does not have an abnormal return premium relative to the CAPM.

6 Summary and Conclusion

Following the pioneering work of Rosenberg (1974), Fama and French (1993) and others, characteristic-based factor models have played a leading role in explaining the comovements of individual equity returns. This paper applies a new weighted additive nonparametric estimation procedure to estimate characteristic-based factor models more data-efficiently than existing nonparametric methods. The methodology we develop extends existing results to the large cross-section long time series setting. We obtain a variety of statistical results that are useful for conducting inference. We think this methodology can be useful elsewhere.

We estimate a characteristic-based factor model with five factors: a market factor, size factor, value factor, momentum factor and own-volatility factor. Although much of the existing literature has focused on the three-factor Fama-French model (market, size and value) we find that the momentum and own-volatility factors are at least, if not more, important than size and value in explaining return comovements. The univariate functions mapping characteristics to factor betas are monotonic but not linear, the deviation from linearity is particularly strong for size and value, less so for momentum and own-volatility.

Our methodology provides a new nonparametric test of multi-beta asset pricing theory. We estimate nonparametrically a set of additive mispricing functions based on the four security characteristics. We find little evidence against the five-factor asset pricing model.

We also examine the time-series behaviour of the estimated factor returns and their squared values using vector autoregressions. There are strong, multivariate dynamics in the volatilities of the factor returns.

We consider the case in which the market factor is observed (set equal to a capitalization-weighted index) and test an embedded CAPM pricing relationship against the more general five-factor pricing model. We reject the CAPM restriction relative to the five-factor alternative, due to abnormal return premia associated with the size, value and own-volatility factors.

A Appendix

A.1 Population Integral Equation

Here we study the properties of the population equations (14) in order to relate our procedure to Mammen, Linton, and Nielsen (1999) and Carrasco, Florens and Renault (2006). Define

$$m_j^T(x) = \frac{\sum_{t=1}^T f_{jt} E\left[(y_{it} - f_{ut})\,|X_{ji} = x\right]}{\sum_{t=1}^T f_{jt}^2}$$

$$\mathcal{H}_{jk}^T(x, x') = \rho_{jk}^T \times \frac{p_{k,j}(x, x')}{p_j(x) p_k(x')}, \text{ where } \rho_{jk}^T = \frac{\sum_{t=1}^T f_{jt} f_{kt}}{\sum_{t=1}^T f_{jt}^2}$$

and $p_{k,j}$ is the joint density of (X_{ji}, X_{ki}). Assumption A5 implies that ρ_{jk}^T are uniformly bounded, since

$$\left(\sum_{t=1}^T f_{jt} f_{kt}\right)^2 \leq \left(\sum_{t=1}^T f_{jt}^2\right)\left(\sum_{t=1}^T f_{kt}^2\right). \tag{35}$$

by the Cauchy Schwarz inequality. We drop the T superscript in $\mathcal{H}_{jk}^T$, m_j^T, and ρ_{jk}^T for the remainder of this subsection. The equations (14) can be rewritten as

$$\begin{pmatrix} I & \mathcal{H}_1 & \cdots & \mathcal{H}_1 \\ \mathcal{H}_2 & I & \mathcal{H}_2 & \cdots \\ \vdots & & \ddots & \\ \mathcal{H}_J & \cdots & & I \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_J \end{pmatrix} = \mathcal{H} g = m = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_J \end{pmatrix} \tag{36}$$

where I is the identity operator and $(\mathcal{H}_j g_k)(x) = \int \mathcal{H}_{jk}(x, x') g_k(x') p_k(x') dx'$. Note that (36) is a system of linear type 2 integral equations in the functions g_j for each T. It is very similar to the standard equations associated with additive nonparametric regression with two exceptions. First, in that case the intercept function is just an unweighted conditional expectation $E\left[y_i | X_{ji} = x\right]$. Second, the operator in that case does not have the weighting ρ_{jk}. The first difference is irrelevant for the studying of the existence and uniqueness of solutions, since only the operator is required for that. The second difference is rather minor since the weighting factors ρ_{jk} do not vary with the covariates. Suppose also that the Hilbert-Schmidt condition holds:

$$\int \frac{p_{k,j}(x,x')^2}{p_j(x)p_k(x')}dxdx' < \infty, \quad \text{for all } j,k. \tag{37}$$

This is satisfied under our assumption A2. Then let Ψ_j be the operator such that $\Psi_j f_j(x_j) = 0$ and $\Psi_j f_k(x_k) = f_k(x_k) - \int \mathcal{H}_{jk}(x_k,u)f_k(u)p_k(u)du$ for any functions f_j, f_k, and define $\mathcal{T} = \Psi_J \cdots \Psi_1$. This operator represents a population version of one cycle of the iteration (17). It can be shown from Lemma 1 of Mammen, Linton, and Nielsen (1999) that $\mathcal{T}$ is a positive self-adjoint linear operator with operator norm less than one, which implies that the population iterations converge to the solution of (36) at a geometric rate. Consider the special case of $J = 2$, in which case it suffices to show that the operator norm of $\mathcal{H}_1\mathcal{H}_2$ is strictly less than one. Consider the classical bivariate additive regression on covariates X_1, X_2, Mammen, Linton, and Nielsen (1999), and associated operators $\widetilde{\mathcal{H}}_1$ and $\widetilde{\mathcal{H}}_2$. Then $\mathcal{H}_1\mathcal{H}_2 = \rho_{12}\rho_{21}\widetilde{\mathcal{H}}_1\widetilde{\mathcal{H}}_2$, where

$$\rho_{12}\rho_{21} = \frac{\left(\sum_{t=1}^T f_{1t}f_{2t}\right)^2}{\sum_{t=1}^T f_{1t}^2 \sum_{t=1}^T f_{2t}^2} \leq 1,$$

by (35). It follows that the operator norm of $\mathcal{H}_1\mathcal{H}_2$ is strictly less than one if and only if the operator norm of $\widetilde{\mathcal{H}}_1\widetilde{\mathcal{H}}_2$ is strictly less than one. This is implied by the Hilbert-Schmidt condition (37).

A.2 Proof of Results

Proof of Theorem 1. The proof strategy is to first establish the properties of the initial consistent estimators and then to work with iterations from these starting points. As it turns out, this strategy obviates the need to establish the convergence of the algorithm and to deal with the integral equation (36) in any detail. Take the time averaged data (**??**) and estimate the functions $\overline{g}_j(.)$ by the smooth backfitting method and then renormalize. The properties of the resulting estimator $\widetilde{g}_j(x_j)$, $j = 1,\ldots,J$, are as in Mammen, Linton, and Nielsen (1999) except that the error term is $O_p(T^{-1/2})$. In particular, we have for an interior point x_j

$$\widetilde{g}_j(x_j) - g_j(x_j) = \frac{1}{np_j(x_j)}\sum_{i=1}^n K_h(x_j - X_{ji})\overline{\varepsilon}_i + h^2\beta_{n,j}(x_j) + \frac{1}{n}\sum_{i=1}^n s_n(X_i,x_j)\overline{\varepsilon}_i + R_{nj}(x_j), \tag{38}$$

where $\sup_{x_j}|R_{nj}(x_j)| = o_p(n^{-1/2}T^{-1/2})$, $\beta_{n,j}(x_j)$ is stochastically bounded with $\sup_{x_j}|\beta_{n,j}(x_j) - \beta_j(x_j)| = o_p(1)$, where $\beta_j(x_j)$ is a deterministic bounded continuous function. The function s_n is stochastically bounded (and depends only on $X_1,\ldots,X_n$) with $Es_n(X_i,x_j)^2 < \infty$. The error term $\overline{\varepsilon}_i$ has variance of order T^{-1}, which implies that the leading term in the expansion is $O_p(n^{-1/2}h^{-1/2}T^{-1/2})+O_p(h^2)$. Note that the remainder contains terms of the form $O_p(n^{-1/2}h^{-1/2}) \times O_p(n^{-1/2}h^{-1/2}T^{-1/2})$, which are $o_p(n^{-1/2}T^{-1/2})$ under our bandwidth conditions. It follows that $\widetilde{g}_j(x_j)$ is $\sqrt{nTh}$ consistent and asymptotically normal, and asymptotically independent of $\widetilde{g}_k(x_k)$. Note that the renormalization only affects the bias and not the variance due to the effect that integration has on variance.

The proof of our main result is given in the following lemmas. Lemma 1 and 2 give the pointwise performance of the initial factor estimator, denoted $\widetilde{f}_t$, while Lemmas 3 and 4 give the uniform (over t) performance of $\widetilde{f}_t$. Lemma 5 and 6 give the pointwise expansion of the update $\widetilde{g}_j^{[1]}(x_j)$ of $\widetilde{g}_j(x_j)$.

Consider the infeasible estimator $f_t^\dagger$ that is any solution of the system of linear equations $A^\dagger f_t^\dagger = b_t^\dagger$, where

$$b_t^\dagger = \frac{1}{n}\sum_{i=1}^n G(X_i)y_{it} \quad ; \quad A^\dagger = \frac{1}{n}\sum_{i=1}^n G(X_i)G(X_i)^\top.$$

This is just a standard OLS estimator with regressors one and $g_j(X_{ji})$. Let also $A = E[G(X_i)G(X_i)^\top]$.

LEMMA 1. *Under our assumptions, for any t*

$$\sqrt{n}(f_t^\dagger - f_t) \Longrightarrow N\left(0, V_{t,t}\right),$$

and $f_t^\dagger, f_s^\dagger$ are asymptotically independent for any $s \neq t$.

PROOF. We have

$$\sqrt{n}\begin{bmatrix} f_t^\dagger - f_t \\ f_s^\dagger - f_s \end{bmatrix} = (A^\dagger)^{-1}\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n G(X_i)\varepsilon_{it} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n G(X_i)\varepsilon_{is} \end{pmatrix}.$$

Then by the cross-sectional independence, $A^\dagger = A + o_p(1)$. Let

$$u_{i,ts}(c) = c^\top \begin{bmatrix} G(X_i)\varepsilon_{it} \\ G(X_i)\varepsilon_{is} \end{bmatrix}$$

for some vector c. Then by the Lindeberg CLT,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n u_{i,ts}(c) \Longrightarrow N(0, \sigma_{u,ts}^2(c)),$$

where $\sigma_{u,ts}^2(c) = \mathrm{var}[u_{i,ts}(c)] = c^\top E[\varepsilon_{it}^2 G(X_i)G(X_i)^\top]c < \infty$. The result then follows by the Cramer-Rao device, Slutsky theorem, and the fact that $\mathrm{cov}(\varepsilon_{it}, \varepsilon_{is}) = 0$. ■

Now consider the feasible factor estimator based on the time-averaged backfitting estimator $\widetilde{f}_t = \widetilde{A}^{-1}\widetilde{b}_t$, where

$$\widetilde{b}_t = \frac{1}{n}\sum_{i=1}^n \widetilde{G}(X_i)y_{it} \quad ; \quad \widetilde{A} = \frac{1}{n}\sum_{i=1}^n \widetilde{G}(X_i)\widetilde{G}(X_i)^\top,$$

where $\widetilde{G}(X_i) = [1, \widetilde{g}_1(X_{1i}), \ldots, \widetilde{g}_J(X_{Ji})]^\top$.

LEMMA 2. *Under our assumptions, for any t there is a stochastically bounded sequence $\delta_{n,t}$ such that*

$$\sqrt{n}(\widetilde{f}_t - f_t^\dagger - h^2\delta_{n,t}) = o_p(1).$$

PROOF. We use the matrix expansion $(I+\Delta)^{-1} = I - \Delta + (I+\Delta)^{-1}\Delta^2$ to obtain

$$\begin{aligned}
\widetilde{f}_t - f_t^\dagger &= \widetilde{A}^{-1}\widetilde{b}_t - A^{\dagger -1}b_t^\dagger \\
&= A^{\dagger -1}(\widetilde{b}_t - b_t^\dagger) - A^{\dagger -1}(\widetilde{A} - A^\dagger)A^{\dagger -1}b_t^\dagger \\
&\quad - A^{\dagger -1}(\widetilde{A} - A^\dagger)A^{\dagger -1}(\widetilde{b}_t - b_t^\dagger) \\
&\quad + A^{\dagger -1/2}\left[I + A^{\dagger -1/2}(\widetilde{A} - A^\dagger)A^{\dagger -1/2}\right]^{-1}\widetilde{A}^{-1/2}(\widetilde{A} - A^\dagger)A^{\dagger -1}(\widetilde{A} - A^\dagger)A^{\dagger -1}\widetilde{b}_t.
\end{aligned} \tag{39}$$

The error $||\widetilde{f}_t - f_t^\dagger||$ is majorized by the errors $||\widetilde{b}_t - b_t^\dagger||$ and $||\widetilde{A} - A^\dagger||$ times constants due to the invertibility of $A^\dagger$. For example,

$$\left\|A^{\dagger -1}(\widetilde{b}_t - b_t^\dagger)\right\| \leq \lambda_{\max}(A^{\dagger -1})\left\|\widetilde{b}_t - b_t^\dagger\right\| = \frac{1}{\lambda_{\min}(A^\dagger)}\left(\sum_{j=0}^{J}(\widetilde{b}_{jt} - b_{jt}^\dagger)^2\right)^{1/2},$$

$$\left\|A^{\dagger -1}(\widetilde{A} - A^\dagger)A^{\dagger -1}b_t^\dagger\right\| \leq \frac{\lambda_{\max}^{1/2}(((\widetilde{A} - A^\dagger)^\top(\widetilde{A} - A^\dagger))}{\lambda_{\min}^2(A^\dagger)}\left\|b_t^\dagger\right\| \leq \frac{\left(\sum_{j,k=0}^{J}(\widetilde{A}_{jk} - A_{jk}^\dagger)^2\right)^{1/2}}{\lambda_{\min}^2(A^\dagger)}\left\|b_t^\dagger\right\|,$$

where $\lambda_{\max}(.)$ and $\lambda_{\min}(.)$ denote the largest and smallest (respectively) eigenvalues of a square symmetric matrix. Furthermore, $\lambda_{\min}(A^\dagger) \geq \lambda_{\min}(A) - o_p(1)$, where by assumption, $\lambda_{\min}(A) > 0$. We establish the order in probability of the terms $\widetilde{b}_{jt} - b_{jt}^\dagger$ and $\widetilde{A}_{jk} - A_{jk}^\dagger$.

Consider the typical element in $\widetilde{b}_t - b_t^\dagger$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} y_{it}[\widetilde{g}_j(X_{ji}) - g_j(X_{ji})] &= \frac{1}{n}\sum_{i=1}^{n}\left[f_{ut} + \sum_{j=1}^{J} g_j(X_{ji})f_{jt}\right][\widetilde{g}_j(X_{ji}) - g_j(X_{ji})] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{it}\left[\widetilde{g}_j(X_{ji}) - g_j(X_{ji})\right] \\
&= T_{n1} + T_{n2}.
\end{aligned}$$

We consider the term T_{n1}. From (38), we have

$$\begin{aligned}
T_{n1} &= \frac{1}{n}\sum_{i=1}^{n}\left[f_{ut}+\sum_{j=1}^{J}g_j(X_{ji})f_{jt}\right][\widetilde{g}_j(X_{ji})-g_j(X_{ji})] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[f_{ut}+\sum_{j=1}^{J}g_j(X_{ji})f_{jt}\right]\frac{1}{np_j(X_{ji})}\sum_{i'=1}^{n}K_h\left(X_{ji}-X_{ji'}\right)\overline{\varepsilon}_{i'} \\
&\quad +h^2\frac{1}{n}\sum_{i=1}^{n}\left[f_{ut}+\sum_{j=1}^{J}g_j(X_{ji})f_{jt}\right]\beta_{n,j}(X_{ji}) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\left[f_{ut}+\sum_{j=1}^{J}g_j(X_{ji})f_{jt}\right]\frac{1}{n}\sum_{i'=1}^{n}s_n(X_{i'},x_j)\overline{\varepsilon}_{i'} \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\left[f_{ut}+\sum_{j=1}^{J}g_j(X_{ji})f_{jt}\right]R_{nj}(X_{ji}) \\
&= T_{n11}+T_{n12}+T_{n13}+T_{n14}.
\end{aligned}$$

The first term T_{n11} is a degenerate U-statistic, Powell, Stock, and Stoker (1989), i.e., $T_{n11}=\sum\sum_{i,i'}\varphi_{ni,i'}$ with

$$\varphi_{ni,i'}=\frac{1}{n^2}\left[f_{ut}+\sum_{j=1}^{J}g_j(X_{ji})f_{jt}\right]\frac{K_h\left(X_{ji}-X_{ji'}\right)}{p_j(X_{ji})}\overline{\varepsilon}_{i'},$$

where $E[\varphi_{ni,i'}|X_i,\overline{\varepsilon}_i]=0$. Therefore, we can write

$$T_{n11}=\sum_{i=1}^{n}\varphi_{ni,i}+\sum_{i'=1}^{n}E[\varphi_{ni,i'}|X_{i'},\overline{\varepsilon}_{i'}]+\sum\sum_{i\neq i'}\widetilde{\varphi}_{ni,i'},$$

where $\widetilde{\varphi}_{ni,i'}=\varphi_{ni,i'}-E[\varphi_{ni,i'}|X_{i'},\overline{\varepsilon}_{i'}]$ and by construction $E[\widetilde{\varphi}_{ni,i'}|X_i,\overline{\varepsilon}_i]=E[\widetilde{\varphi}_{ni,i'}|X_{i'},\overline{\varepsilon}_{i'}]=0$. By straightforward moment calculations it can be shown that $\sum_{i=1}^{n}\varphi_{ni,i}=O_p(n^{-3/2}h^{-1}T^{-1/2})=o_p((nT)^{-1/2})$. Furthermore, $\mathrm{var}(\widetilde{\varphi}_{ni,i'})=O(n^{-4}T^{-1}h^{-1})$ so that $\sum\sum_{i\neq i'}\widetilde{\varphi}_{ni,i'}=O_p(n^{-1}T^{-1/2}h^{-1/2})=o_p((nT)^{-1/2})$. Finally,

$$\sum_{i'=1}^{n}E[\varphi_{ni,i'}|X_{i'},\overline{\varepsilon}_{i'}]\simeq\frac{1}{n}\sum_{i=1}^{n}\overline{\varepsilon}_i\left[f_{ut}+\sum_{j'=1}^{J}E[g_{j'}(X_{j'i})|X_{ji}]f_{j't}\right]=O_p((nT)^{-1/2}).$$

Furthermore, $T_{n12}=O_p(h^2)$ and $T_{n13},T_{n14}=o_p((nT)^{-1/2})$. Therefore, $\widetilde{b}_t-b_t^{\dagger}=O_p((nT)^{-1/2})+O_p(h^2)$.

Likewise the typical element in $\widetilde{A}-A^{\dagger}$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}[\widetilde{g}_j(X_{ji})\widetilde{g}_k(X_{ki})-g_j(X_{ji})g_k(X_{ki})]=O_p(h^2)+O_p((nT)^{-1/2}).$$

It follows that provided $nh^4\to 0$, $\sqrt{n}(\widetilde{f}_t-f_t^{\dagger})=o_p(1)$. More generally, letting

$$\delta_{n,t}=\frac{1}{n}\sum_{i=1}^{n}\left[f_{ut}+\sum_{j=1}^{J}g_j(X_{ji})f_{jt}\right]\beta_{n,j}(X_{ji})$$

we have $\sqrt{n}(\widetilde{f}_t - f_t^\dagger - h^2\delta_{n,t}) = o_p(1)$. ■

We now turn to the uniform over t properties, (28). By the triangle inequality

$$\max_{1\le t\le T}\left\|\widetilde{f}_t - f_t\right\| \le \max_{1\le t\le T}\left\|\widetilde{f}_t - f_t^\dagger\right\| + \max_{1\le t\le T}\left\|f_t^\dagger - f_t\right\|.$$

We first examine $\max_{1\le t\le T}\left\|f_t^\dagger - f_t\right\|$.

Lemma 3. *Under our assumptions*

$$\max_{1\le t\le T}\left\|f_t^\dagger - f_t\right\| = O_p(n^{-1/2}(\log T)^\rho).$$

Proof. By the above arguments, there is a finite constant C such that

$$\max_{1\le t\le T}\left\|f_t^\dagger - f_t\right\| \le (C + o_p(1))\max_{j=u,1,\ldots,J}\max_{1\le t\le T}\left|\frac{1}{n}\sum_{i=1}^n \varepsilon_{it}g_j(X_{ji})\right|.$$

So it suffices to show that

$$\max_{1\le t\le T}\left|\frac{1}{n}\sum_{i=1}^n \varepsilon_{it}g_j(X_{ji})\right| = O_p(n^{-1/2}(\log T)^\rho) \tag{40}$$

for some $\rho > 0$. Let $\varepsilon_{it}^+ = \varepsilon_{it}1(|\varepsilon_{it}| \le (nT)^{1/\kappa}) - E[\varepsilon_{it}1(|\varepsilon_{it}| \le (nT)^{1/\kappa})]$. Then, $1 - \Pr[|\varepsilon_{it}| \le (nT)^{1/\kappa}$ for $1 \le t \le T$ and $1 \le i \le n] \le nT\Pr[|\varepsilon_{it}| > (nT)^{1/\kappa}] \le E[|\varepsilon_{it}|^\kappa 1(|\varepsilon_{it}| > (nT)^{1/\kappa})] \to 0$. We now apply the Bonferroni and exponential inequalities to $\max_{1\le t\le T}|\frac{1}{n}\sum_{i=1}^n \varepsilon_{it}^+ g_j(X_{ji})|$. In particular, letting $\tau_{nT}^2 = \inf_{1\le t\le T}\operatorname{var}[\sum_{i=1}^n \varepsilon_{it}^+ g_j(X_{ji})]$, we have

$$\begin{aligned}
\Pr\left[\max_{1\le t\le T}\left|\sum_{i=1}^n \varepsilon_{it}^+ g_j(X_{ji})\right| > Kn^{1/2}\right] &\le \sum_{t=1}^T \Pr\left[\left|\sum_{i=1}^n \varepsilon_{it}^+ g_j(X_{ji})\right| > Kn^{1/2}\right] \\
&\le 2T\exp\left(-\frac{nK^2}{2\tau_{nT}^2 + 2(nT)^{1/\kappa}n^{1/2}K/3}\right).
\end{aligned}$$

By taking $K = (\log T)^\rho$ the right hand side is $o(1)$ provided $\kappa > 4$. ■

Lemma 4. *Under our assumptions*

$$\max_{1\le t\le T}\left\|\widetilde{f}_t - f_t^\dagger\right\| = O_p((nT)^{-1/2}(\log T)^\rho) + O_p(h^2(\log T)^{\rho'}).$$

Proof. As before we apply the triangle inequality again to each term in (39). We have $\max_{1\le t\le T}||b_t^\dagger|| = O_p((\log T)^{\rho\prime})$, so it suffices to bound the terms $\max_{1\le t\le T}|\widetilde{b}_{jt} - b_{jt}^\dagger|$ and $\max_{1\le t\le T}|\widetilde{A}_{jk} - A_{jk}^\dagger|$. We just show that

$$\max_{1\le t\le T}\left|\frac{1}{n}\sum_{i=1}^n\left[f_{ut} + \sum_{j=1}^J g_j(X_{ji})f_{jt}\right][\widetilde{g}_j(X_{ji}) - g_j(X_{ji})]\right| = O_p(n^{-1/2}(\log T)^\rho) + O_p(h^2(\log T)^{\rho\prime}) \tag{41}$$

$$\max_{1\le t\le T}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{it}[\widetilde{g}_j(X_{ji})-g_j(X_{ji})]\right| = O_p(n^{-1/2}(\log T)^{\rho}). \tag{42}$$

This uses the same type of techniques as above. In particular, we have

$$\begin{aligned}
&\max_{1\le t\le T}\left|h^2\frac{1}{n}\sum_{i=1}^{n}\left[f_{ut}+\sum_{j=1}^{J}g_j(X_{ji})f_{jt}\right]\beta_j(X_{ji})\right| \\
\le\ & h^2\left(\max_{1\le t\le T}|f_{ut}|\frac{1}{n}\sum_{i=1}^{n}|\beta_j(X_{ji})|+\sum_{j=1}^{J}\max_{1\le t\le T}|f_{jt}|\frac{1}{n}\sum_{i=1}^{n}|g_j(X_{ji})||\beta_j(X_{ji})|\right) \\
=\ & O_p(h^2(\log T)^{\rho'}).
\end{aligned}$$

Furthermore,

$$\max_{1\le t\le T}\left|\frac{1}{n}\sum_{i=1}^{n}\overline{\varepsilon}_i\left[f_{ut}+\sum_{j'=1}^{J}E[g_{j'}(X_{j'i})|X_{ji}]f_{j't}\right]\right| = O_p((nT)^{-1/2}(\log T)^{\rho'}).$$

In conclusion we have shown $\max_{1\le t\le T}||\widetilde{f}_t-f_t|| = O_p((nT)^{-1/2}(\log T)^{\rho})+O_p(h^2(\log T)^{\rho'})$. ■

Finally, we establish the asymptotic distribution of $\widehat{g}_j(x_j)$. Consider the one-step estimator

$$\widehat{g}_j^{[1]}(x_j)=\frac{\sum_{t=1}^{T}\widetilde{f}_{jt}\left[\widehat{\lambda}_{1t}(j,x_j)-\widetilde{f}_{ut}-\sum_{k\neq j}\widetilde{f}_{kt}\widetilde{\lambda}_2(j,k,x_j)\right]}{\sum_{t=1}^{T}\widetilde{f}_{jt}^2},$$

where

$$\widetilde{\lambda}_2(j,k,x_j)=\frac{\sum_{i=1}^{n}K_h(X_{ji}-x_j)\widetilde{g}_k(X_{ki})}{\sum_{i=1}^{n}K_h(X_{ji}-x_j)}.$$

Define

$$\widetilde{g}_j^{[1]}(x_j)=\frac{\sum_{t=1}^{T}f_{jt}\left[\widehat{\lambda}_{1t}(j,x_j)-f_{ut}-\sum_{k\neq j}f_{kt}\widetilde{\lambda}_2(j,k,x_j)\right]}{\sum_{t=1}^{T}f_{jt}^2}.$$

Lemma 5. *Under our assumptions*

$$\widehat{g}_j^{[1]}(x_j)-g_j(x_j)=\widetilde{g}_j^{[1]}(x_j)-g_j(x_j)+O_p((nT)^{-1/2})+O_p(h^2).$$

Proof. We expand out $\widehat{g}_j^{[1]}(x_j)$ about $\widetilde{g}_j^{[1]}(x_j)$ in a Taylor expansion in $\widetilde{f}_{jt}-f_{jt}$ and $\widetilde{g}_k(X_{ki})-g_k(X_{ki})$ obtaining many terms. A typical term is

$$\frac{\sum_{t=1}^{T}\left(\widetilde{f}_{jt}-f_{jt}\right)f_{jt}g_j(x_j)}{\sum_{t=1}^{T}f_{jt}^2}.$$

Then

$$\frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{f}_{jt}-f_{jt}\right)f_{jt}=\frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{f}_{jt}-f_{jt}^{\dagger}\right)f_{jt}+\frac{1}{T}\sum_{t=1}^{T}\left(f_{jt}^{\dagger}-f_{jt}\right)f_{jt}, \tag{43}$$

where

$$\frac{1}{T}\sum_{t=1}^{T}\left(f_{jt}^{\dagger}-f_{jt}\right)f_{jt}=(A^{\dagger})^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}G(X_i)\varepsilon_{it}\left(f_{jt}^{\dagger}-f_{jt}\right)f_{jt}=O_p((nT)^{-1/2}).$$

The expansion for $T^{-1}\sum_{t=1}^{T}(\widetilde{f}_{jt}-f_{jt}^{\dagger})f_{jt}$ is more complicated but basically one obtains terms like

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}f_{jt}\frac{1}{n}\sum_{i=1}^{n}y_{it}[\widetilde{g}_j(X_{ji})-g_j(X_{ji})]\\
=\;&\frac{1}{T}\sum_{t=1}^{T}f_{jt}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{it}[\widetilde{g}_j(X_{ji})-g_j(X_{ji})]\\
&+\frac{1}{T}\sum_{t=1}^{T}f_{jt}\frac{1}{n}\sum_{i=1}^{n}\left[f_{ut}+\sum_{j=1}^{J}g_j(X_{ji})f_{jt}\right][\widetilde{g}_j(X_{ji})-g_j(X_{ji})].
\end{aligned}$$

The double averaging makes the stochastic terms $O_p((nT)^{-1/2})$. The bias terms are always $O_p(h^2)$. ■

Define

$$\widetilde{U}_{n1}=\frac{1}{p_j(x_j)}\frac{1}{n}\sum_{i=1}^{n}K_h(x_j-X_{ji})\widetilde{\varepsilon}_i, \tag{44}$$

$$\widetilde{\varepsilon}_i=\frac{\sum_{t=1}^{T}f_{jt}\varepsilon_{it}}{\sum_{t=1}^{T}f_{jt}^2}.$$

Lemma 6. *Under our assumptions*

$$\widetilde{g}_j^{[1]}(x_j)-g_j(x_j)=\widetilde{U}_{n1}+O_p(h^2)+o_p(n^{-1/2}T^{-1/2}h^{-1/2}).$$

Proof. Consider

$$\begin{aligned}
&\frac{\sum_{t=1}^{T}f_{jt}\left[\widehat{\lambda}_{1t}(j,x_j)-f_{ut}-\sum_{k\neq j}f_{kt}\widetilde{\lambda}_2(j,k,x_j)\right]}{\sum_{t=1}^{T}f_{jt}^2}-g_j(x_j)\\
=\;&\frac{1}{\sum_{t=1}^{T}f_{jt}^2}\sum_{t=1}^{T}f_{jt}\left\{\frac{\sum_{i=1}^{n}K_h(X_{ji}-x_j)\left[y_{it}-f_{ut}-\sum_{k\neq j}f_{kt}\widetilde{g}_k(X_{ki})\right]}{\sum_{i=1}^{n}K_h(X_{ji}-x_j)}-f_{jt}g_j(x_j)\right\}\\
=\;&\frac{1}{\sum_{t=1}^{T}f_{jt}^2}\sum_{t=1}^{T}f_{jt}\left\{\frac{\sum_{i=1}^{n}K_h(X_{ji}-x_j)\varepsilon_{it}}{\sum_{i=1}^{n}K_h(X_{ji}-x_j)}+\frac{\sum_{i=1}^{n}K_h(X_{ji}-x_j)\left[g_j(X_{ji})-g_j(x_j)\right]}{\sum_{i=1}^{n}K_h(X_{ji}-x_j)}\right\}\\
&-\frac{1}{\sum_{t=1}^{T}f_{jt}^2}\sum_{t=1}^{T}f_{jt}\frac{\sum_{i=1}^{n}K_h(X_{ji}-x_j)\sum_{k\neq j}f_{kt}\left[\widetilde{g}_k(X_{ki})-g_k(X_{ki})\right]}{\sum_{i=1}^{n}K_h(X_{ji}-x_j)}\\
=\;&U_{n1}+U_{n2}+U_{n3}.
\end{aligned}$$

The term U_{n2} is a standard bias term of order h^2. The term U_{n3} can be shown to be $O_p(h^2) + o_p(n^{-1/2}T^{-1/2}h^{-1/2})$ as in Linton (1997), where the $O_p(h^2)$ is a bias term. Interchanging summations and approximating $\sum_{i=1}^n K_h(X_{ji} - x_j)/n$ by $p_j(x_j)$ we obtain an approximation to the leading term U_{n1}, i.e., $U_{n1} = \widetilde{U}_{n1}(1 + O_p(n^{-1/2}h^{-1/2}))$. The term $\widetilde{U}_{n1}$ is a sum of independent random variables and is $O_p(n^{-1/2}h^{-1/2}T^{-1/2})$ with mean zero and variance as stated in the theorem. Specifically, $E[\widetilde{\varepsilon}_i | X_{ji} = x_j] = 0$ and

$$\mathrm{var}[\widetilde{\varepsilon}_i | X_{ji} = x_j] = \frac{\sum_{t=1}^T f_{jt}^2 \sigma_{jt}^2(x_j)}{\left(\sum_{t=1}^T f_{jt}^2\right)^2} \leq \frac{C}{T}$$

for some $C < \infty$ for large enough T. Therefore, we can apply the Lindeberg's CLT. Furthermore, $\widetilde{U}_{n1j}(x_j)$ and $\widetilde{U}_{n1k}(x_k)$ are asymptotically independent by standard arguments for kernels.

■

Now define $\widehat{f}_t^{[1]}$ as in (19). Using the above expansion it can be shown that the results of Lemma 2 and 4 continue to hold with $\widehat{f}_t^{[1]}$ replacing $\widetilde{f}_t$ with a difference sequence $\delta_{n,t}$. Then we can show that the conclusion of Lemma 5 and 6 continue to hold with $\widehat{g}_j^{[2]}$ replacing $\widehat{g}_j^{[1]}$. This process can be continued for any finite number of iterations, see Linton, Nielsen, and Van der Geer (2004). The only thing that changes in the expansions is the bias function, although it can still be approximated by some bounded continuous function.

■

Proof of Theorem 2. By the triangle inequality

$$\sup_{\theta\in\Theta} \left\| \widehat{M}_T(\theta) - M(\theta) \right\| \leq \sup_{\theta\in\Theta} \left\| \widehat{M}_T(\theta) - M_T(\theta) \right\| + \sup_{\theta\in\Theta} \left\| M_T(\theta) - M(\theta) \right\|,$$

where $M_T(\theta) = T^{-1}\sum_{t=k+1}^T \psi(F_t, Z_t; \theta)$. Applying uniform laws of large numbers for classes of smooth functions (Andrews (1987)) we obtain that $\sup_{\theta\in\Theta} \|M_T(\theta) - M(\theta)\| = o_p(1)$. Furthermore, by a first order expansion for each $\ell = 1, \ldots, q$,

$$\widehat{M}_{T\ell}(\theta) - M_{T\ell}(\theta) = \frac{1}{T}\sum_{t=k+1}^T \frac{\partial\psi_\ell}{\partial F_t}(\overline{F}_t, Z_t; \theta)(\widehat{F}_t - F_t),$$

where $\overline{F}_t$ are intermediate values. We next use the standard inequality $\Pr[C] \leq \Pr[C \cap D] + \Pr[D^c]$ with $D = \{\max_{1\leq t\leq T} |\widehat{f}_t - f_t| \leq \delta_{n,T}\}$ and $\delta_{n,T} \to 0$ chosen such that $\Pr[D^c] \to 0$; this allows us to restrict attention to the event D. It follows that on this set by crude bounding using assumption B3 and (28) we have for some $C < \infty$,

$$\begin{aligned}
&\left\| \frac{1}{T} \sum_{t=k+1}^{T} \frac{\partial \psi_\ell}{\partial F_t}(\overline{F}_t, Z_t; \theta)(\widehat{F}_t - F_t) \right\| \\
\leq\ & C \left(\max_{1 \leq t \leq T} \left\| \widehat{f}_t - f_t \right\| \right) \frac{1}{T} \sum_{t=k+1}^{T} \sup_{||\theta - \theta_0|| \leq \delta, ||F' - F_t|| \leq \delta_{n,T}} \left\| \frac{\partial \psi_\ell}{\partial F_t}(F', Z_t; \theta) \right\| \\
=\ & o_p(1).
\end{aligned}$$

It follows that $\sup_{\theta \in \Theta} ||\widehat{M}_T(\theta) - M_T(\theta)|| = o_p(1)$. Finally, the unique minimum condition B2 implies that $\widehat{\theta} - \theta_0 = o_p(1)$.

By a Taylor expansion

$$\widehat{M}_T(\widehat{\theta}) = \widehat{M}_T(\theta_0) + \frac{\partial \widehat{M}_T}{\partial \theta}(\overline{\theta})(\widehat{\theta} - \theta_0), \tag{45}$$

where $\overline{\theta}$ are intermediate values. Furthermore, for each $\ell = 1, \ldots, q$,

$$\begin{aligned}
\widehat{M}_{T\ell}(\theta_0) &= M_{T\ell}(\theta_0) + \frac{1}{T} \sum_{t=k+1}^{T} \frac{\partial \psi_\ell}{\partial F_t}(F_t, Z_t; \theta_0)(\widehat{F}_t - F_t) \\
&\quad + \frac{1}{2T} \sum_{t=k+1}^{T} (\widehat{F}_t - F_t)^\top \frac{\partial^2 \psi_\ell}{\partial F_t \partial F_t^\top}(\overline{F}_t, Z_t; \theta_0)(\widehat{F}_t - F_t),
\end{aligned}$$

where $\overline{F}_t$ are intermediate values. By substituting in the expansion for $\widehat{F}_t - F_t$ it is easy to see that

$$\frac{1}{T} \sum_{t=k+1}^{T} \frac{\partial \psi_\ell}{\partial F_t}(F_t, Z_t; \theta_0)(\widehat{F}_t - F_t) = O_p((nT)^{-1/2}) + O_p(h^2). \tag{46}$$

We next use the standard inequality $\Pr[C] \leq \Pr[C \cap D] + \Pr[D^c]$ with $D = \{\max_{1 \leq t \leq T} |\widehat{f}_t - f_t| \leq \delta_{n,T} n^{-1/2} (\log T)^\rho\}$ and $\delta_{n,T} \to 0$ chosen such that $\Pr[D^c] \to 0$; this allows us to restrict attention to the event D. It follows that on this set by crude bounding using assumption B3 we have for some $C < \infty$,

$$\begin{aligned}
&\left\| \frac{1}{2T} \sum_{t=k+1}^{T} (\widehat{F}_t - F_t)^\top \frac{\partial^2 \psi_\ell}{\partial F_t \partial F_t^\top}(\overline{F}_t, Z_t; \theta_0)(\widehat{F}_t - F_t) \right\| \\
\leq\ & C \left(\max_{1 \leq t \leq T} \left\| \widehat{f}_t - f_t \right\| \right)^2 \frac{1}{T} \sum_{t=k+1}^{T} \sup_{||\theta - \theta_0|| \leq \delta_{n,T}, ||F' - F_t|| \leq \delta_{n,T}} \left\| \frac{\partial^2 \psi_\ell}{\partial F_t \partial F_t^\top}(F', Z_t; \theta) \right\| \\
=\ & O_p(n^{-1} (\log T)^{2\rho}) = o_p(T^{-1/2}).
\end{aligned}$$

Therefore, $\widehat{M}_T(\theta_0) = M_T(\theta_0) + o_p(T^{-1/2})$. Similarly

$$\frac{\partial \widehat{M}_T}{\partial \theta}(\overline{\theta}) = \frac{\partial M_T}{\partial \theta}(\theta_0) + o_p(1) = \Gamma_0 + o_p(1).$$

In conclusion, we have $o_p(T^{-1/2}) = M_T(\theta_0) + \Gamma_0(\widehat{\theta} - \theta_0)$, and the result follows from arguments of Pakes and Pollard (1989, pp 1041-1042). In particular, a CLT for stationary mixing random variables is applied to $\sqrt{T}M_T(\theta_0)$, whence $\sqrt{T}(\widehat{\theta} - \theta_0)$ is asymptotically normal as stated. ■

References

[1] Andrews, D. W. K., (1987), Consistency in nonlinear econometric models: A generic uniform law of large numbers," *Econometrica*, 55, 1465-1471.

[2] Ang, A., R. Hodrick, Y. Xing and X. Zhang, 2006a, The cross-section of volatility and expected returns, Journal of Finance, 61, 259-299.

[3] Ang, A., R. Hodrick, Y. Xing and X. Zhang, 2006b, High idiosyncratic risk and low returns: international and further U.S. evidence, working paper, Columbia Business School, Columbia University.

[4] Bai, J. 2004. Inferential theory for factor models of large dimension. *Econometrica*

[5] Bai, J. 2005. Estimating cross-section common stochastic trends in nonstationary panel data. *Econometrica.*

[6] Bai, J., and S. Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* **70**: 191-221.

[7] Baltagi, B.H., J. Hidalgo, and Q. Li (1996). A nonparametric test for poolability using panel data. Journal of Econometrics 75, 345-367.

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

[9] 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.

[10] 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).

[11] Borak, S., W. Hærdle, E. Mammen, and B.U. Park (2007). Time series modelling with semiparametric factor dynamics.

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

[13] Campbell, J., M. Lettau, B. Malkiel and Y. Xu, Have individual stocks become more volatile?, Journal of finance, 56, 1-43.

[14] Carrasco, M., J.P. Florens, and E. Renault (2002), Linear inverse problems in structural econometrics," forthcoming in Handbook of Econometrics, volume 6, eds. J.J. Heckman and E. Leamer.

[15] 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.

[16] 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.

[17] Connor, G, R.A. Korjaczyk and O.B. Linton, 2006, The common and specific components of dynamic volatility, 132, 231-255.

[18] Connor, G. and O.B. Linton, 2007, Semiparametric estimation of a characteristic-based factor model of stock returns, Journal of Empirical Finance, forthcoming.

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

[20] 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.

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

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

[23] Fan, J., and Q. Yao (1998): Efficient estimation of conditional variance functions in Stochastic Regression, Biometrika 85, 645-660.

[24] Gibbbon, M. S. Ross and J. Shanken (1989). A test of the efficiency of a given portfolio, Econometrica 57, 1121-1152.

[25] Goyal, A. and P. Santa-Clara (2003), Idiosyncratic risk matters!, Journal of finance, 58, 975-1007.

[26] Gozalo, P.L. and O.B. Linton (2001). Testing additivity in generalized nonparametric regression models with estimated parameters. Journal of Econometrics 104, 1-48.

[27] Haag, B.R. (2007). Nonparametric regression rests based on additive model estimation. Manuscript, Mannheim University.

[28] Hansen, L. H., B. Nielsen and J. P. Nielsen (2004). Two sided analysis of variance with a latent time series. Nuffield College, Oxford University 2004-W25.

[29] Horowitz, J.L., J. Klemelä, and E. Mammen (2002): Optimal estimation in additive regression," Manuscript, Heidelberg University.

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

[31] Jagadeesh, N. and S. Titman (1993) Returns to buying winners and selling losers: implications of stock market efficiency, Journal of Finance, 48, 65-92.

[32] Jones, C.S. (2001) Extracting factors from heteroskedastic asset returns, Journal of financial economics, 62, 293-325.

[33] Li, Q., and J.S. Racine (2007). Nonparametric Econometrics. Princeton University Press, Princeton.

[34] Linton, O.B. (1997). Efficient estimation of additive nonparametric regression models. Biometrika, 84, 469-474.

[35] Linton, O.B., and E. Mammen (2005): Estimating semiparametric ARCH(∞) models by kernel smoothing methods, Econometrica 73, 771-836.

[36] Linton, O.B., and E. Mammen (2006): Nonparametric Transformation to White Noise. Forthcoming in Journal of Econometrics.

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

[38] Linton, O.B., J.P. Nielsen and S. van de Geer (2004). Estimating Multiplicative and Additive Hazard Functions by Kernel Methods. *The Annals of Statistics* 31, 2, 464-492.

[39] Lo, A. and A.C. MacKinlay, 1990, When are contrarian profits due to stock market overreaction?, Review of financial studies, 3, 175-205.

[40] Mammen, E., O. Linton, and Nielsen, J. P. (1999): The existence and asymptotic properties of a backfitting projection algorithm under weak conditions, *Annals of Statistics* 27, 1443-1490.

[41] Mammen, E., and B. Park (2005): Bandwidth selection for smooth backfitting in additive models, *Annals of Statistics* 33, 1260-1294.

[42] Mammen, E., B. Støve, and D. Tjøstheim (2006): Nonparametric Additive Models for Panels of Time Series, manuscript.

[43] Opsomer, J. D. and D. Ruppert (1997): Fitting a bivariate additive model by local polynomial regression, *Annals of Statistics* 25, 186 - 211.

[44] Pakes, A., and D. Pollard (1989): Simulation and the asymptotics of optimization estimators, *Econometrica* 57, 1027-1057.

[45] Pesaran, M.H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure *Econometrica* 74, 967-1012.

[46] Phillips. P.C.B., and H.R. Moon. 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* **67**: 1057-1113.

[47] Porter, J. (1996). Nonparametric Regression Estimation for a Flexible Panel Data Model. PhD thesis, Department of Economics, MIT.

[48] Powell, J. L., J. H. Stock, and T. M. Stoker (1989): Semiparametric estimation of index coefficients." *Econometrica*, 57, 1403-1430.

[49] Robinson, P.M. (1983). Nonparametric Estimators for Time Series. Journal of Time Series Analysis 4, 185-207.

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

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

[52] Rosenblatt, M. (1956). A central limit theorem and strong mixing conditions, *Proc. Nat. Acad. Sci.* **4**, 43-47.

[53] Ross, S.A (1976) The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341-360.

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

[55] Stock, J., and M. Watson (1998). Diffusion indexes. NBER working paper no 6702.

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

Table 1: Sample Statistics of Raw Security Characteristics

		Mean				Standard Deviation				Skewness				Excess kurtosis			
Year	Firms	Size	Value	Mom	Vol	Size	Value	Mom	Vol	Size	Value	Mom	Vol	Size	Value	Mom	Vol
1965	467	4.84	0.69	0.01	0.06	1.69	0.40	0.02	0.03	-0.25	1.04	0.18	1.21	-0.14	0.47	0.59	1.04
1970	1562	4.62	0.52	0.00	0.10	1.56	0.29	0.03	0.04	0.23	0.82	0.23	0.74	-0.47	0.19	0.46	0.09
1975	3394	3.44	1.40	-0.01	0.14	1.81	0.90	0.03	0.06	0.47	0.91	0.03	0.59	-0.38	0.27	0.19	-0.14
1980	3465	3.72	1.21	0.02	0.12	1.86	0.69	0.03	0.06	0.20	0.79	0.44	0.69	-0.38	0.15	0.40	0.00
1985	4017	4.20	0.73	-0.02	0.10	1.95	0.43	0.03	0.05	0.17	0.69	-0.37	0.76	-0.42	-0.04	0.22	0.06
1990	4595	4.13	0.76	0.01	0.11	2.18	0.48	0.03	0.06	0.24	0.91	-0.02	1.00	-0.33	0.47	0.59	0.39
1995	5583	4.78	0.62	0.01	0.11	2.00	0.40	0.03	0.06	0.37	0.91	-0.07	0.94	-0.31	0.28	0.57	0.33
2000	6003	5.25	0.65	0.01	0.17	2.12	0.46	0.04	0.09	0.38	1.01	0.55	0.90	-0.27	0.47	0.73	0.31
2005	4952	6.06	0.52	0.03	0.11	2.03	0.31	0.03	0.07	0.23	0.85	0.58	1.08	-0.28	0.39	0.66	0.42
Med	4017	4.62	0.69	0.01	0.11	1.95	0.43	0.03	0.06	0.23	0.91	0.18	0.90	-0.33	0.28	0.57	0.31

Table 1 shows some descriptive statistics of the cross-sectional data for July at five-year intervals: the number of securities in the annual cross section, and the first four cross-sectional moments of the four raw characteristics. Separately provided are the time series medians over the full 42 year period, also using July data.

Figure 1: Time Series Plots of Cross-sectional Correlations between the Characteristics

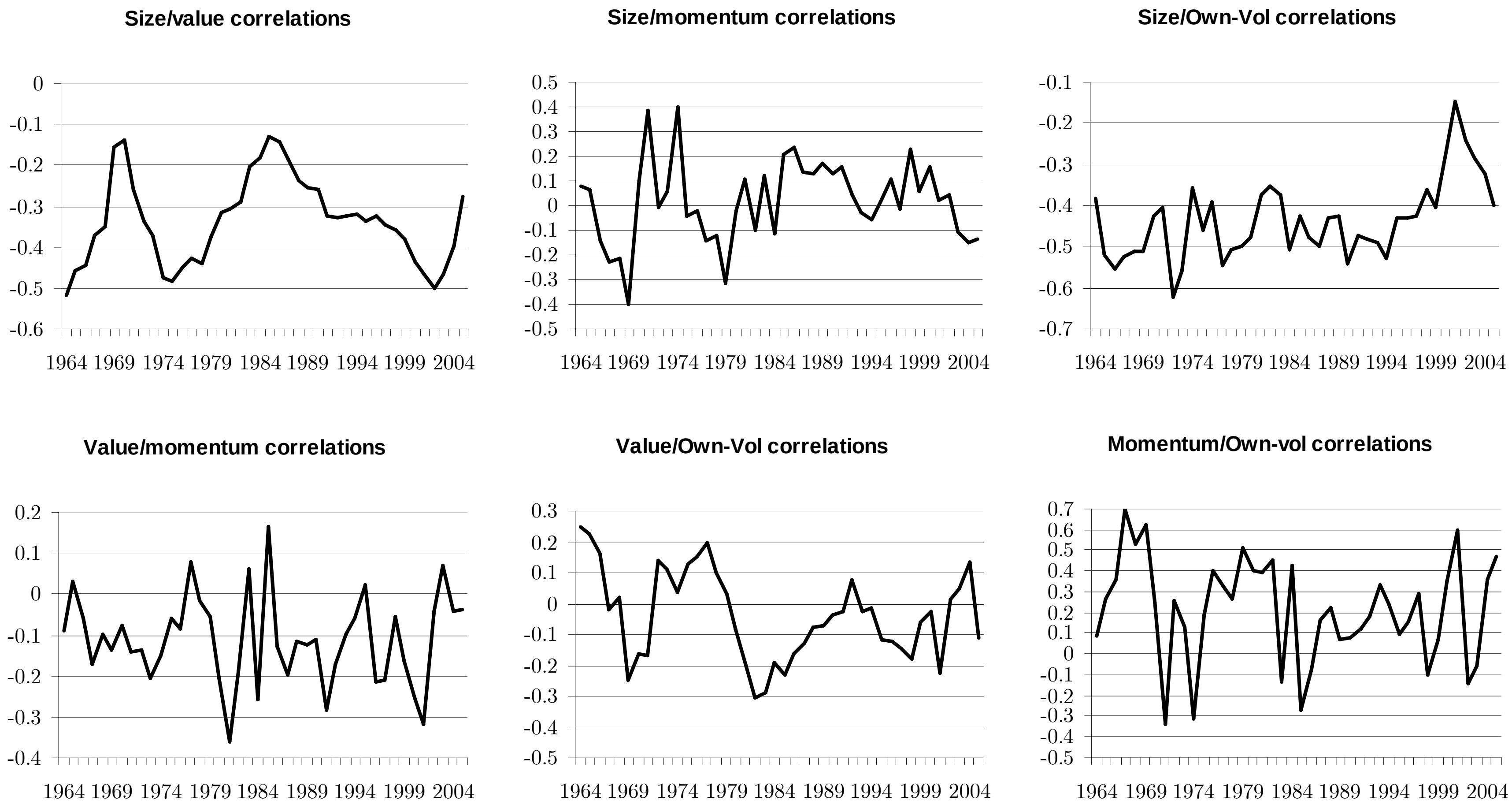


Figure 1 illustrates the time varying nature of the correlations between characteristics by showing the cross-sectional correlation for each pair of characteristics each July.

$\frac{1}{\sqrt{n_t}}$

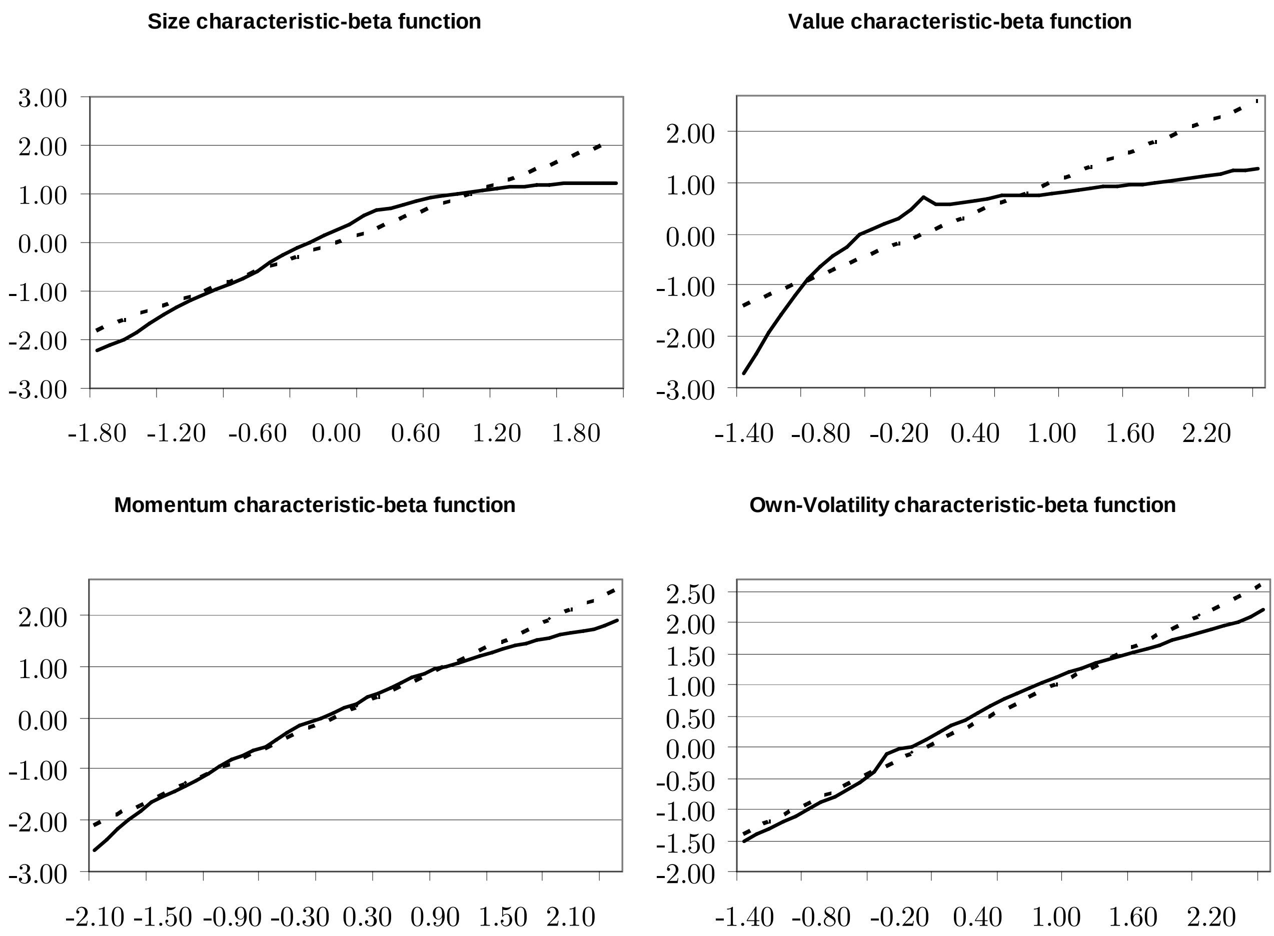
Table 2: Characteristic-Beta Functions and Standard Errors

	Size		Value		Momentum		Own-Volatility	
Grid	Value	SE	Value	SE	Value	SE	Value	SE
-2.00	n.c	n.c	n.c	n.c	-2.39	0.048	n.c	n.c
-1.50	-1.86	0.036	n.c	n.c	-1.53	0.046	n.c	n.c
-1.00	-1.07	0.030	-1.21	0.074	-0.94	0.034	-1.12	0.016
-0.50	-0.42	0.028	-0.01	0.059	-0.42	0.030	-0.56	0.024
0.00	0.28	0.025	0.71	0.067	0.10	0.026	0.13	0.034
0.50	0.78	0.022	0.70	0.059	0.58	0.025	0.67	0.041
1.00	1.04	0.019	0.79	0.058	1.01	0.029	1.11	0.044
1.50	1.19	0.016	0.93	0.063	1.34	0.032	1.47	0.047
2.00	1.22	0.014	1.08	0.059	1.61	0.035	1.78	0.045
2.50	n.c	0.012	1.25	0.089	1.89	0.063	2.10	0.071

n.c = not computed

Table 2 shows the estimates of the characteristic-beta functions at some selected characteristic values and the heteroskedasticity-consistent standard errors for each of these estimates. The table reports results for each characteristic over a support ranging from the empirical 2.5% to the 97.5% quantile.

Figure 2: The Characteristic-beta Functions



The solid lines in Figure 2 display the estimated characteristic-beta functions over all grid points. For comparative purposes the dashed line depicts the linear Rosenberg-type model with the same identification conditions imposed.

Figure 3: Characteristic-Beta Functions on Four 126-Month Subperiods

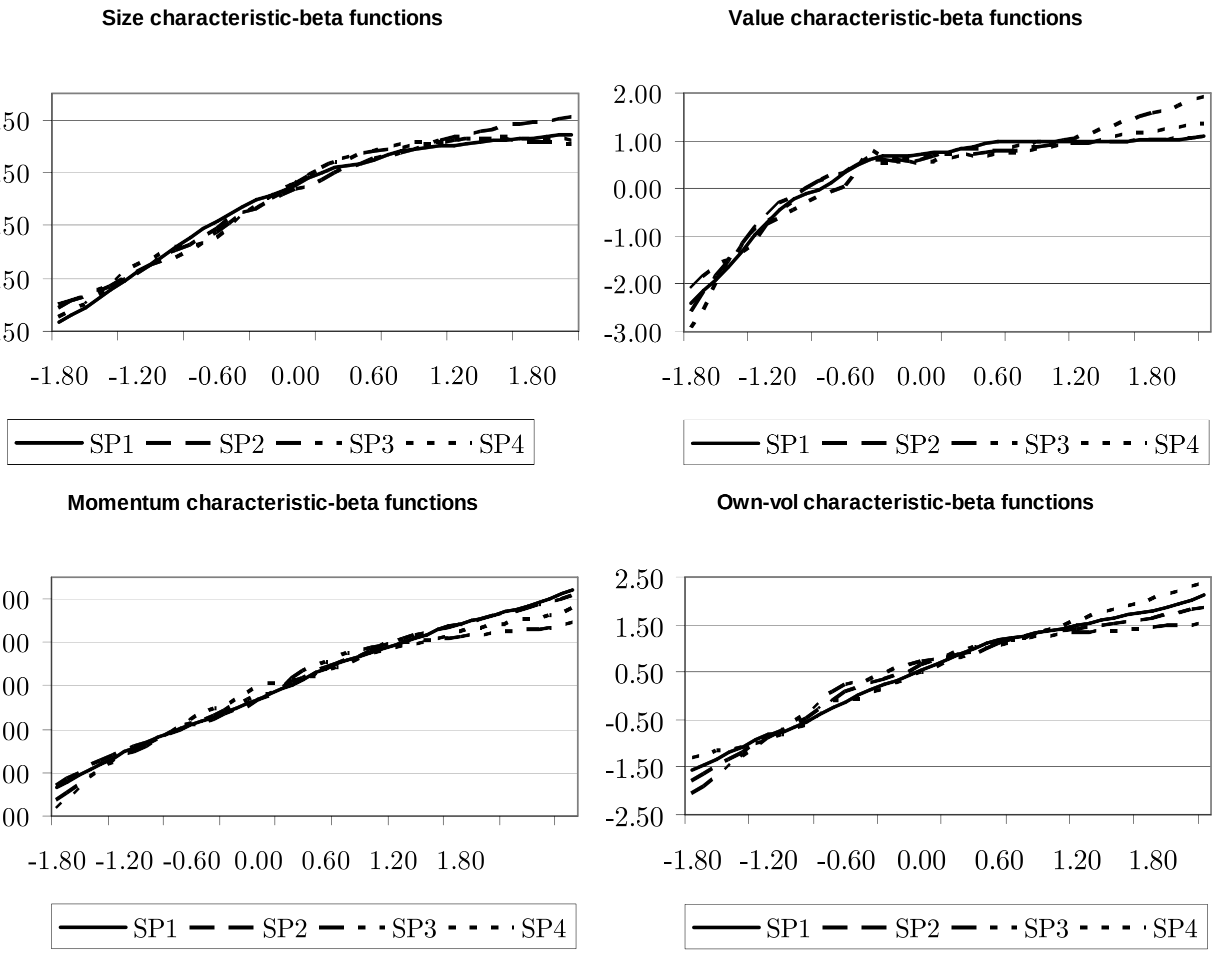


Figure 3 shows the characteristic-beta functions from four subsamples. SP1 line is a function for 1963 July-1973 December period, SP2 for 1974 January-1984 June period, while SP3 and SP4 for 1984 July-1994 December and 1995 January-2005 July periods, respectively.

Table 3: Uncentered R-squared Statistics Using Subsets of the Characteristics

	Marginal UR2 Statistics when Adding Factors First or Last to the Model				
	Market	Size	Value	Momentum	Volatility
Adding first	12.23%	1.50%	0.66%	1.26%	2.01%
Adding last	12.12%	0.89%	0.34%	0.74%	1.09%

	UR2 Linear and Nonlinear Five Factor Model	
	Linear	Nonlinear
UR2	16.20%	16.44%

Table 3 shows the time-series averages of uncentered cross-sectional R^2 (UR2) statistics as a measure of the explanatory power of the factor model. The upper part of the table shows average UR2 statistics from cross-sectional regressions of excess returns on each characteristic-beta function singly as well as their marginal contribution given the other four. The lower part of the table shows the UR2 statistic based on the model with all five characteristic-beta functions.

Table 4: Factor Return Statistics and Comparison to Fama French Factor Mimicking Portfolios

Panel a: Factor Return Statistics

	Market	Size	Value	Momentum	Volatility
Annualized Mean	9.65%	-2.66%	0.80%	-0.28%	-1.27%
Annualized Volatility	19.47%	5.78%	3.54%	5.85%	7.29%
% Periods significant*	91.07%	74.40%	57.34%	66.47%	75.20%
overall p-value	0.00%	0.00%	0.00%	0.00%	0.00%

Panel b: Empirical Factor Return and Fama French Mimicking Portfolio Return Correlations

	Market	Size	Value	Momentum	Volatility	RMRF	SMB	HML	FF_MOM
Market	1.00	0.00	-0.16	-0.39	0.80	0.86	0.66	-0.30	-0.17
Size	0.00	1.00	0.19	0.15	-0.01	0.34	-0.34	-0.10	-0.11
Value	-0.16	0.19	1.00	-0.12	-0.32	-0.23	-0.18	0.78	-0.21
Momentum	-0.39	0.15	-0.12	1.00	-0.41	-0.17	-0.18	-0.08	0.80
Volatility	0.80	-0.01	-0.32	-0.41	1.00	0.65	0.63	-0.44	-0.17
RMRF	0.86	0.34	-0.23	-0.17	0.65	1.00	0.30	-0.41	-0.06
SMB	0.66	-0.34	-0.18	-0.18	0.63	0.30	1.00	-0.28	0.02
HML	-0.30	-0.10	0.78	-0.08	-0.44	-0.41	-0.28	1.00	-0.12
FF_MOM	-0.17	-0.11	-0.21	0.80	-0.17	-0.06	0.02	-0.12	1.00

* defined as abs(t-value) > 1.96.

The upper part of Table 4 shows the mean, volatility, and statistical significance of each factor. The statistical significance is calculated as the percentage of significant t-statistics for each factor by estimating for each cross-sectional regression, the t-statistic for each estimated coefficient, based on Hansen-White heteroskedasticity consistent standard errors. Then for each factor, finding the average number of cross-sectional regression t-statistics that are significant at a 95% confidence level across the 504 time periods. The aggregate p-value is also provided. The lower part of Table 4 displays the correlations between the estimated factors, along with the three Fama-French factors, RMRF, SMB, and HML and a momentum factor FF-MOM created by Ken French. RMRF is the Fama-French market factor, it is the return to the value-weighted market index minus the riskfree return; SMB is the return to a small capitalization portfolio minus the return to a large-capitalization portfolio; HML is the return to a high book-to-price portfolio minus the return to a low book-to-price portfolio. FF-MOM is the return to a portfolio with high cumulative returns over the past twelve months minus the return to a portfolio with low cumulative returns over the past twelve months, adjusted to have roughly equal average capitalization.

Figure 4: Characteristic-based Mispricing Function

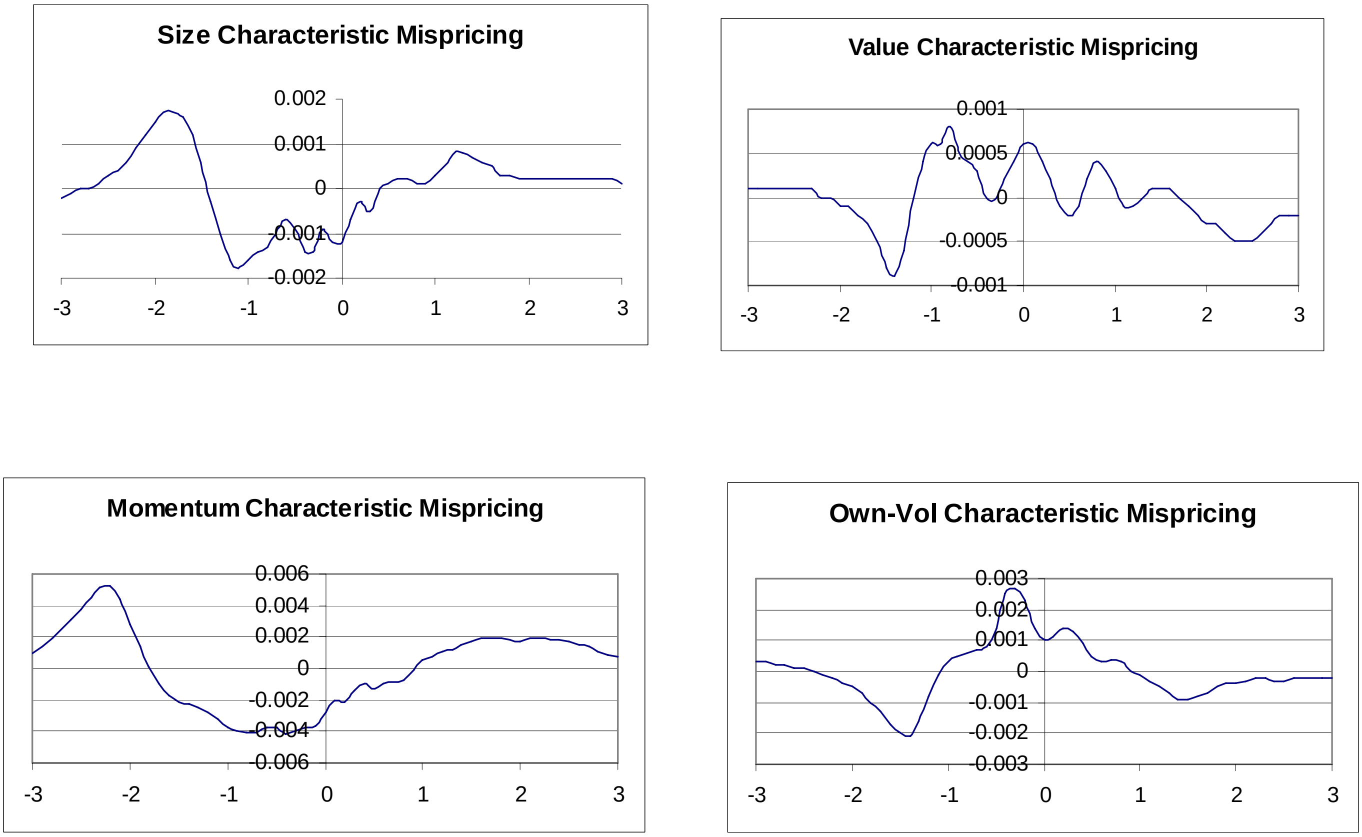


Figure 4 shows the four additive nonparametric characteristic-based mispricing functions over the grid of points between -3 and 3 at equally-spaced intervals of 0.1.

Table 5: Vector Autoregression for the Factor Returns

	Market	Size	Value	Momentum	Own-Volatility
Market(-1)	.122	-.074	-.0017	-.014	-.0049
	.086	.0017	.0013	.016	.020
Size(-1)	.733	.133	-.0008	-.116	.171
	.239	.047	.0036	.046	.056
Value(-1)	-2.18	-.496	.292	-.475	-1.22
	3.80	.751	.057	.723	.889
Mom(-1)	.201	-.144	-.0049	-.073	.118
	.287	.057	.0043	.055	.067
Own-Vol (-1)	.244	-.047	.0084	-.123	.200
	.431	.085	.0064	.082	.101
Constant	.0082	-.00066	.00006	.00031	.00019
	.0029	.00056	.00004	.00054	.00067
R^2	.0544	.129	.0683	.0539	.0605

Table 5 shows the results from a first-order vector autoregression of the five factors returns on their lagged values.

Table 6: Vector Autoregression for the Squared Factor Returns

	Market^2	Size^2	Value^2	Momentum^2	Own-Volatility^2
Market(-1)^2	-.0020	.0040	-6.12E-06	.0014	-.0039
	.052	.0020	1.3E-05	.0027	.0035
Size(-1) ^2	1.40	.238	.0024	.280	.520
	1.37	.052	.0003	.070	.090
Value(-1) ^2	-76.03	32.39	.230	14.08	19.07
	248.3	9.48	.062	12.61	16.37
Mom(-1) ^2	.699	-.102	-7.94E-05	.126	.014
	1.42	.054	.00035	.072	.094
Own-Vol (-1) ^2	-.390	-.013	4.48E-06	-.110	.041
	1.36	.052	.00034	.069	.090
Constant	.0037	9.89E-05	3.08E-07	8.09E-05	.00011
	.0005	1.8E-05	1.2E-07	2.4E-04	3.1E-05
R^2	.0030	.145	.257	.067	.141

Table 6 shows a first-order vector autoregression of squared factor returns on their lagged values.

Figure 5: Time series of cross-sectional root-mean-square asset-specific return

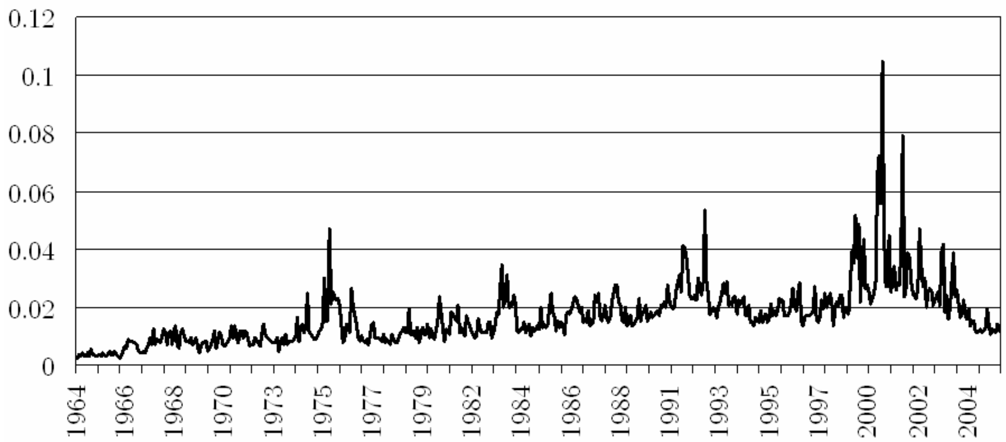


Figure 5 shows the time-series of cross-sectional mean-square asset-specific return over the sample period July 1964 to June 2005.

Figure 6: The Characteristic-beta Functions with Observed Market Index

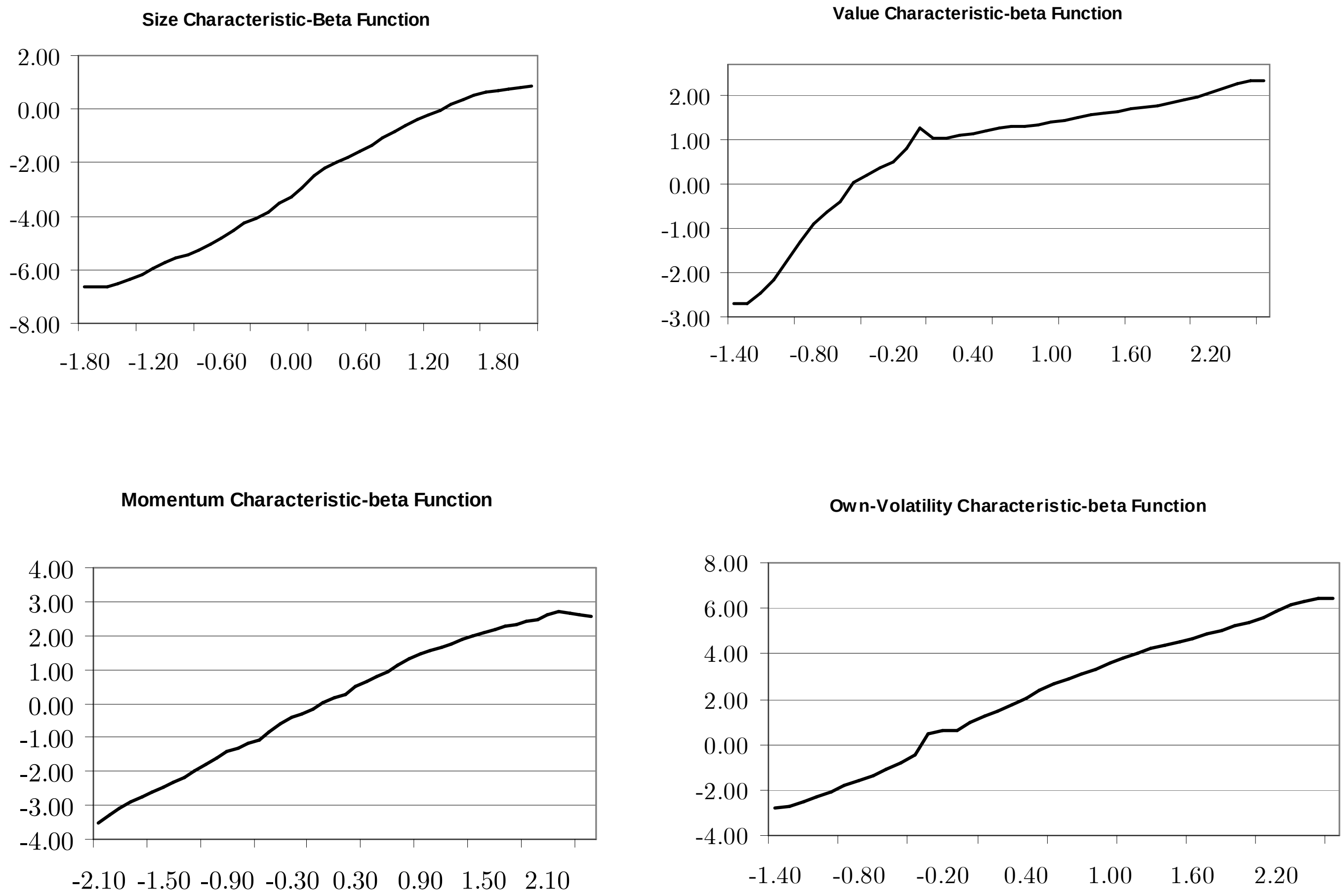


Figure 6 displays the estimated characteristic-beta functions over the range of grid points covering 95% of characteristic values. The model is estimated with an observed market index. The characteristic-beta functions are standardized to have a capitalization-weighted mean of zero and capitalization-weighted variance of one.

Table 7: Tests of the Capital Asset Pricing Model as a Restriction on the Non-Market Factor Returns

	α	β	R^2
Full Period: 7/1963 - 6/2005			
Size	-.0011 (-3.40)	.0065 (.875)	.002
Value	.0006 (2.09)	-.057 (-9.06)	.140
Momentum	-.0004 (.755)	-.039 (-3.41)	.023
Own-Volatility	-.0008 (-3.21)	.105 (17.0)	.367
First Subperiod: 7/1963-12/1973			
Size	-.0006 (-1.07)	-.061 (-4.08)	.118
Value	.0001 (.275)	-.040 (-2.92)	.065
Momentum	.0012 (1.54)	-.056 (-2.68)	.055
Own-Volatility	-.0006 (1.69)	.104 (10.4)	.465
Second Subperiod: 1/1974 – 6/1984			
Size	-.0027 (-3.78)	-.010 (-.723)	.004
Value	.0003 (.499)	-.063 (-5.73)	.207
Momentum	-.0007 (-.734)	-.021 (-1.09)	.009
Own-Volatility	-.0010 (-2.97)	.074 (10.5)	.469
Third Subperiod: 7/1984 - 12/1994			
Size	.0002 (.440)	.035 (2.78)	.059
Value	.0010 (2.15)	-.042 (-3.83)	.106
Momentum	.0003 (.497)	.004 (.285)	.001
Own-Volatility	-.0007 (-1.99)	.054 (6.55)	.257
Fourth Subperiod: 1/1995 - 6/2005			
Size	-.0017 (-2.40)	.047 (3.04)	.070
Value	.0010 (1.36)	-.080 (-5.20)	.182
Momentum	.0007 (.0480)	-.085 (-2.72)	.057
Own-Volatility	-.0012 (-1.46)	.186 (10.5)	.477

Table 7 shows the results from univariate time-series regression of each of the four characteristic-based factors on a constant and the excess return to the market index. The CAPM implies that the set of four intercepts from these univariate regressions should jointly equal zero. T-statistics are shown in parentheses next to each estimated coefficient. The regressions are run over the full 504 month sample period and for each of four 126-month subperiods.