Testing conditional factor models*

Andrew Ang[†]
Columbia University and NBER

Dennis Kristensen[‡]
UCL and IFS

This Version: 20 February, 2012

JEL classification: C12, C13, C14, C32, G12.

Keywords: Nonparametric estimator; Time-varying beta;
Conditional alpha; Book-to-market premium; Value and momentum.

*We thank Pengcheng Wan for research assistance and Kenneth French for providing data. For helpful comments and suggestions, we thank an anonymous referee, Matias Cattaneo, Will Goetzmann, Jonathan Lewellen, Serena Ng, Jay Shanken, Masahiro Watanabe and Guofu Zhou, as well as seminar participants at Columbia University, Dartmouth University, Georgetown University, Emory University, Federal Reserve Board, Massachusetts Institute of Technology, Oxford-Man Institute of Quantitative Finance, Princeton University, State University in New York at Albany, Washington University in St. Louis, Yale University, University of Montréal, the American Finance Association 2010 meeting, the Banff International Research Station 2009 Conference on "Semiparametric and Nonparametric Methods in Econometrics", the Econometric Society Australasian 2009 meeting, the Humboldt-Copenhagen 2009 Conference on "Recent Developments in Financial Econometrics", the 2010 National Bureau of Economic Research Summer Institute, and the New York University Five Star 2009 conference. Andrew Ang acknowledges funding from the Network for Studies on Pension, Aging and Retirement (Netspar). Dennis Kristensen acknowledges funding from the Danish National Research Foundation through a grant to Center for Research in Econometric Analysis of Time Series (CREATES) and the National Science Foundation (grant no. SES-0961596). Corresponding author: Dennis Kristensen

[†]Email: aa610@columbia.edu

[‡]Email: d.kristensen@ucl.ac.uk

Abstract

Using nonparametric techniques, we develop a methodology for estimating and testing conditional alphas and betas and long-run alphas and betas, which are the averages of conditional alphas and betas, respectively, across time. The estimators and tests can be implemented for a single asset or jointly across portfolios. The traditional Gibbons, Ross, and Shanken (1989) test arises as a special case of no time variation in the alphas and factor loadings and homoskedasticity. As applications of the methodology, we estimate conditional CAPM and multifactor models on book-to-market and momentum decile portfolios. We reject the null that long-run alphas are equal to zero even though there is substantial variation in the conditional factor loadings of these portfolios.

1 Introduction

Under the null of a factor model, an asset's expected excess return should be zero after controlling for that asset's systematic factor exposure. Traditional regression tests of whether an alpha is equal to zero, such as the widely used Gibbons, Ross, and Shanken (1989) test, assume that the factor loadings are constant. However, overwhelming empirical evidence shows that factor loadings, especially for the standard capital asset pricing model (CAPM) and Fama and French (1993) models, vary substantially over time. Factor loadings exhibit variation even at the portfolio level (see, among others, Fama and French, 1997; Lewellen and Nagel, 2006; and Ang and Chen, 2007). Time-varying factor loadings can distort standard factor model tests for whether the alphas are equal to zero and, thus, render traditional statistical inference for the validity of a factor model to be possibly misleading.

We introduce a methodology to estimate time-varying alphas and betas in conditional factor models. Conditional on the realized alphas and betas, our factor specification can be regarded as a regression model with changing regression coefficients. We impose no parametric assumptions on the nature of the realized time variation of the alphas and betas and estimate them nonparametrically based on techniques similar to those found in the literature on realized volatility (see, e.g., Foster and Nelson, 1996; and Andersen, Bollerslev, Diebold and Wu, 2006).[1] We also develop estimators of the long-run alphas and betas, defined as the averages of the conditional alphas or factor loadings, respectively, across time. Our estimators are highly robust due to their nonparametric nature.

Based on the conditional and long-run estimators, we propose short- and long-run tests for the asset pricing model. Major advantages of our estimators and tests are that they are straightforward to apply, powerful, and involve no more than running a series of kernel-weighted ordinary least-squares (OLS) regressions for each asset. The tests can be applied to a single asset or jointly across a system of assets. In the special case in which betas are constant and there is no heteroskedasticity, our long-run tests for whether the long-run alphas equal zero are asymptotically equivalent to Gibbons, Ross, and Shanken (1989).

We analyze the estimators and tests in a continuous-time setting where the conditional al-

[1] Other papers in finance developing nonparametric estimators include Stanton (1997), Aït-Sahalia (1996), and Bandi (2002), who estimate drift and diffusion functions of the short rate. Bansal and Viswanathan (1993), Aït-Sahalia and Lo (1998), and Wang (2003) characterize the pricing kernel by nonparametric estimation. Brandt (1999) and Aït-Sahalia and Brandt (2007) present applications of nonparametric estimators to portfolio choice and consumption problems.

phas and betas can be thought of as the instantaneous drift of the assets and the covariance between asset and factor returns, respectively. All our estimations, however, are of discrete-time models and so our methodology is widely applicable to the majority of empirical asset studies that estimate factor models in discrete time. As is well known from the literature on drift and volatility estimation in continuous time, one can learn about volatilities or covariances from data for any fixed time span with increasingly dense observations, while pinning down the drift requires a long span of data (see, e.g., Merton, 1980). We obtain similar results in our setting: The conditional beta estimators are consistent under very weak restrictions on the data-generating process for fixed time spans, while the conditional alpha estimators are, in general, inconsistent.

Whereas a large number of applied papers implement rolling-window estimators of conditional alphas and use them in statistical inference (see the summary by Ferson and Qian, 2004), under a continuous-time model we show that conditional alphas cannot be estimated consistently. However, we demonstrate that under additional restrictions on the model involving certain time normalizations of the model parameters, a formal asymptotic theory of the conditional alpha estimators can, in fact, be developed. These additional assumptions supply conditions under which the popular rolling-window conditional alphas have well-defined (asymptotic) distributions, but the required time normalization is economically counterintuitive when interpreted in the context of a continuous-time factor model. In contrast to the conditional alphas, the long-run alphas are identified from data without any time normalizations. We develop an asymptotic theory for our long-run alpha and beta estimators, which converge at standard rates as found in parametric diffusion models.

Our approach builds on a literature advocating the use of short windows with high-frequency data to estimate time-varying second moments or betas, such as French, Schwert, and Stambaugh (1987) and Lewellen and Nagel (2006). In particular, Lewellen and Nagel estimate time-varying factor loadings and infer conditional alphas. In the same spirit of Lewellen and Nagel, we use local information to obtain estimates of conditional alphas and betas without having to instrument time-varying factor loadings with macroeconomic and firm-specific variables.[2] Our work extends this literature in several important ways.

First, we provide a formal distribution theory for conditional and long-run estimators which

[2] The instrumental variables approach is taken by Shanken (1990) and Ferson and Harvey (1991), among others. As Ghysels (1998) and Harvey (2001) note, the estimates of the factor loadings obtained using instrumental variables are very sensitive to the variables included in the information set. Furthermore, many conditioning variables, especially macro and accounting variables, are only available at coarse frequencies.

the earlier literature did not provide. For example, the Lewellen and Nagel (2006) procedure identifies the time variation of conditional betas and provides period-by-period estimates of conditional alphas on short, fixed windows equally weighting all observations in that window. We show this is a special case (a one-sided filter) of our general estimator and so our theoretical results apply. Lewellen and Nagel further test whether the average conditional alpha is equal to zero using a Fama and MacBeth (1973) procedure. Because this is nested as a special case of our methodology, we provide formal arguments for the validity of this procedure. We also develop data-driven methods for choosing optimal window widths used in estimation.

Second, by using kernel methods to estimate time-varying betas we are able to use all the data efficiently in the estimation of conditional alphas and betas at any particular time. Naturally, our methodology allows for any valid kernel and so nests the one-sided, equal-weighted filters used by French, Schwert, and Stambaugh (1987), Andersen, Bollerslev, Diebold and Wu (2006), Lewellen and Nagel (2006), and others, as special cases. All of these studies use truncated, backward-looking windows to estimate second moments that have larger mean square errors (MSE's) compared with estimates based on two-sided kernels.[3]

Third, we develop tests for the significance of conditional and long-run alphas jointly across assets in the presence of time-varying betas. Earlier work incorporating time-varying factor loadings restricts attention to only single assets, whereas our methodology can incorporate a large number of assets. Our procedure can be viewed as the conditional analogue of Gibbons, Ross, and Shanken (1989), who jointly test whether alphas are equal to zero across assets, where we now permit the alphas and betas to vary over time. Joint tests are useful for investigating whether a relation between conditional alphas and firm characteristics strongly exists across many portfolios and have been extensively used by Fama and French (1993) and many others.

Our work is most similar to tests of conditional factor models contemporaneously examined by Li and Yang (2011). Li and Yang also use nonparametric methods to estimate conditional parameters and formulate a test statistic based on average conditional alphas. However, they do this in a discrete-time setting, do not investigate conditional or long-run betas, and do not develop tests of constancy of conditional alphas or betas. One important issue is the bandwidth

[3] Foster and Nelson (1996) derive optimal two-sided filters to estimate time-varying covariance matrices for a general class of time series models. Foster and Nelson's exponentially declining weights can be replicated by special choice kernel weights. An advantage of using a nonparametric procedure is that we obtain efficient estimates of betas without having to specify a particular data generating process, whether this is generalized autoregressive conditional heteroskedastic (GARCH) model (see, for example, Bekaert and Wu, 2000) or a stochastic volatility model (see, for example, Jostova and Philipov, 2005; Ang and Chen, 2007).

selection procedure, which requires different bandwidths for conditional or long-run estimates. Li and Yang do not provide an optimal bandwidth selection procedure. They also do not derive specification tests jointly across assets as in Gibbons, Ross, and Shanken (1989), which we nest as a special case, or present a complete distribution theory for their estimators.

The rest of this paper is organized as follows. Section 2 lays out our empirical methodology. Section 3 discusses our data. In Sections 4 and 5 we investigate tests of conditional CAPM and Fama and French models on the book-to-market and momentum portfolios, respectively. Section 6 concludes. We relegate all technical proofs to the Appendix.

2 Statistical methodology

We present our conditional factor model and develop nonparametric estimators and test statistics for the conditional alphas and betas.

2.1 *Conditional factor model*

Let $R = (R_1, ..., R_M)'$ denote a vector of excess returns of M assets observed at n time points, $0 < t_1 < t_2 < ... < t_n < T$, within a time span $T > 0$. We wish to explain the returns through a set of J common tradeable factors, $f = (f_1, ..., f_J)'$, which are observed at the same time points. We assume the following conditional factor model explains the returns of stock k ($k = 1, ..., M$) at time t_i ($i = 1, ..., n$):

$$R_{k,i} = \alpha_k (t_i) + \beta_k (t_i)' f_i + \omega_{kk} (t_i) z_{k,i}, \tag{1}$$

where $R_{k,i}$ and f_i are the observed return and factors respectively at time t_i. This can be rewritten in matrix notation:

$$R_i = \alpha (t_i) + \beta (t_i)' f_i + \Omega^{1/2} (t_i) z_i, \tag{2}$$

where $\alpha (t) = (\alpha_1 (t), ..., \alpha_M (t))' \in \mathbb{R}^M$ is the vector of conditional alphas across stocks $k = 1, ..., M$ and $\beta (t) = (\beta_1 (t), ..., \beta_M (t))' \in \mathbb{R}^{J\times M}$ is the corresponding matrix of conditional betas. The alphas and betas can take on any sample path in the data, subject to the (weak) restrictions in Appendix A, including nonstationary and discontinuous cases, and time-varying dependence of conditional betas and factors. The vector $z_i = (z_{1,i}, ..., z_{M,i})' \in \mathbb{R}^M$ contains the errors and the covariance matrix $\Omega (t) = [\omega_{jk}^2 (t)]_{j,k} \in \mathbb{R}^{M\times M}$ allows for both heteroskedasticity and time-varying cross-sectional correlations.

Letting $\mathcal{F}_i = \mathcal{F}\left\{R_j, f_j, \alpha\left(t_j\right), \beta\left(t_j\right) : j \leq i\right\}$ denote the filtration up to time t_i, we assume the error term satisfies

$$\mathrm{E}\left[z_i|\mathcal{F}_i\right] = 0 \quad \text{and} \quad \mathrm{E}\left[z_i z_i'|\mathcal{F}_i\right] = I_M, \tag{3}$$

where I_M denotes the M-dimensional identity matrix. Eq. (3) is the identifying assumption of the model and rules out non-zero correlations between the factor and the errors. This orthogonality assumption is an extension of standard OLS, which specifies that errors and factors are orthogonal.[4] Importantly, this condition does not rule out the alphas and betas being correlated with the factors. That is, the conditional factor loadings can be random processes in their own right and exhibit (potentially time-varying) dependence with the factors. Thus, we allow for a rich set of dynamic trading strategies of the factor portfolios.

We are interested in time series estimates of the realized conditional alphas, $\alpha(t)$, and the conditional factor loadings, $\beta(t)$, along with their standard errors. Under the null of a factor model, the conditional alphas are equal to zero, or $\alpha(t) = 0$. As Jagannathan and Wang (1996) point out, if the correlation of the factor loadings, $\beta(t_i)$, with factors, f_i, is zero, then the unconditional pricing errors of a conditional factor model are mean zero and an unconditional OLS methodology could be used to test the conditional factor model. When the betas are correlated with the factors then the unconditional alpha reflects both the true conditional alpha and the covariance between the betas and the factor (see Jagannathan and Wang, 1996 and Lewellen and Nagel, 2006).

Given the realized alphas and betas, we define the long-run alphas and betas for asset $k = 1, ..., M$ as

$$\begin{aligned} \alpha_{\mathrm{LR},k} &\equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \alpha_k(t_i) \in \mathbb{R}, \\ \beta_{\mathrm{LR},k} &\equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \beta_k(t_i) \in \mathbb{R}^J, \end{aligned} \tag{4}$$

We use the terminology "long run" (LR) to distinguish the conditional alpha at a particular time, $\alpha_k(t)$, from the conditional alpha averaged over the sample, $\alpha_{\mathrm{LR},k}$. When the factors are

[4] The strict factor structure rules out leverage effects and other nonlinear relations between asset and factor returns, which Boguth, Fisher, and Simutin (2011) argue could lead to additional biases in the estimators. We expect that our theoretical results are still applicable under weaker assumptions, but this requires specification of the appropriate correlation structure between the alphas, betas, and error terms. Appendix A details our technical assumptions and Appendix B contains proofs. We leave these extensions to further research. Simulation results show that our long-run alpha estimators perform well under mild misspecification of modestly correlated betas and error terms.

correlated with the betas, the long-run alphas are potentially different from OLS alphas.

We test the hypothesis that the long-run alphas are jointly equal to zero across M assets:

$$H_0 : \alpha_{\mathrm{LR},k} = 0, \quad k = 1, ..., M. \tag{5}$$

In a setting with constant alphas and betas, Gibbons, Ross, and Shanken (1989) develop a test of the null H_0. Our methodology can be considered the conditional version of the Gibbons, Ross and Shanken test when both conditional alphas and betas potentially vary over time. In addition, we test the stronger hypothesis of the conditional alphas being zero at any given point in time,

$$H_{0,k} : \alpha_k(t) = 0 \text{ for all } t. \tag{6}$$

2.2 *Conditional estimators*

Our analysis of the model and estimators is done conditional on the particular realization of alphas and betas that generated data. That is, our analysis relies on the following conditional relation between the observations and the parameters of interest that holds under the orthogonality condition in Eq. (3):

$$\left[\alpha\left(t_i\right), \beta\left(t_i\right)'\right]' = \Lambda^{-1}\left(t_i\right) E\left[X_i R_i' | \mathcal{F}_i\right], \quad X_i = \left(1, f_i'\right)', \tag{7}$$

where $\Lambda\left(t_i\right)$ denotes the conditional second moment of the regressors:

$$\Lambda(t_i) \equiv \mathrm{E}\left[X_i X_i' | \mathcal{F}_i\right]. \tag{8}$$

Eq. (7) identifies the particular realization of alphas and betas that generated data.

The time variation in $\Lambda\left(t\right)$ reflects potential correlation between factors and betas. If there is zero correlation (and the factors are stationary), then $\Lambda\left(t\right) = \Lambda$ is constant over time, but in general $\Lambda\left(t\right)$ varies over time. One advantage of conducting the analysis conditional on the sample is that we can tailor our estimates of the particular realization of alphas and betas.

A natural way to estimate $\alpha\left(t\right)$ and $\beta\left(t\right)$ is by replacing the population moments in Eq. (7) by their sample versions. Given observations of returns and factors, we propose the following local least squares estimators of $\alpha_k\left(t\right)$ and $\beta_k\left(t\right)$ for asset k in Eq. (1) at any time $0 \leq t \leq T$:

$$[\hat{\alpha}_k\left(t\right), \hat{\beta}_k\left(t\right)']' = \arg\min_{(\alpha,\beta)} \sum_{i=1}^{n} K_{h_k T}\left(t_i - t\right)\left(R_{k,i} - \alpha - \beta' f_i\right)^2, \tag{9}$$

where $K_{h_kT}(z) \equiv K(z/(h_kT))/(h_kT)$ with $K(\cdot)$ being a kernel and $h_k > 0$ a bandwidth. The optimal estimators solving Eq. (9) are simply kernel-weighted least squares:

$$[\hat{\alpha}_k(t), \hat{\beta}_k(t)']' = \left[\sum_{i=1}^{n} K_{h_kT}(t_i - t) X_i X_i'\right]^{-1} \left[\sum_{i=1}^{n} K_{h_kT}(t_i - t) X_i R_{k,i}\right]. \qquad (10)$$

The proposed estimators are sample analogues to Eq. (7) giving weights to the individual observations according to how close in time they are to the time point of interest, t. The shape of the kernel, K, determines how the different observations are weighted. For most of our empirical work we choose the Gaussian density as kernel,

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right),$$

but also examine one-sided and uniform kernels that have been used in the literature by Andersen, Bollerslev, Diebold and Wu (2006) and Lewellen and Nagel (2006), among others. In common with other nonparametric estimation methods, as long as the kernel is symmetric, the most important choice is not so much the shape of the kernel as the bandwidth, h_k. The bandwidth, $h_k \in (0, 1)$, controls the proportion of data obtained in the sample span $[0, T]$ that is used in the computation of the estimated alphas and betas. A small bandwidth means only observations very close to t are included in the estimation. The bandwidth controls the bias and variance of the estimator and it should, in general, be sample specific. In particular, as the sample size grows, the bandwidth should shrink toward zero at a suitable rate in order for any finite-sample biases and variances to vanish. We discuss the bandwidth choice in Section 2.8.

We run the kernel regression in Eq. (9) separately stock by stock for $k = 1, ..., M$. This is a generalization of the regular OLS estimators, which are also run stock by stock in the Gibbons, Ross, and Shanken (1989) test. If the same bandwidth h is used for all stocks, our estimator of alphas and betas across all stocks take the simple form of a weighted multivariate OLS estimator,

$$[\hat{\alpha}(t), \hat{\beta}(t)']' = \left[\sum_{i=1}^{n} K_{hT}(t_i - t) X_i X_i'\right]^{-1} \left[\sum_{i=1}^{n} K_{hT}(t_i - t) X_i R_i\right]. \qquad (11)$$

In practice it is not advisable to use one common bandwidth across all assets. We use different bandwidths for different stocks because the variation and curvature of the conditional alphas and betas could differ widely across stocks and each stock could have a different level of heteroskedasticity. We show below that, for book-to-market and momentum test assets, the patterns of conditional alphas and betas are dissimilar across portfolios. Choosing stock-specific bandwidths allows us to better adjust the estimators for these effects. However, to avoid cumbersome notation, we present the asymptotic results for the estimators $\hat{\alpha}(t)$ and $\hat{\beta}(t)$ assuming one common bandwidth, h, across all stocks. The asymptotic results are identical in the case with multiple bandwidths under the assumption that these all converge at the same rate as $n \to \infty$.

2.3 *Continuous-time factor model*

For the theoretical analysis of the proposed estimators, we introduce a continuous-time version of the discrete-time factor model. Suppose that the vector of log-prices of the M risky assets (in excess of the risk-free asset), $s(t) = \log S(t) \in \mathbb{R}^M$, solve the stochastic differential equation

$$ds(t) = \alpha(t)\, dt + \beta(t)' \, dF(t) + \Sigma^{1/2}(t)\, dB(t), \tag{12}$$

where $F(t)$ are J factors and $B(t)$ is a M-dimensional Brownian motion. This is the ANOVA (analysis of variance) model considered in Andersen, Bollerslev, Diebold and Wu (2006) and Mykland and Zhang (2006). Suppose we have observed $s(t)$ and $F(t)$ over the time span $[0, T]$ at n discrete time points, $0 \leq t_0 < t_1 < ... < t_n \leq T$. We wish to estimate the spot alphas, $\alpha(t) \in \mathbb{R}^M$, and betas, $\beta(t) \in \mathbb{R}^{J \times M}$, which can be interpreted as the realized instantaneous drift of $s(t)$ and (co-)volatility of $(s(t), F(t))$, respectively. For simplicity, we assume that the observations are equidistant in time such that $\Delta \equiv t_i - t_{i-1}$ is constant; in particular, $n\Delta = T$.

To facilitate the analysis of the estimators in this diffusion setting, we introduce a discretized version of the continuous-time model,

$$\Delta s_i = \alpha(t_i)\, \Delta + \beta(t_i)' \, \Delta F_i + \Sigma^{1/2}(t_i) \sqrt{\Delta} z_i, \quad i = 1, 2, ..., n, \tag{13}$$

where z_i are independently and identically distributed (i.i.d.) with mean zero and covariance I_M,

$$\Delta s_i = s(t_i) - s(t_{i-1}) \text{ and } \Delta F_i = F(t_i) - F(t_{i-1}).$$

In the following we treat Eq. (13) as the true, data-generating model. The extension to treat (12) as the true model would require some extra effort to ensure that the discretized version in Eq. (13) is an asymptotically valid approximation of Eq. (12). The analysis would involve controlling the various discretization biases that would need to vanish sufficiently fast as $\Delta \to 0$. This could be done along the lines of Bandi and Phillips (2003) and Kristensen (2010), among others.

Defining

$$R_i \equiv \Delta s_i / \Delta, \quad f_i \equiv \Delta F_i / \Delta, \quad \Omega_t \equiv \Sigma(t) / \Delta, \tag{14}$$

we can rewrite the discretized diffusion model in the form of Eq. (2). Natural estimators of $\alpha(t)$ and $\beta(t)$, therefore, take on the same form as the discrete-time estimators in Eq. (11).

2.4 *Conditional beta estimators*

We now analyze the properties of our estimator $\hat{\beta}(t)$ under the assumption that the discretized version of the diffusion model (13) is the data-generating process. As is well known in the literature on estimation of diffusion models (see, e.g., Merton, 1980; Bandi and Phillips, 2003; and Kristensen, 2010), we can consistently estimate the instantaneous betas, $\beta(t)$, as $\Delta \to 0$ under weak regularity conditions. In addition to Eq. (13), we assume that the factors satisfy the discretized diffusion model

$$\Delta F_i = \mu_F(t_i)\,\Delta + \Lambda_{FF}^{1/2}(t_i)\,\sqrt{\Delta} u_i, \tag{15}$$

where $u_i \sim$ i.i.d$(0, I_J)$ and $\mu(\cdot)$ and $\Lambda_{FF}(\cdot)$ are r times differentiable (possibly random) functions.

Under regularity conditions stated in Appendix A, the bias and variance of the estimator are

$$E[\hat{\beta}(t)] \simeq \beta(t) + (hT)^2\,\beta^{(2)}(t) \quad \text{and} \quad \mathrm{Var}(\hat{\beta}(t)) \simeq \frac{1}{nh} \times \kappa_2 \Lambda_{FF}^{-1}(t) \otimes \Sigma(t),$$

where $\beta^{(2)}(t)$ denotes the second derivative of $\beta(t)$ and $\kappa_2 = \int K^2(z)\,dz$ $(= 0.2821$ for the normal kernel).[5] These expressions show the usual trade-off between bias and variance for kernel regression estimators with h needing to be chosen to balance the two. In particular, as $hT \to 0$ and $nh \to \infty$, $\hat{\beta}(t) \xrightarrow{p} \beta(t)$. Letting $h \to 0$ at a suitable rate, the bias term can be ignored and we obtain the following asymptotic result:

Theorem 1 *Assume that assumptions A.1–A.3 given in Appendix A hold and the bandwidth is chosen such that* $nh \to \infty$ *and* $nT^4h^5 \to 0$*. Then, for any* $t \in [0, T]$*,*

$$\sqrt{nh}\{\hat{\beta}(t) - \beta(t)\} \sim N\left(0, \kappa_2 \Lambda_{FF}^{-1}(t) \otimes \Sigma(t)\right) \quad \textit{in large samples.} \tag{16}$$

Moreover, the conditional estimators are asymptotically independent across any set of distinct time points.

[5] We assume that $t \mapsto \beta(t)$ is twice differentiable. This assumption could be replaced by, for example, a Lipschitz condition, $\|\beta(s) - \beta(t)\| \leq C|s-t|^{\lambda}$ for some $\lambda > 0$, in which case the bias component would change and be of order $O((hT)^{\lambda})$.

This result is a multivariate extension of the asymptotic distribution for kernel-based estimators of spot volatilities found in Kristensen (2010, Theorem 1). Andersen, Bollerslev, Diebold and Wu (2006) develop estimators of integrated factor loadings, $\int_0^T \beta(s)\,ds$, which implicitly ignore the variation of beta within each window. Our estimator is a local version of the asymptotics for the integrated beta (see also Foster and Nelson, 1996). By choosing a flat kernel and the bandwidth, $h > 0$, to match the chosen time window, our proposed estimators nest the realized beta estimators. But, while Andersen, Bollerslev, Diebold and Wu (2006) develop an asymptotic theory for a fixed window width, Theorem 1 establishes results in which the time window shrinks with sample size. This allows us to recover the instantaneous conditional betas.

In Theorem 1, the rate of convergence of $\hat{\beta}(t)$ is $\sqrt{nh}$, which is slower than parametric estimators because $h \to 0$. This is common to all non-parametric estimators. The asymptotic analysis and properties of the estimators are closely related to the kernel-regression type estimators of diffusion models proposed in Bandi and Phillips (2003), Kanaya and Kristensen (2010), and Kristensen (2010). Bandi and Phillips (2003) focus on univariate Markov diffusion processes and use the lagged value of the observed process as kernel regressor, while Kanaya and Kristensen (2010) and Kristensen (2010) consider estimation of univariate stochastic volatility models. In contrast, we model time-inhomogenous, multivariate processes in which the observation times, $t_1, ..., t_n$, are used as the kernel regressor. Because we only smooth over the univariate time variable t, increasing the number of regressors, J, or the number of stocks, M, does not affect the performance of the estimator.

Estimators of the two terms appearing in the asymptotic variance in eq. (16) are obtained as:

$$\begin{aligned}\hat{\Lambda}_{FF}(t) &= \frac{\Delta \sum_{i=1}^{n} K_{hT}(t_i - t)\left[f_i - \hat{\mu}_F(t_i)\right]\left[f_i - \hat{\mu}_F(t_i)\right]'}{\sum_{i=1}^{n} K_{hT}(t_i - t)} \\ \hat{\Sigma}(t) &= \frac{\Delta \sum_{i=1}^{n} K_{hT}(t_i - t)\,\hat{\varepsilon}_i\hat{\varepsilon}_i'}{\sum_{i=1}^{n} K_{hT}(t_i - t)},\end{aligned} \tag{17}$$

where $\hat{\varepsilon}_i = R_i - \hat{\alpha}(t_i) - \hat{\beta}(t_i)' f_i$ are the residuals and $\hat{\mu}_F(t)$ is an estimator of the instantaneous drift in the factors,

$$\hat{\mu}_F(t) = \frac{\sum_{i=1}^{n} K_{hT}(t_i - t)\, f_i}{\sum_{i=1}^{n} K_{hT}(t_i - t)}. \tag{18}$$

Due to the asymptotic independence across different values of t, confidence bands over a given grid of time points can easily be computed.

2.5 *Conditional alpha estimators*

Unlike conditional betas, conditional alphas are not identified in data without additional restrictions on the time series variation and without increasing the data over long time spans ($T \to \infty$), which was first demonstrated by Merton (1980). Without further restrictions, the estimator of $\alpha(t)$ satisfies

$$E[\hat{\alpha}(t)] \simeq \alpha(t) + (Th)^2 \alpha^{(2)}(t), \quad \text{Var}(\hat{\alpha}(t)) \simeq \frac{1}{Th} \times \kappa_2 \Sigma(t), \tag{19}$$

as $\Delta \to 0$. Relative to $\hat{\beta}(t)$, the bias of $\hat{\alpha}(t)$ is of the same order but its variance vanishes slower, $1/(Th)$ versus $1/(nh)$. The slower rate of convergence of Var$(\hat{\alpha}(t))$ is a well-known feature of nonparametric drift estimators in diffusion models, as in Bandi and Phillips (2003), and is due to the smaller amount of information regarding the drift relative to the volatility found in data.

Observe that the bias and variance of $\hat{\alpha}(t)$ are perfectly balanced. To remove the bias, we have to let $Th \to 0$, but with this bandwidth choice the variance explodes. This simply mirrors the well-known fact that, in continuous time, the local variation of observed returns is too noisy to extract information about the drift. As such, we cannot state any formal results regarding the asymptotic distribution of $\hat{\alpha}(t)$. However, informally, with h chosen small enough such that the bias is negiglible, we have

$$\sqrt{Th}\{\hat{\alpha}(t) - \alpha(t)\} \sim N(0, \kappa_2 \Sigma(t)) \quad \text{in large samples.} \tag{20}$$

It should be stressed, though, that without further restriction on the data-generating process, the conditional alpha estimates can be interpreted only as noisy estimates of the underlying conditional alpha process. In particular, the compuation of standard errors and confidence bands for the conditional alphas based on Eq. (20) ignores the bias component which might be substantial. As such, standard errors and confidence bands for conditional alphas should be interpreted with caution.

A large empirical asset pricing literature interprets constant terms in OLS regressions estimated over different sample periods as conditional alphas, at least since Gibbons and Ferson (1985).[6] Given the wide-spread use of conditional alpha estimators, it is of interest to provide conditions under which the statement in Eq. (20) is formally (i.e., asymptotically) correct.

[6] See, among many others, Shanken (1990), Ferson and Schadt (1996), Christopherson, Ferson and Glassman (1998), and more recently Mamaysky, Spiegel and Zhang (2008). Ferson and Qian (2004) provide a summary of this large literature.

One such condition is to impose a recurrency restriction used in the literature on nonparametric estimation of diffusion models. In particular, Bandi and Phillips (2003) assume that the instantaneous drift (in our case, the spot alpha) is a function of a recurrent process, say $Z(t)$that visits any given point in its domain, say z, infinitely often. Thus, under recurrence, there is increasing local information around z that allows identification of the drift function at this value. A similar idea in our setting is to assume there exists functions $a : [0,1] \mapsto \mathbb{R}^M$ and $S : [0,1] \mapsto \mathbb{R}^{M \times M}$ such that the processes $\alpha(t)$ and $\Sigma(t)$ are generated by

$$\alpha(t) = a(t/T) \quad \text{and} \quad \Sigma(t) = S(t/T). \tag{21}$$

Then, the spot alpha would become a function of $Z(t) \equiv t/T \in [0,1]$; in particular, $Z_i \equiv Z(t_i)$, $i = 1, ..., n$, can be thought of as i.i.d. draws from the uniform distribution on $[0,1]$ with observations growing more and more dense in $[0,1]$ as $T \to \infty$. Thus, with this assumption, we would accomplish the same increase in local information about $\alpha(t)$ around a given point t in each successive model as $T \to \infty$. In contrast, the un-normalized time, $Z(t) = t$, is not a recurrent process, and so without the restriction given in Eq. (21), we would not be able to identify $\alpha(t)$.

Under the time normalization assumption in Eq. (21), there is a sequence of models as the sample changes (T increases), and this sequence of models is constructed so that the asymptotic distribution of $\hat{\alpha}(t)$ is well defined. While the time normalization is a widely used statistical tool to construct valid asymptotic distributions, and is used extensively in the large structural change literature, the restriction imposed in Eq. (21) is counterintuitive because the underlying economic structure changes as we sample over larger time spans. The time normalization is needed only to obtain formal asymptotic results for the conditional alpha estimators and not necessary for the asymptotic analysis of the conditional beta estimators (Theorem 1) or the long-run alpha and beta estimators developed in subsequent sections. As a consequence, we relegate the asymptotic theory for the conditional alpha estimators under the time normalization in Eq. (21) to Appendix C.

Finally, it is worth noting that one could alternatively analyze the conditional alpha and beta estimators in a discrete-time setting, where it is necessary to impose a time normalization. This approach is pursued in a previous working paper version of this paper (Ang and Kristensen, 2011) and in Kristensen (2011). The normalization is similar to Eq. (21), but in a discrete-time setting the normalization restriction has to be imposed on both the alphas and betas. This is due to the fact that, in contrast to the continuous-time setting where $\Omega(t) = \Sigma(t)/\sqrt{\Delta} \to \infty$ as we sample more frequently, the variance in the discrete-time model does not change as we collect more data over time. This mirrors the fact that in discrete time we rely only on long-span asymptotics, and so cannot nonparametrically learn about the local variation of the conditional alphas and betas without imposing some type of time normalization. However, as demonstrated in Ang and Kristensen (2011), estimators and finite-sample standard errors obtained in a discrete-time and continuous-time setting, respectively, are numerically identical. Thus, while the asymptotic theory is different, the empirical implementation is the same. The fact that the estimators can both be given a discrete-time and continuous-time interpretation is a convenient feature of the estimators because the vast majority of empirical studies are carried out in a discrete-time setting.

2.6 *Long-run alphas and betas*

To test the null of whether the long-run (LR) alphas are equal to zero [H_0 in Eq. (5)], we construct estimators of the long-run alphas and betas in Eq. (4). A natural way to estimate the long-run alphas and betas for stock k is to simply plug the pointwise kernel estimators into the expressions found in Eq. (4):

$$\hat{\alpha}_{\mathrm{LR},k} = \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_k\left(t_i\right) \quad \text{and} \quad \hat{\beta}_{\mathrm{LR},k} = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_k\left(t_i\right).$$

Given that we can identify the conditional spot betas, we can also identify the LR betas, and so $\hat{\beta}_{\mathrm{LR},k}$ is consistent. But more important, we can identify the LR alphas even if we cannot identify the instantaneous ones. In particular, without the time normalization given in Eq. (21), $\hat{\alpha}_k\left(t\right)$ is an inconsistent estimator of $\alpha_k\left(t\right)$, but $\hat{\alpha}_{\mathrm{LR},k}$ is a consistent estimator of $\alpha_{\mathrm{LR},k}$. Thus, we can consistenly estimate the long-run alphas without imposing the time-normalization used in the theoretical analysis of the conditional alphas. The intuition behind this feature is that our estimator of $\alpha_{\mathrm{LR},k}$ involves additional averaging over time. This averaging reduces the overall sampling error of $\hat{\alpha}_{\mathrm{LR},k}$ and enables consistency as $T \to \infty$.

Theorem 2 states the joint distribution of $\hat{\alpha}_{\mathrm{LR}} = (\hat{\alpha}_{\mathrm{LR},1}, ..., \hat{\alpha}_{\mathrm{LR},M})' \in \mathbb{R}^M$ and $\hat{\beta}_{\mathrm{LR}} = (\hat{\beta}_{\mathrm{LR},1}, ..., \hat{\beta}_{\mathrm{LR},M})' \in \mathbb{R}^{J\times M}$:

Theorem 2 *Assume that assumptions A.1–A.5 given in Appendix A hold. Then the long-run estimators satisfy as $T \to \infty$:*

$$\sqrt{T}(\hat{\alpha}_{\mathrm{LR}} - \alpha_{\mathrm{LR}}) \sim N\left(0, \Sigma_{\mathrm{LR},\alpha\alpha}\right), \quad \sqrt{n}(\hat{\beta}_{\mathrm{LR}} - \beta_{\mathrm{LR}}) \sim N\left(0, \Sigma_{\mathrm{LR},\beta\beta}\right) \tag{22}$$

in large samples, where

$$\begin{aligned}
\alpha_{\mathrm{LR}} &= \lim_{T\to\infty}\frac{1}{T}\int_0^T \alpha(t)\,dt \equiv\; E[\alpha(t)],\\
\beta_{\mathrm{LR}} &= \lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\,dt \equiv \mathrm{E}[\beta(t)],\\
\Sigma_{\mathrm{LR},\alpha\alpha} &= \lim_{T\to\infty}\frac{1}{T}\int_0^T \Sigma(t)\,dt \equiv \mathrm{E}[\Sigma(t)],\quad \textit{and}\\
\Sigma_{\mathrm{LR},\beta\beta} &= \lim_{T\to\infty}\frac{1}{T}\int_0^T \Lambda_{FF}^{-1}(t)\otimes\Sigma(t)\,dt \equiv \mathrm{E}\left[\Lambda_{FF}^{-1}(t)\otimes\Sigma(t)\right].
\end{aligned}$$

The long-run estimators converge at standard parametric rates $\sqrt{n}$ and $\sqrt{T}$, despite the fact that they are based on preliminary estimators that converge at slower, nonparametric rates. That is, inference of the long-run alphas and betas involves the standard Central Limit Theorem (CLT) convergence properties even though the point estimates of the conditional alphas and betas converge at slower rates. Intuitively, this is due to the additional smoothing taking place when we average over the preliminary estimates in Eq. (10). This occurs in other semiparametric estimators involving integrals of kernel estimators (see, for example, Newey and McFadden, 1994, Section 8; and Powell, Stock, and Stoker, 1989).

Consistent estimators of the asymptotic variances are obtained by simply plugging the point estimates of $\Lambda_{FF}(t)$ and $\Sigma(t)$ given in Eq. (17) into the sample versions of $\Sigma_{\mathrm{LR},\alpha\alpha}$ and $\Sigma_{\mathrm{LR},\beta\beta}$:

$$\hat{\Sigma}_{\mathrm{LR},\alpha\alpha} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Sigma}(t_i),\quad \Sigma_{\mathrm{LR},\beta\beta} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_{FF}^{-1}(t)\otimes\hat{\Sigma}(t_i).$$

We can test $H_0:\alpha_{\mathrm{LR}}=0$ by the following Wald-type statistic:

$$W_{\mathrm{LR}} = T\hat{\alpha}_{\mathrm{LR}}'\hat{\Sigma}_{\mathrm{LR},\alpha\alpha}^{-1}\hat{\alpha}_{\mathrm{LR}} \sim \chi_M^2 \quad \text{in large samples,} \tag{23}$$

as a direct consequence of Theorem 2. This is a conditional analogue of Gibbons, Ross, and Shanken (1989) and tests if long-run alphas are jointly equal to zero across all $k=1,...,M$ portfolios. A special case of Theorem 2 is Lewellen and Nagel (2006), who use a uniform kernel and the Fama and MacBeth (1973) procedure to compute standard errors of long-run estimators. Theorem 2 formally validates these procedures, and extend them to allow for general kernels, and joint tests across stocks, and tests for long-run betas.

Our model includes the case in which the factor loadings are constant with $\beta(t)=\beta\in\mathbb{R}^{J\times M}$ for all t. Under the null that the beta's are constant, $\beta(t)=\beta$, and with no heteroskedasticity, $\Sigma(t)=\Sigma$ for all t, the asymptotic distribution of $\hat{\alpha}_{\mathrm{LR}}$ is identical to the standard Gibbons, Ross, and Shanken (1989) test. This is shown in Appendix D. Thus, we pay no price asymptotically for the added robustness of our estimator. Furthermore, only in a setting where the factors are uncorrelated with the betas is the Gibbons, Ross and Shanken estimator of α_{LR} consistent. This is not surprising given the results of Jagannathan and Wang (1996) and others who show that in the presence of time-varying betas, OLS alphas do not yield estimates of conditional or long-run alphas.

2.7 *Tests for constancy of alphas and betas*

We wish to test for constancy of the potentially time-varying conditional alphas and betas. The two null hypotheses of interest are formally

$$\begin{aligned} H_k(\alpha) &: \alpha_k(t) = \alpha_k \in \mathbb{R}, \quad \text{for all } t \in [0,T], \\ H_k(\beta) &: \beta_k(t) = \beta_k \in \mathbb{R}^J, \quad \text{for all } t \in [0,T]. \end{aligned} \tag{24}$$

We propose to test each of the two hypotheses through Hausman-type statistics where we compare two estimators. The first is chosen to be consistent both under the relevant null and the alternative while the second one is consistent only under the null. If the null is true, the test statistic is expected to be small and vice versa. A natural choice for the former estimator is the nonparametric estimator developed in Subsection 2.2. For the latter, we use the long-run estimator because under the relevant null the long-run estimator is a consistent estimator of the constant coefficient. To be more precise, we define our test stastistics for $H_k(\alpha)$ and $H_k(\beta)$, respectively, as the following two weighted least-squares statistics:

$$W_k(\alpha) \equiv \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}_{kk}^{-2}(t_i)\left[\hat{\alpha}_k(t_i) - \hat{\alpha}_{\mathrm{LR},k}\right]^2,$$

and

$$W_k(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}_{kk}^{-2}(t_i)\left[\hat{\beta}_k(t_i) - \hat{\beta}_{\mathrm{LR},k}\right]'\hat{\Lambda}_{FF}(t_i)\left[\beta_k(t_i) - \hat{\beta}_{\mathrm{LR},k}\right]. \tag{25}$$

The weights have been chosen to ensure that the asymptotic distributions of the statistics are nuisance parameter-free. The two proposed test statistic are related to the generalized likelihood-ratio test statistics advocated in Fan, Zhang, and Zhang (2001).

The test statistic $W_k(\alpha)$ depends on $\hat{\alpha}_k(t)$, which is in general an inconsistent estimator of $\alpha_k(t)$ as discussed in Subsection 2.6. However, under the null of constant alphas, $\mathrm{E}[\hat{\alpha}_k(t)] \simeq \alpha_k(t)$ [c.f. Eq. (19)], and we are, therefore, able to derive a formal large-sample distribution of $W_k(\alpha)$. An important hypothesis nested within $H_k(\alpha)$ is the asset pricing hypothesis that $\alpha_{k,t} = 0$ for all $t \in [0, T]$ [$H_{0,k}$ in Eq. (6)]. This can be tested by simply setting $\hat{\alpha}_{\mathrm{LR},k} = 0$ in $W_k(\alpha)$ yielding:

$$W_k(0) \equiv \frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}_{kk}^{-2}(t_i)\, \hat{\alpha}_k^2(t_i)\,. \tag{26}$$

The proposed test statistics follow normal distributions in large samples as stated in Theorem 3.

Theorem 3 *Assume that assumptions A.1–A.5 given in Appendix A hold and the bandwidth satisfies A.6. Then,*

$$\begin{aligned} \textit{Under } H_k(\alpha) &: \quad \frac{W_k(\alpha) - m(\alpha)}{v(\alpha)} \sim N(0,1)\,, \\ \textit{Under } H_k(\beta) &: \quad \frac{W_k(\beta) - m(\beta)}{v(\beta)} \sim N(0,1) \end{aligned} \tag{27}$$

in large samples, where, with $(K * K)(z) \equiv \int K(y)\, K(z+y)\, dy$ *and* $J = \dim(f_i)$*,*

$$\begin{aligned} m(\alpha) &= \frac{\kappa_2}{Th} \textit{ and } v^2(\alpha) = \frac{2\int (K*K)^2(z)\, dz}{T^3 h}, \\ m(\beta) &= \frac{\kappa_2 J \Delta}{Th} \textit{ and } v^2(\beta) = \frac{2J\int (K*K)^2(z)\, dz}{n^2 T h}. \end{aligned}$$

For Gaussian kernels, $\int (K*K)^2(z)\, dz = 0.1995$ *and* $\kappa_2 = 0.2821$.

A convenient feature of the limiting distributions of the test statistics is that they are nuisance parameter–free because $m(\alpha)$, $m(\beta)$, $v(\alpha)$, and $v(\beta)$ depend on known quantities only. Moreover, Fan, Zhang, and Zhang (2001) demonstrate in a cross-sectional setting that test statistics of the form of $W_k(\alpha)$ and $W_k(\beta)$ are, in general, asymptotically optimal and can even be adaptively optimal, and so we expect them to be able to easily detect departures from the null. As a straightforward corollary of Theorem 3, one can show that the test statistic $W_k(0)$ has the same asymptotic distribution as $W_k(\alpha)$ and so is not affected by setting $\hat{\alpha}_{\mathrm{LR},k} = 0$.

The above test procedures can easily be adapted to construct joint tests of parameter constancy across multiple stocks. For example, to jointly test for constant alphas across all stocks, we would simply redefine the above least squares statistic to include alpha estimates across all stocks,

$$\bar{W}(\alpha) \equiv \frac{1}{n} \sum_{i=1}^{n} [\hat{\alpha}(t_i) - \hat{\alpha}_{\mathrm{LR}}]' \, \hat{\Sigma}^{-1}(t_i)\, [\hat{\alpha}(t_i) - \hat{\alpha}_{\mathrm{LR}}]\,.$$

The asymptotic distribution of this would be the same as for $W_k(\alpha)$, except that now $m(\alpha) = \kappa_2 M/(Th)$ and $v^2(\alpha) = 2M \int (K*K)^2(z)\, dz/(T^3 h)$ because we are testing M hypotheses jointly.

2.8 *Choice of kernel and bandwidth*

As is common to all nonparametric estimators, the kernel and bandwidth need to be selected. Our theoretical results are based on using a kernel centered around zero and our main empirical results use the Gaussian kernel. Other authors using high frequency data to estimate covariances or betas, such as Andersen, Bollerslev, Diebold and Wu (2006) and Lewellen and Nagel (2006), have used one-sided filters. For example, the rolling window estimator employed by Lewellen and Nagel corresponds to a uniform kernel on $[-1, 0]$ with $K(z) = \mathbb{I}\{-1 \leq z \leq 0\}$. For the estimator to be consistent, we have to let the sequence of bandwidths shrink toward zero as the sample size grows, $h \equiv h_n \to 0$ as $n \to \infty$ to remove any biases of the estimator.[7] However, a given sample requires a particular choice of h.

Because our interest lies in the in-sample estimation and testing, we advocate using two-sided symmetric kernels because in this case the bias from two-sided symmetric kernels is lower than for one-sided filters. In our data where n is over ten thousand daily observations, the improvement in the integrated root mean squared error (RMSE) using a Gaussian filter over a backward-looking uniform filter can be substantial. For the symmetric kernel the integrated RMSE is of order $O\left(n^{-2/5}\right)$, whereas the corresponding integrated RMSE is at most of order $O\left(n^{-1/3}\right)$ for a one-sided kernel. We provide further details in Appendix E.

Bias at end points is a well-known issue common to all kernel estimators. Symmetric kernels suffer from excess bias at the beginning and end of the sample. This can be handled in a number of different ways. The easiest way, which is also the procedure we follow in the empirical work, is to simply refrain from reporting estimates close to the two boundaries. All our theoretical results are established under the assumption that our sample has been observed across (normalized) time points $t \in [-c, T + c]$ for some $c > 0$ and we then estimate the alphas and betas only for $t \in [0, T]$. In the empirical work, we do not report the time-varying

[7] For very finely sampled data, especially intra day data, nonsynchronous trading could induce bias. A large literature exists on methods to handle nonsynchronous trading going back to Scholes and Williams (1977) and Dimson (1979). These methods can be employed in our setting. As an example, consider the one-factor model in which $f_t = R_{m,t}$ is the market return. As an ad hoc adjustment for nonsynchronous trading, we can augment the one-factor regression to include the lagged market return, $R_t = \alpha_t + \beta_{1,t} R_{m,t} + \beta_{2,t} R_{m,t-1} + \varepsilon_t$, and add the combined betas, $\hat{\beta}_t = \hat{\beta}_{1,t} + \hat{\beta}_{2,t}$. This is done by Li and Yang (2011). More recently, a literature has been growing on how to adjust for nonsynchronous effects in the estimation of realized volatility. Again, these can be carried over to our setting. For example, the methods proposed in, for example, Hayashi and Yoshida (2005) or Barndorff-Nielsen, Hansen, Lunde and Shephard (2009) can be adapted to our setting to adjust for biases due to non-synchronous observations. In our empirical work, non-synchronous trading should not be a major issue as we work with value-weighted, not equal-weighted, portfolios at the daily frequency.

alphas and betas during the first and last year of our post-1963 sample. Alternatively, adaptive estimators, such as boundary kernels and locally linear kernel estimators, that control for the boundary bias could be used. Usage of these estimators does not affect the asymptotic distributions in Theorems 1–3 or the asymptotic distributions we derive for long-run alphas and betas in Subsection 2.6.

Two bandwidths have to be chosen: One for the conditional estimators and another for the long-run estimators. The two different bandwidths are necessary because in our theoretical framework the conditional estimators and the long-run estimators converge at different rates. In particular, the asymptotic results suggest that for the integrated long-run estimators we need to undersmooth relative to the point-wise conditional estimates; that is, we should choose our long-run bandwidths to be smaller than the conditional bandwidths. Our strategy is to determine optimal conditional bandwidths and then adjust the conditional bandwidths for the long-run alpha and beta estimates. We propose data-driven rules for choosing the bandwidths. We conducted simulation studies showing that the proposed methods work well in practice.

2.8.1 *Bandwidth for conditional estimators*

To estimate the conditional bandwidths, we develop a global plug-in method that is designed to mimic the optimal, infeasible bandwidth. The bandwidth selection criterion is chosen as the integrated (across all time points) mean square error (MSE), and so the resulting bandwidth is a global one. In some situations, local bandwidth selection procedures that adapt to local features at a given time could be more useful. The following procedure can be adapted to this purpose by replacing all sample averages by subsample ones in the expressions.

For a symmetric kernel with $\int K(z)\,dz = \int K(z)\,z^2 dz = 1$, the optimal global bandwidth that minimizes the (integrated over $[0,T]$) MSE of $\hat{\beta}_k(t)$ is

$$h^*_{\beta,k} = \left(\frac{V_k(\beta)}{B_k(\beta)}\right)^{1/5} n^{-1/5}, \tag{28}$$

where $V_k(\beta) = \frac{1}{T}\int_0^T v_k(s;\beta)\,ds$ and $B_k(\beta) = \frac{1}{T}\int_0^T b_k^2(s;\beta)\,ds$ are the integrated time-varying variance and squared-bias components. Similarly, under the time normalization, the optimal bandwidth for the estimation of $\alpha_k(t)$ in terms of integrated MSE is

$$h^*_{\alpha,k} = \left(\frac{V_k(\alpha)}{B_k(\alpha)}\right)^{1/5} T^{-1/5}, \tag{29}$$

where $V_k(\alpha) = \frac{1}{T}\int_0^T v_k(s;\alpha)\,ds$ and $B_k(\alpha) = \frac{1}{T}\int_0^T b_k^2(s;\alpha)\,ds$ are the integrated time-varying variance and squared-bias components of $\hat{\alpha}_k(t)$. The functions appearing in the integrals are given by

$$\begin{aligned} v_k(t;\beta) &= \kappa_2 \Lambda_{FF}^{-1}(t)\,\sigma_{kk}^2(t) \quad \text{and} \quad b_k(t;\beta) = \beta_k^{(2)}(t)\,; \\ v_k(t;\alpha) &= \kappa_2 \sigma_{kk}^2(t) \quad \text{and} \quad b_k(t;\alpha) = \alpha_k^{(2)}(t)\,. \end{aligned}$$

Ideally, we would compute v_k and b_k to obtain the optimal bandwidth given in eqs. (28)-(29). However, these depend on unknown components, α, β, Λ_{FF}, and Σ. To implement the bandwidth choice we propose a two-step method to provide preliminary estimates of these unknown quantities.[8] Because the proposed procedures for choosing the bandwidth choice for $\hat{\beta}(t)$ and $\hat{\alpha}(t)$ follow along the same lines, we describe only the one for $\hat{\beta}(t)$.

1. Choose as prior $\Lambda_{FF}(t) = \Lambda$ and $\sigma_{kk}(t) = \sigma_{kk}$ being constants, and $\beta_k(t) = b_{k0} + b_{k1}t + ... + b_{kp}t^p$ a polynomial of order $p \geq 2$. We obtain parametric least-squares estimates $\tilde{\Lambda}_{FF}$, $\tilde{\sigma}_{kk}^2$ and $\tilde{\beta}_k(t) = \tilde{b}_{k0} + \tilde{b}_{k1}t + ... + \tilde{b}_{kp}t^p$. Compute for each stock ($k = 1, ..., M$)

   $$\tilde{V}_k(\beta) = \frac{\kappa_2}{T}\tilde{\Lambda}_{FF}^{-1}\tilde{\sigma}_{kk}^2 \quad \text{and} \quad \tilde{B}_k(\beta) = \frac{1}{n}\sum_{i=1}^{n} ||\tilde{\beta}_k^{(2)}(t_i)||^2,$$

   where $\tilde{\beta}_{k,t}^{(2)} = 2\tilde{b}_{k2} + 6\tilde{b}_{k3}(t/n) + ... + p(p-1)\,\tilde{b}_{kp}(t/n)^{p-2}$. Then, using these estimates we compute the first-pass bandwidth

   $$\tilde{h}_k = \left[\frac{\tilde{V}_k(\beta)}{\tilde{B}_k(\beta)}\right]^{1/5} \times n^{-1/5}. \tag{30}$$

2. Given $\tilde{h}_k$, compute the kernel estimators $\hat{\beta}_k(t)$ and the variance components given in Eq. (17) with $h_k = \tilde{h}_k$. Use these to compute

   $$\hat{V}_k(\beta) = \kappa_2 \frac{1}{n}\sum_{i=1}^{n} \hat{\Lambda}_{FF}^{-1}(t_i)\,\hat{\sigma}_{kk}^2(t_i) \quad \text{and} \quad \hat{B}_k(\beta) = \frac{1}{n}\sum_{i=1}^{n} ||\hat{\beta}_k^{(2)}(t_i)||^2,$$

   with $\hat{\beta}_k^{(2)}(t)$ being the second derivative of the kernel estimator. These are in turn used to obtain the second-pass bandwidth:

   $$\hat{h}_k = \left[\frac{\hat{V}_k(\beta)}{\hat{B}_k(\beta)}\right]^{1/5} \times n^{-1/5}. \tag{31}$$

[8] Ruppert, Sheather, and Wand (1995) discuss in detail how this can done in a standard kernel regression framework. This bandwidth selection procedure takes into account the (time-varying) correlation structure between betas and factors through Λ_t and Ω_t.

We compute conditional alphas and betas using the bandwidths obtained by the described two-step procedure.

Our motivation for using a plug-in bandwidth is as follows. We believe that the betas for our portfolios vary slowly and smoothly over time as argued both in economic models such as Gomes, Kogan, and Zhang (2003) and from previous empirical estimates such as Petkova and Zhang (2005), Lewellen and Nagel (2006), Ang and Chen (2007), and others. The plug-in bandwidth accommodates this prior information by allowing us to specify a low-level polynomial order. In our empirical work we choose a polynomial of degree $p = 6$ and find little difference in the choice of bandwidths when p is below ten.

One could alternatively use cross-validation (CV) procedures to choose the bandwidth. These procedures are completely data driven and, in general, yield consistent estimates of the optimal bandwidth. However, we find that in our data these can produce bandwidths that are extremely small, corresponding to a time window as narrow as three-to-five days with corresponding huge time variation in the estimated factor loadings. We believe these bandwidth choices are not economically sensible. The poor performance of the CV procedures is likely due to a number of factors. First, it is wellknown that cross-validated bandwidths could exhibit very inferior asymptotic and practical performance even in a cross-sectional setting (see, for example, Härdle, Hall, and Marron, 1988). This problem is further enhanced when CV procedures are used on time series data as found in various studies (Diggle and Hutchinson, 1989; Hart, 1991; and Opsomer, Wang, and Yang, 2001).

2.8.2 *Bandwidth for long-run estimators*

To estimate the long-run alphas and betas we re-estimate the conditional coefficients by undersmoothing relative to the bandwidth in Eq. (31). The reason for this is that the long-run estimates are themselves integrals and the integration imparts additional smoothing. Using the same bandwidth as for the conditional alphas and betas results in over-smoothing.

Ideally, we would choose optimal long-run bandwidths to minimize the MSE's $\mathrm{E}\left[(\hat{\alpha}_{\mathrm{LR},k} - \alpha_{\mathrm{LR},k})^2\right]$ and $\mathrm{E}[(\hat{\beta}_{\mathrm{LR},k} - \beta_{\mathrm{LR},k})^2]$, which we derive in Appendix F. As demonstrated there, the bandwidths used for the long-run estimators should be chosen to be of order $h_{\mathrm{LR},k} = O\left(n^{-1/3}\right)$ and $h_{\mathrm{LR},k} = O\left(T^{-1/3}\right)$ for the long-run betas and alphas, respectively. Thus, the optimal bandwidth for the long-run estimates is required to shrink at a faster rate than the one used for pointwise estimates above.

In our empirical work, we select the bandwidth for the long-run alphas and betas by first computing the optimal second-pass conditional bandwidth $\hat{h}_k$ in Eq. (31) and then scaling this down by setting

$$\hat{h}_{\mathrm{LR},k} = \hat{h}_k \times n^{-2/15}. \tag{32}$$

3 Data

In our empirical work, we consider two specifications of conditional factor models: a conditional CAPM with a single factor, which is the market excess return, and a conditional version of the Fama and French (1993) model with the three factors being the market excess return (MKT) and two zero-cost mimicking portfolios (a size factor, SMB, and a value factor, HML).

We apply our methodology to decile portfolios sorted by book-to-market ratios and decile portfolios sorted on past returns constructed by Kenneth French. We use the Fama and French (1993) factors MKT, SMB, and HML as explanatory factors. All our data are at the daily frequency from July 1963 to December 2007, and we choose to measure time in days such that $\Delta = 1$. We use this whole span of data to compute optimal bandwidths. However, in reporting estimates of conditional factor models we truncate the first and last years of daily observations to avoid endpoint bias, so our conditional estimates of alphas and factor loadings and our estimates of long-run alphas and betas span July 1964 to December 2006. Our summary statistics in Table 1 cover this truncated sample, as do all of our results in the next sections.

Panel A of Table 1 reports annualized means and standard deviations of our factors. The market premium is 5.32% compared with a small size premium for SMB at 1.84% and a value premium for HML at 5.24%. Both SMB and HML are negatively correlated with the market portfolio with correlations of -23% and -58%, respectively, but have a low correlation with each other of only -6%. In Panel B, we list summary statistics of the book-to-market and momentum decile portfolios. We also report OLS estimates of a constant alpha and constant beta in the last two columns using the market excess return factor. The book-to-market portfolios have average excess returns of 3.84% for growth stocks (decile 1) to 9.97% for value stocks (decile 10). We refer to the zero-cost strategy 10–1 that goes long value stocks and shorts growth stocks as the book-to-market strategy. The book-to-market strategy has an average return of 6.13%, an OLS alpha of 7.73% and a negative OLS beta of -0.301. Similarly, for the momentum portfolios we refer to a 10–1 strategy that goes long past winners (decile 10) and goes short past losers (decile 1) as the momentum strategy. The momentum strategy's returns are particularly impressive with a mean of 17.07% and an OLS alpha of 16.69%. The momentum strategy has an OLS beta

Table 1:

Summary statistics of factors and portfolios

Notes. We report summary statistics of Fama and French (1993) factors and book-to-market and momentum portfolios in Panel A and B. Data is at a daily frequency and spans July 1964–December 2006 and are from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. We annualize means and standard deviations by multiplying daily estimates by 252 and $\sqrt{252}$, respectively. The portfolio returns are in excess of the daily Ibbotson risk-free rate except for the 10-1 book-to-market and momentum strategies which are simply differences between portfolio 10 and portfolio 1. The last two columns in Panel B report OLS estimates of constant alphas ($\hat{\alpha}_{OLS}$) and betas ($\hat{\beta}_{OLS}$).

Panel A: Factors

			Correlations		
Factor	Mean	Standard deviation	MKT	SMB	HML
MKT	0.0532	0.1414	1.0000	-0.2264	-0.5821
SMB	0.0184	0.0787	-0.2264	1.0000	-0.0631
HML	0.0524	0.0721	-0.5812	-0.0631	1.0000

Panel B: Portfolios

			OLS estimates	
Portfolio	Mean	Standard deviation	$\hat{\alpha}_{OLS}$	$\hat{\beta}_{OLS}$
Book-to-Market				
1 Growth	0.0384	0.1729	-0.0235	1.1641
2	0.0525	0.1554	-0.0033	1.0486
3	0.0551	0.1465	0.0032	0.9764
4	0.0581	0.1433	0.0082	0.9386
5	0.0589	0.1369	0.0121	0.8782
6	0.0697	0.1331	0.0243	0.8534
7	0.0795	0.1315	0.0355	0.8271
8	0.0799	0.1264	0.0380	0.7878
9	0.0908	0.1367	0.0462	0.8367
10 Value	0.0997	0.1470	0.0537	0.8633
10-1 Book-to-market strategy	0.0613	0.1193	0.0773	-0.3007
Momentum				
1 Losers	-0.0393	0.2027	-0.1015	1.1686
2	0.0226	0.1687	-0.0320	1.0261
3	0.0515	0.1494	0.0016	0.9375
4	0.0492	0.1449	-0.0001	0.9258
5	0.0355	0.1394	-0.0120	0.8934
6	0.0521	0.1385	0.0044	0.8962
7	0.0492	0.1407	0.0005	0.9158
8	0.0808	0.1461	0.0304	0.9480
9	0.0798	0.1571	0.0256	1.0195
10 Winners	0.1314	0.1984	0.0654	1.2404
10-1 Momentum Strategy	0.1707	0.1694	0.1669	0.0718

close to zero of 0.072.

We first examine the conditional and long-run alphas and betas of the book-to-market portfolios and the book-to-market strategy in Section 4. Then, we test the conditional Fama and French (1993) model on the momentum portfolios in Section 5.

4 Portfolios sorted on book-to-market ratios

For portfolios sorted on book-to-market ratios, we first test the conditional CAPM model, then examine the time-variation in the conditional betas and finally test the Fama and French (1993) three-factor model.

4.1 *Tests of the conditional CAPM*

We report estimates of bandwidths, conditional alphas and betas, and long-run alphas and betas in Table 2 for the decile book-to-market portfolios. The last row contains results for the 10-1 book-to-market strategy. The columns labeled "Bandwidth" list the second-pass bandwidth $\hat{h}_{k,2}$ in Eq. (31). The column headed "Fraction" reports the bandwidths as a fraction of the entire sample, which is equal to one. In the column titled "Months" we transform the bandwidth to a monthly equivalent unit. For the normal distribution, 95% of the mass lies between $(-1.96, 1.96)$. If we were to use a flat uniform distribution, 95% of the mass would lie between $(-0.975, 0.975)$. Thus, to transform to a monthly equivalent unit we multiply by $533 \times 1.96/0.975$, where there are 533 months in the sample. We annualize the alphas in Table 2 by multiplying the daily estimates by 252.

For the decile 8–10 portfolios, which contain predominantly value stocks, and the value-growth strategy 10–1, the optimal bandwidth is around 20 months.

For these portfolios, significant time variation in beta exists and the relatively tighter windows allow this variation to be picked up with greater precision. In contrast, growth stocks in deciles 1–2 have optimal windows of 51 and 106 months, respectively. Growth portfolios do not exhibit much variation in beta so the window estimation procedure picks a much longer bandwidth. Overall, our estimated bandwidths are somewhat longer than the commonly used 12-month horizon to compute betas using daily data (see, for example, Ang, Chen, and Xing, 2006). At the same time, our 20-month window is shorter than the standard 60-month window often used at the monthly frequency (see, for example, Fama and French, 1993, 1997).

Table 2:

Alphas and Betas of Book-to-Market Portfolios

Notes. The table reports conditional bandwidths [$\hat{h}_k$ in Eq. (31)] and various statistics of conditional and long-run alphas and betas from a conditional capital asset pricingmodel of the book-to-market portfolios. The bandwidths are reported in fractions of the entire sample, which corresponds to one, and in monthly equivalent units. We transform the fraction to a monthly equivalent unit by multiplying $533 \times 1.96/0.975$, where there are 533 months in the sample, and the intervals $(-1.96, 1.96)$ and $(-0.975, 0.975)$ correspond to cumulative probabilities of 95% for the unscaled normal and uniform kernel, respectively. The conditional alphas and betas are computed at the end of each calendar month, and we report the standard deviations of the monthly conditional beta estimates following Theorem 1 using the conditional bandwidths in the columns labeled "Bandwidth." The long-run estimates, with standard errors in parentheses, are computed following Theorem 2 and average daily estimates of conditional alphas and betas. The long-run bandwidths apply the transformation in Eq. (32) with $n = 11202$ days. Both the conditional and the long-run alphas are annualized by multiplying by 252. The joint test for long-run alphas equal to zero is given by the Wald test statistic in Eq. (23). The full data sample is from July 1963 to December 2007, but the conditional and long-run estimates span July 1964 to December 2006 to avoid the bias at the endpoints.

	Bandwidth			Long-Run Estimates	
Portfolio	Fraction	Months	Standard deviation of conditional betas	Alpha	Beta
1 Growth	0.0474	50.8	0.0558	-0.0226	1.1705
				(0.0078)	(0.0039)
2	0.0989	105.9	0.0410	-0.0037	1.0547
				(0.0069)	(0.0034)
3	0.0349	37.4	0.0701	0.0006	0.9935
				(0.0072)	(0.0034)
4	0.0294	31.5	0.0727	0.0043	0.9467
				(0.0077)	(0.0035)
5	0.0379	40.6	0.0842	0.0077	0.8993
				(0.0083)	(0.0039)
6	0.0213	22.8	0.0871	0.0187	0.8858
				(0.0080)	(0.0038)
7	0.0188	20.1	0.1144	0.0275	0.8767
				(0.0084)	(0.0038)
8	0.0213	22.8	0.1316	0.0313	0.8444
				(0.0082)	(0.0039)
9	0.0160	17.2	0.1497	0.0373	0.8961
				(0.0094)	(0.0046)
10 Value	0.0182	19.5	0.1911	0.0461	0.9556
				(0.0112)	(0.0055)
10–1 Book-to-market strategy	0.0217	23.3	0.2059	0.0681	-0.2180
				(0.0155)	(0.0076)

Joint test for $\alpha_{LR,i} = 0$, $i = 1, ..., 10$.
Wald statistic $W = 31.6$, and p-value = 0.0005.

We report the standard deviation of conditional betas at the end of each month. Below, we further characterize the time variation of these monthly conditional estimates. The conditional betas of the book-to-market strategy have a standard deviation of 0.206. The majority of this time variation comes from value stocks, as decile 1 betas have a standard deviation of only 0.056, while decile 10 betas have a standard deviation of 0.191.

Lewellen and Nagel (2006) argue that the magnitude of the time variation of conditional betas is too small for a conditional CAPM to explain the value premium. The estimates in Table 2 overwhelmingly confirm this. Lewellen and Nagel suggest that an approximate upper bound for the unconditional OLS alpha of the book-to-market strategy, which Table 1 reports as 0.644% per month or 7.73% per annum, is given by $\sigma_\beta \times \sigma_{\mathrm{E}_t[r_{m,t+1}]}$, where σ_β is the standard deviation of conditional betas and $\sigma_{\mathrm{E}_t[r_{m,t+1}]}$ is the standard deviation of the conditional market risk premium. Conservatively assuming that $\sigma_{\mathrm{E}_t[r_{m,t+1}]}$ is 0.5% per month following Campbell and Cochrane (1999), we can explain at most $0.206 \times 0.5 = 0.103\%$ per month or 1.24% per annum of the annual 7.73% book-to-market OLS alpha. We now formally test for this result by computing long-run alphas and betas.

In the last two columns of Table 2, we report estimates of long-run annualized alphas and betas, along with standard errors in parentheses. The long-run alpha of the growth portfolio is -2.26% with a standard error of 0.008 and the long-run alpha of the value portfolio is 4.61% with a standard error of 0.011. Thus, both growth and value portfolios overwhelmingly reject the conditional CAPM. The long-run alpha of the book-to-market portfolio is 6.81% with a standard error of 0.015. Clearly, there is a significant long-run alpha after controlling for time-varying market betas. The long-run alpha of the book-to-market strategy is very similar to, but not precisely equal to, the difference in long-run alphas between the value and growth deciles because of the different smoothing parameters applied to each portfolio. There is no monotonic pattern for the long-run betas of the book-to-market portfolios, but the book-to-market strategy has a significantly negative long-run beta of -0.218 with a standard error of 0.008.

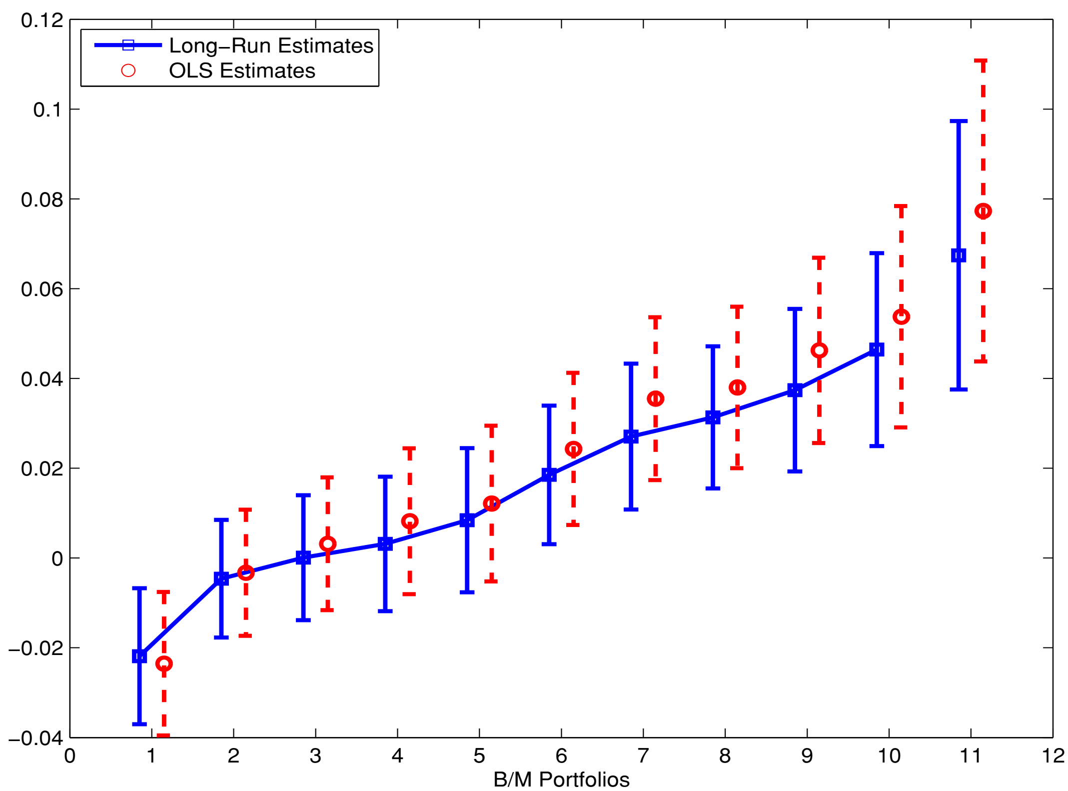
We test if the long-run alphas across all ten book-to-market portfolios are equal to zero using the Wald test of Eq. (23). The Wald test statistic is 31.6 with a p-value less than 0.001. Thus, the book-to-market portfolios overwhelmingly reject the null of the conditional CAPM with time-varying betas.

Fig. 1 compares the long-run alphas with OLS alphas. We plot the long-run alphas using squares with 95% confidence intervals displayed in the solid error bars. The point estimates of the OLS alphas are plotted as circles with 95% confidence intervals in dashed lines. Portfolios

1–10 on the x-axis represent the growth to value decile portfolios. Portfolio 11 is the book-to-market strategy. The spread in OLS alphas is greater than the spread in long-run alphas, but the standard error bands are very similar for both the long-run and OLS estimates, despite our procedure being nonparametric. For the book-to-market strategy, the OLS alpha is 7.73% compared with a long-run alpha of 6.81%. Thus accounting for time-varying betas has reduced the OLS alpha by approximately only 1.1%.

Figure 1:
Long-run alphas versus ordinary-least squares alphas in the conditional capital asset pricing model for the book-to-market portfolios

Notes. We plot long-run alphas implied by a conditional CAPM and OLS alphas for the book-to-market portfolios. We plot the long-run alphas using squares with 95% confidence intervals displayed by the solid error bars. The point estimates of the OLS alphas are plotted as circles with 95% confidence intervals in dashed lines. Portfolios 1–10 on the x-axis represent the growth to value decile portfolios. Portfolio 11 is the book-to-market strategy, which goes long portfolio 10 and short portfolio 1. The long-run conditional and OLS alphas are annualized by multiplying by 252.



4.2 *Time variation of conditional betas*

We now characterize the time variation of conditional betas from the one-factor market model. We begin by testing for constant conditional alphas and betas using the Wald test of Theorem 3. Table 3 shows that for all book-to-market portfolios, we fail to reject the hypothesis that the conditional alphas are constant, with Wald statistics that are far below the 95% critical values. This does not mean that the conditional alphas are equal to zero, as we estimate a highly significant long-run alpha of the book-to-market strategy and reject that the long-run alphas are jointly equal to zero across book-to-market portfolios. In contrast, we reject the null that the conditional betas are constant with p-values that are effectively zero.

Table 3:

Tests of constant conditional alphas and betas of book-to-market portfolios

Notes. We test for constancy of the conditional alphas and betas in a conditional capital asset pricing model using the Wald test of Theorem 3. In the columns labeled "Alpha" ("Beta") we test the null that the conditional alphas (betas) are constant. We report the test statistic W in Theorem 3 and 95% and 99% critical values of the asymptotic distribution. We mark rejections at the 99% level with **.

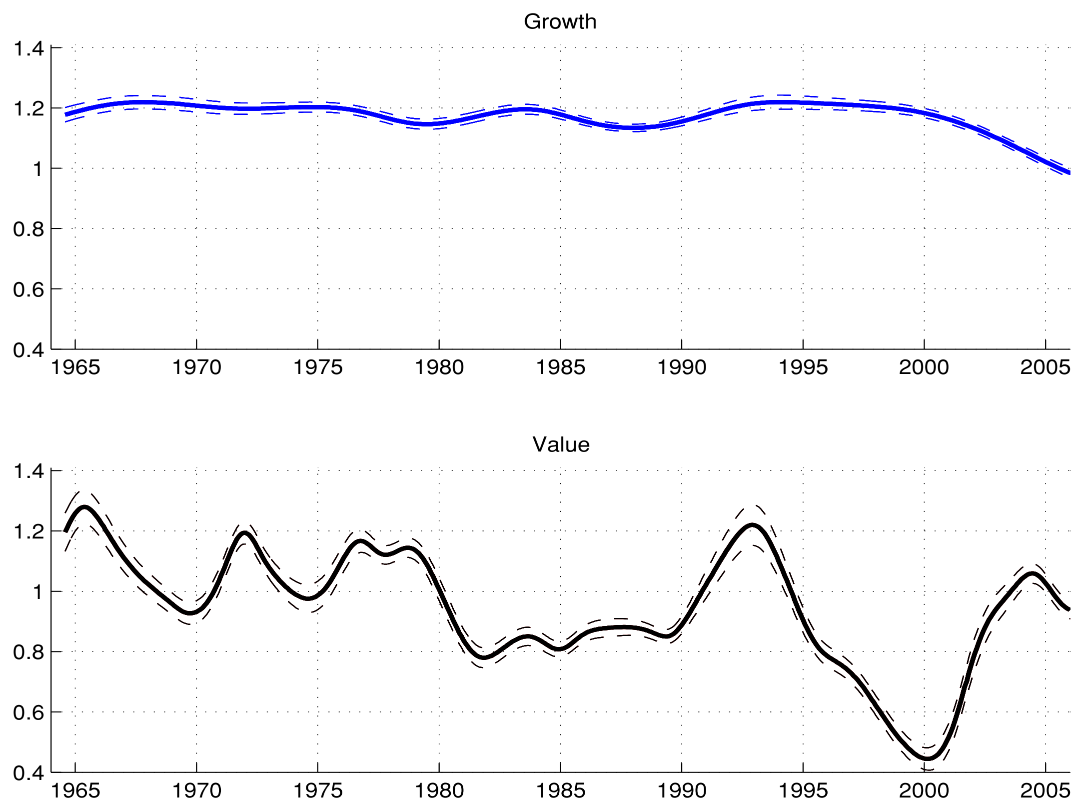
	W		Critical values	
Portfolio	Alpha	Beta	95%	99%
1 Growth	49	424^{**}	129	136
2	9	331^{**}	65	71
3	26	425^{**}	172	180
4	47	426^{**}	202	211
5	30	585^{**}	159	167
6	50	610^{**}	276	286
7	75	678^{**}	311	322
8	70	756^{**}	276	286
9	84	949^{**}	361	373
10 Value	116	1028^{**}	320	331
10-1 Book-to-Market Strategy	114	830^{**}	270	280

Fig. 2 charts the annualized estimates of conditional betas for the growth (decile 1) and value (decile 10) portfolios at a monthly frequency. Conditional factor loadings are estimated relatively precisely with tight 95% confidence bands, shown in dashed lines. Growth betas are largely constant around 1.2, except after 2000, when growth betas decline to around one. In contrast, conditional betas of value stocks are much more variable, ranging from close to 1.3 in 1965 and around 0.45 in 2000. From this low, value stock betas increase to around 1 at the end of the sample. We attribute the low relative returns of value stocks in the late 1990s to the low betas of value stocks at this time.

Figure 2:

Conditional betas of growth and value portfolios

Notes. The figure shows monthly estimates of conditional conditional betas from a conditional capital asset pricing model of the first and tenth decile book-to-market portfolios (growth and value, respectively). We plot 95% confidence bands in dashed lines. The conditional alphas are annualized by multiplying by 252.

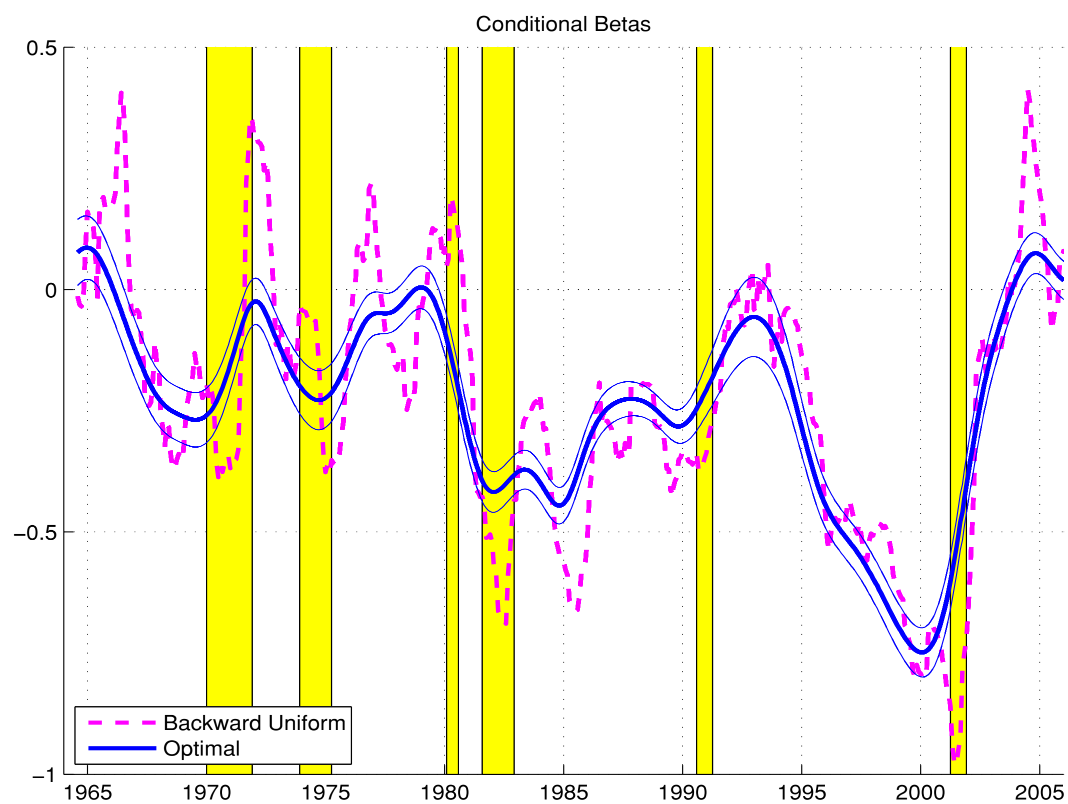


In Fig. 3, we plot betas of the book-to-market strategy, which is the difference in returns between deciles 10 and 1 (value minus growth). Because the conditional betas of growth stocks are fairly flat, almost all of the time variation of the conditional betas of the book-to-market strategy is driven by the conditional betas of the decile 10 value stocks. Fig. 3 also overlays estimates of conditional betas from a backward-looking, flat 12-month filter. Similar filters are employed by Andersen, Bollerslev, Diebold and Wu (2006) and Lewellen and Nagel (2006). Not surprisingly, the 12-month uniform filter produces estimates with larger conditional variation. Some of this conditional variation is smoothed away by using the longer bandwidths of our optimal estimators.[9] However, the unconditional variation over the whole sample of the uniform filter estimates and the optimal estimates are similar. For example, the standard deviation of end-of-month conditional beta estimates from the uniform filter is 0.276, compared with 0.206 for the optimal two-sided conditional beta estimates. This implies that the Lewellen and Nagel (2006) analysis using backward-looking uniform filters is conservative. Using our optimal estimators reduces the overall volatility of the conditional betas, making it even more unlikely that the value premium can be explained by time-varying market factor loadings.

Figure 3:

Conditional betas of the book-to-market strategy

Notes. The figure shows monthly estimates of conditional conditional betas of the book-to-market strategy. We plot the optimal estimates in bold solid lines along with 95% confidence bands in regular solid lines. We also overlay the backward one-year uniform estimates in dashed lines. National Bureau of Economic Research designated recession periods are shaded in horizontal bars.



Several authors such as Jagannathan and Wang (1996) and Lettau and Ludvigson (2001b)

[9] The standard error bands of the uniform filters (not shown) are much larger than the standard error bands of the optimal estimates.

argue that value stock betas increase during times when risk premia are high, causing value stocks to carry a premium to compensate investors for bearing this risk. Theoretical models of risk predict that betas on value stocks should vary over time and be highest during times when marginal utility is high (see, for example, Gomes, Kogan, and Zhang, 2003; and Zhang, 2005). We investigate how betas move over the business cycle in Table 4 where we regress conditional betas of the value-growth strategy onto various macro factors. Kanaya and Kristensen (2010) provide theoretical justification for this two-step procedure in a continuous-time setting.

In Table 4, we find only weak evidence that the book-to-market strategy betas increase during bad times. Regressions I–IX examine the covariation of conditional betas with individual macro factors known to predict market excess returns. When dividend yields are high, the market risk premium is high, and Regression I shows that conditional betas covary positively with dividend yields. However, this is the only variable that has a significant coefficient with the correct sign. When bad times are proxied by high default spreads, high short rates, or high market volatility, conditional betas of the book-to-market strategy tend to be lower. During National Bureau of Economic Research designated recessions conditional betas also move the wrong way and tend to be lower. The industrial production, term spread, the Lettau and Ludvigson (2001a) cay, and inflation regressions have insignificant coefficients. The industrial production coefficient also has the wrong predicted sign.

In Regression X, we find that book-to-market strategy betas do have significant covariation with many macro factors. This regression has an impressive adjusted R^2 of 55%. Except for the positive and significant coefficient on the dividend yield, the coefficients on the other macro variables: the default spread, industrial production, short rate, term spread, market volatility, and cay are either insignificant or have the wrong sign, or both. In Regression XI, we perform a similar exercise to Petkova and Zhang (2005). We first estimate the market risk premium by running a first-stage regression of excess market returns over the next quarter onto the instruments in Regression X measured at the beginning of the quarter. We define the market risk premium as the fitted value of this regression at the beginning of each quarter. We find that, in Regression XI, there is small positive covariation of conditional betas of value stocks with these fitted market risk premia with a coefficient of 0.37 and a standard error of 0.18. But, the adjusted R^2 of this regression is only 0.06. This is smaller than the covariation that Petkova and Zhang (2005) find because they specify betas as linear functions of the same state variables that drive the time variation of market risk premia. In summary, although conditional betas do covary with macro variables, little evidence exists that betas of value stocks are higher during

Table 4:
Characterizing conditional betas of the value-growth strategy

Notes. We regress conditional betas of the value-growth strategy onto various macro variables. The betas are computed from a conditional capital asset pricing model and are plotted in Fig. 3. The dividend yield is the sum of past twelve-month dividends divided by current market capitalization of the Centre for Research in Security Prices (CRSP) value-weighted market portfolio. The default spread is the difference between BAA and ten-year Treasury yields. Industrial production is the log year-on-year change in the industrial production index. The short rate is the three-month T-bill yield. The term spread is the difference between ten-year Treasury yields and three-month T-bill yields. Market volatility is defined as the standard deviation of daily CRSP value-weighted market returns over the past month. We denote the Lettau-Ludvigson (2001a) cointegrating residuals of consumption, wealth, and labor from their long-term trend as cay. Inflation is the log year-on-year change of the CPI index. The National Bureau of Economic Research (NBER) designated recession variable is a zero-one indicator that takes on the value one if the NBER defines a recession that month. All right-hand-side variables are expressed in annualized units. All regressions are at the monthly frequency except regressions VII and XI which are at the quarterly frequency. The market risk premium is constructed in a regression of excess market returns over the next quarter on dividend yields, default spreads, industrial production, short rates, industrial production, short rates, term spreads, market volatility, and cay. The instruments are measured at the beginning of the quarter. We define the market risk premium as the fitted value of this regression at the beginning of each quarter. Robust standard errors are reported in parentheses and we denote 95% and 99% significance levels with * and **, respectively. The data sample is from July 1964 to December 2006.

Regressor	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Dividend yield	4.55									16.5	
	(2.01)*									(2.95)**	
Default spread		-9.65								-1.86	
		(2.14)**								(3.68)	
Industrial production			0.50							0.18	
			(0.21)							(0.33)	
Short rate				-1.83						-7.33	
				(0.50)**						(1.22)**	
Term spread					1.08					-3.96	
					(1.20)					(2.10)	
Market volatility						-1.38				-0.96	
						(0.38)**				(0.40)*	
cay							0.97			-0.74	
							(1.12)			(1.31)	
Inflation								1.01			
								(0.55)			
NBER recession									-0.07		
									(0.03)*		
Market risk premium											0.37
											(0.18)*
Adjusted R^2	0.06	0.09	0.01	0.06	0.01	0.15	0.02	0.02	0.01	0.55	0.06

times when the market risk premium is high.

4.3 *Tests of the conditional Fama and French (1993) model*

We now investigate the performance of a conditional version of the Fama and French (1993) model estimated on the book-to-market portfolios and the book-to-market strategy. Table 5 reports long-run alphas and factor loadings. After controlling for the Fama and French factors with time-varying factor loadings, the long-run alphas of the book-to-market portfolios are still significantly different from zero and are positive for growth stocks and negative for value stocks. The long-run alphas monotonically decline from 2.16% for decile 1 to -1.58% for decile 10. The book-to-market strategy has a long-run alpha of -3.75% with a standard error of 0.010. The joint test across all ten book-to-market portfolios for the long-run alphas equal to zero decisively rejects with a p-value of zero. Thus, the conditional Fama and French (1993) model is overwhelmingly rejected.

Table 5 shows that long-run market factor loadings have only a small spread across growth to value deciles, with the book-to-market strategy having a small long-run market loading of 0.191. In contrast, the long-run SMB loading is relatively large at 0.452 and would be zero if the value effect were uniform across stocks of all sizes. Value stocks have a small size bias (see Loughran, 1997), and this is reflected in the large long-run SMB loading. We expect, and find, that long-run HML loadings increase from -0.672 for growth stocks to 0.792 for value stocks, with the book-to-market strategy having a long-run HML loading of 1.466. The previously positive long-run alphas for value stocks under the conditional CAPM become negative under the conditional Fama and French model. The conditional Fama and French model overcompensates for the high returns for value stocks by producing SMB and HML factor loadings that are relatively too large, leading to a negative long-run alpha for value stocks.

In Table 6, we conduct constancy tests of the conditional alphas and factor loadings. We fail to reject that the conditional alphas are constant for all book-to-market portfolios. Whereas the conditional betas exhibited large time variation in the conditional CAPM, we now cannot reject that the conditional market factor loadings are constant for decile portfolios 3–9. However, the extreme growth and value deciles do have time-varying MKT betas. Table 6 reports rejections at the 99% level that the SMB loadings and HML loadings are constant for the extreme growth and value deciles. For the book-to-market strategy, we find strong evidence that the SMB and HML loadings vary over time, especially for the HML loadings. Consequently, the time variation of conditional betas in the one-factor model is now absorbed by time-varying SMB

Table 5:

Long-run Fama and French (1993) alphas and factor loadings of book-to-market portfolios

Notes. The table reports estimates of long-run alphas and factor loadings from a conditional Fama and French (1993) model applied to decile book-to-market portfolios and the 10–1 book-to-market strategy. The long-run estimates, with standard errors in parentheses, are computed following Theorem 2 and average daily estimates of conditional alphas and betas. The long-run alphas are annualized by multiplying by 252. The joint test for long-run alphas equal to zero is given by the Wald test statistic W_0 in Eq. (23). The full data sample is from July 1963 to December 2007, but the long-run estimates span July 1964 to December 2006 to avoid the bias at the endpoints.

Portfolio	Alpha	MKT	SMB	HML
1 Growth	0.0216	0.9763	-0.1794	-0.6721
	(0.0056)	(0.0041)	(0.0060)	(0.0074)
2	0.0123	0.9726	-0.0634	-0.2724
	(0.0060)	(0.0043)	(0.0063)	(0.0079)
3	0.0072	0.9682	-0.0228	-0.1129
	(0.0067)	(0.0048)	(0.0072)	(0.0087)
4	-0.0057	0.9995	0.0163	0.1584
	(0.0072)	(0.0050)	(0.0073)	(0.0093)
5	-0.0032	0.9668	0.0005	0.2567
	(0.0075)	(0.0053)	(0.0079)	(0.0099)
6	-0.0025	0.9821	0.0640	0.3022
	(0.0072)	(0.0051)	(0.0077)	(0.0094)
7	-0.0113	1.0043	0.0846	0.4294
	(0.0071)	(0.0050)	(0.0074)	(0.0091)
8	-0.0153	1.0352	0.1061	0.7053
	(0.0057)	(0.0041)	(0.0061)	(0.0075)
9	-0.0153	1.1011	0.1353	0.7705
	(0.0069)	(0.0049)	(0.0074)	(0.0090)
10 Value	-0.0158	1.1667	0.2729	0.7925
	(0.0090)	(0.0064)	(0.0095)	(0.0118)
10-1 Book-to-market strategy	-0.0375	0.1911	0.4521	1.4660
	(0.0103)	(0.0073)	(0.0108)	(0.0133)

Joint test for $\alpha_{LR,i} = 0, i = 1, ..., 10.$
Wald statistic $W_0 = 77.5$, and p-value = 0.0000.

and HML loadings in the conditional Fama and French model.

We plot the conditional factor loadings in Fig. 4. Market factor loadings range between zero and 0.5. The SMB loadings generally remain above 0.5 until the mid-1970s and then decline to approximately 0.2 in the mid-1980s. During the 1990s the SMB loadings strongly trended upward, particularly during the late 1990s bull market. This is a period when value

Table 6:

Tests of constant conditional Fama and French (1993) alphas and factor loadings of book-to-market portfolios

Notes. The table reports W test statistics in Theorem 3 of tests of constancy of conditional alphas and factor loadings from a conditional Fama and French (1993) model applied to decile book-to-market portfolios and the 10 -1 book-to-market strategy. Constancy tests are done separately for each alpha or factor loading on each portfolio. We report the test statistic W and 95% critical values of the asymptotic distribution. We mark rejections at the 99% level with **. The full data sample is from July 1963 to December 2007, but the conditional estimates span July 1964 to December 2006 to avoid the bias at the endpoints.

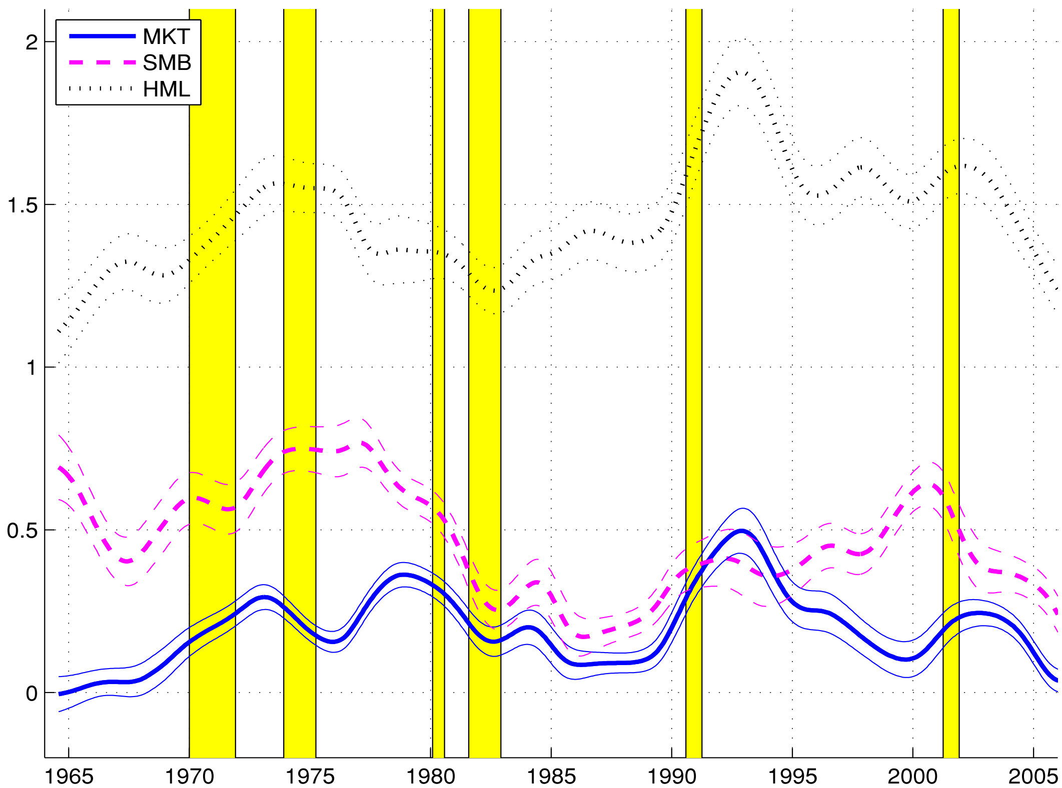
	W				Critical Values	
	Alpha	MKT	SMB	HML	95%	99%
1 Growth	59	441**	729**	3259**	285	295
2	64	369**	429**	1035**	348	360
3	44	184	198	450**	284	294
4	40	219	209	421**	236	246
5	31	212	182	675**	216	225
6	29	201	316**	773**	240	436
7	49	195	440**	1190**	353	365
8	60	221	483**	3406**	369	381
9	37	192	520**	3126**	270	281
10 Value	42	367**	612**	2194**	242	440
10-1 Book-to-Market Strategy	46	283**	1075**	4307**	257	267

stocks performed poorly and the high SMB loadings translate into larger negative conditional Fama and French alphas during this time. After 2000, the SMB loadings decrease to end the sample around 0.25.

Fig. 4 shows that the HML loadings are well above one for the whole sample and reach a high of 1.91 in 1993 and end the sample at 1.25. Value strategies perform well coming out of the early 1990s recession and the early 2000s recession, and HML loadings during these periods decrease for the book-to-market strategy. One could expect that the HML loadings should be constant because HML is constructed by Fama and French (1993) as a zero-cost mimicking portfolio to go long value stocks and short growth stocks, which is what the book-to-market strategy does. However, the breakpoints of the HML factor are different, at approximately thirds, compared to the first and last deciles of firms in the book-to-market strategy. The fact that the HML loadings vary so much over time indicates that growth and value stocks in the

Figure 4: Conditional Fama and French (1993) loadings of the book-to-market strategy

Notes. The figure shows monthly estimates of conditional Fama and French (1993) factor loadings of the book-to-market strategy, which goes long the tenth book-to-market decile portfolio and short the fist book-to-market decile portfolio. We plot the optimal estimates in bold lines along with 95% confidence bands in regular lines. National Bureau of Economic Research designated recession periods are shaded in horizontal bars.



10% extremes covary differently with average growth and value stocks in the middle of the distribution. Put another way, the 10% tail value stocks are not simply levered versions of value stocks with lower and more typical book-to-market ratios.

5 Portfolios sorted on past returns

We end by testing the conditional Fama and French (1993) model on decile portfolios sorted by past returns. These portfolios are well known to strongly reject the null of the standard Fama and French model with constant alphas and factor loadings. In Table 7, we report long-run estimates of alphas and Fama and French factor loadings for each portfolio and the 10–1 momentum strategy. The long-run alphas range from -4.68% with a standard error of 0.014 for the first loser decile to 2.97% with a standard error of 0.010 to the tenth loser decile. The momentum strategy has a long-run alpha of 8.11% with a standard error of 0.019. A joint test that the long-run alphas are equal to zero rejects with a p-value of zero. Thus, a conditional version of the Fama and French model cannot price the momentum portfolios.

Table 7 shows that no pattern emerges in the long-run market factor loading across the momentum deciles and the momentum strategy is close to market neutral in the long run with a long-run beta of 0.074. The long-run SMB loadings are small, except for the loser and winner deciles at 0.385 and 0.359, respectively. These effectively cancel in the momentum strategy, which is effectively SMB neutral. Finally, the long-run HML loading for the winner portfolio is noticeably negative at -0.175. The momentum strategy long-run HML loading is -0.117, and the negative sign means that controlling for a conditional HML loading exacerbates the momentum effect, as firms with negative HML exposures have low returns on average.

We can judge the impact of allowing for conditional Fama and French loadings in Fig. 5 which graphs the long-run alphas of the momentum portfolios 1–10 and the long-run alpha of the momentum strategy (portfolio 11). We overlay the OLS alpha estimates which assume constant factor loadings. The momentum strategy has a Fama-French OLS alpha of 16.7% with a standard error of 0.026. Table 7 reports that the long-run alpha controlling for time-varying factor loadings is 8.11%. Thus, the conditional factor loadings have lowered the momentum strategy alpha by almost 9% but this still leaves a large amount of the momentum effect unexplained. Fig. 5 shows that the reduction of the absolute values of OLS alphas compared with the long-run conditional alphas is particularly large for both the extreme loser and winner deciles.

Fig. 6 shows that all the Fama and French conditional factor loadings vary significantly over time and their variation is larger than the case of the book-to-market portfolios. Formal constancy tests (not reported) overwhelmingly reject the null of constant Fama and French factor loadings. Whereas the standard deviation of the conditional betas is around 0.2 for the book-to-market strategy (see Table 2), the standard deviations of the conditional Fama and French betas are 0.386, 0.584, and 0.658 for MKT, SMB, and HML, respectively. Fig. 6 also shows a marked common covariation of these factor loadings, with a correlation of 0.61 between conditional MKT and SMB loadings and a correlation of 0.43 between conditional SMB and HML loadings. During the early 1970s all factor loadings generally increased and all factor loadings also generally decrease during the late 1970s and through the 1980s. Beginning in 1990, all factor loadings experience a sharp run up and also generally trend downwards over the

Figure 5: Long-run alphas versus ordinary least-squares alphas in the Fama and French (1993) model for the momentum portfolios

Notes. We plot long-run alphas from a conditional Fama and French (1993) model and OLS Fama and French alphas for the momentum portfolios. We plot the long-run alphas using squares with 95% confidence intervals displayed in the error bars. The point estimates of the OLS alphas are plotted as circles with 95% confidence intervals in dashed lines. Portfolios 1–10 on the x-axis represent the loser to winner decile portfolios. Portfolio 11 is the momentum strategy, which goes long portfolio 10 and short portfolio 1. The long-run conditional and OLS alphas are annualized by multiplying by 252.

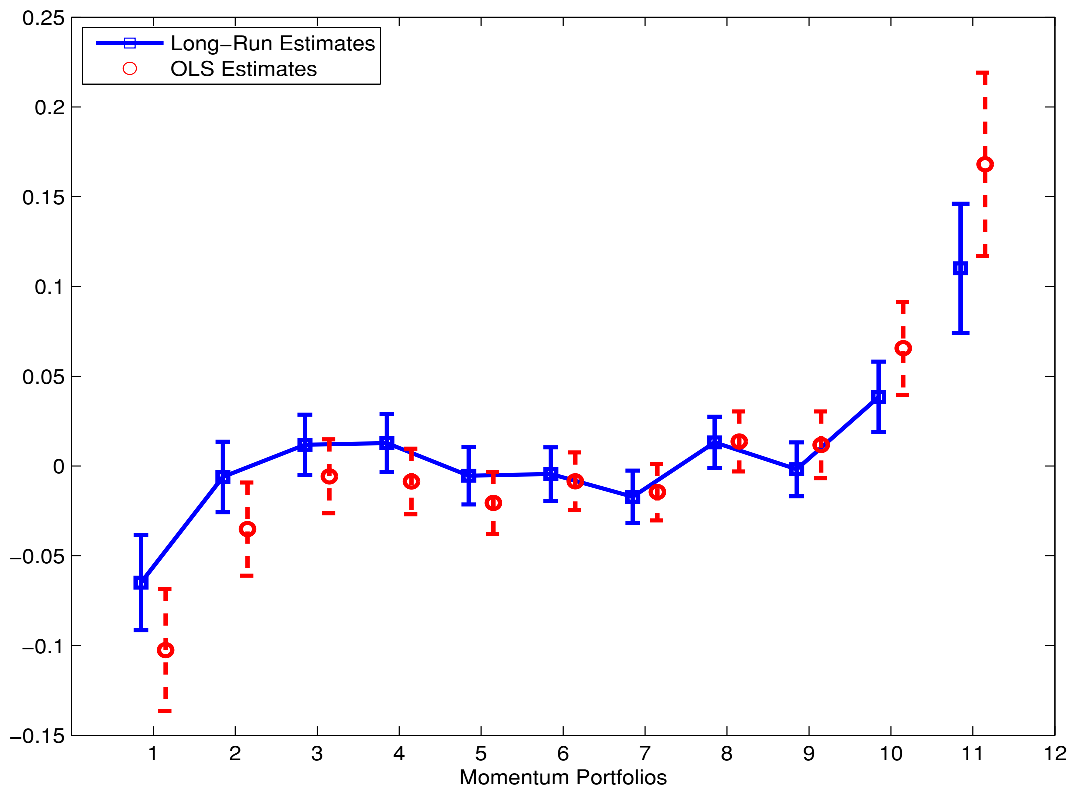


Table 7:

Long-Run Fama and French (1993) alphas and factor loadings of momentum portfolios

Notes. The table reports estimates of long-run alphas and factor loadings from a conditional Fama and French (1993) model applied to decile momentum portfolios and the 10-1 momentum strategy. The long-run estimates, with standard errors in parentheses, are computed following Theorem 2 and average daily estimates of conditional alphas and betas. The long-run alphas are annualized by multiplying by 252. The joint test for long-run alphas equal to zero is given by the Wald test statistic W_0 in Eq. (23). The full data sample is from July 1963 to December 2007, but the long-run estimates span July 1964 to December 2006 to avoid the bias at the endpoints.

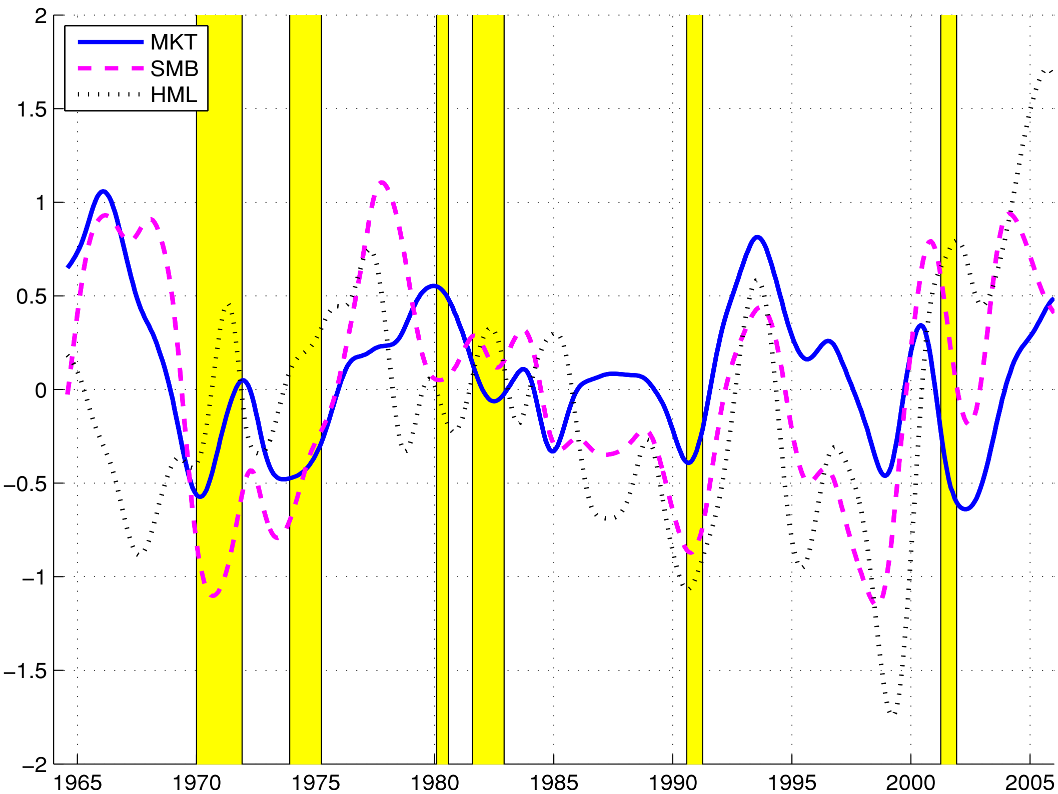
Portfolio	Alpha	MKT	SMB	HML
1 Losers	-0.0468	1.1762	0.3848	-0.0590
	(0.0138)	(0.0091)	(0.0132)	(0.0164)
2	0.0074	1.0436	0.0911	0.0280
	(0.0103)	(0.0072)	(0.0102)	(0.0128)
3	0.0226	0.9723	-0.0263	0.0491
	(0.0087)	(0.0061)	(0.0088)	(0.0111)
4	0.0162	0.9604	-0.0531	0.0723
	(0.0083)	(0.0058)	(0.0084)	(0.0104)
5	-0.0056	0.9343	-0.0566	0.0559
	(0.0082)	(0.0056)	(0.0082)	(0.0101)
6	-0.0043	0.9586	-0.0363	0.1111
	(0.0076)	(0.0053)	(0.0079)	(0.0097)
7	-0.0176	0.9837	-0.0270	0.0979
	(0.0074)	(0.0052)	(0.0077)	(0.0094)
8	0.0082	1.0258	-0.0238	0.1027
	(0.0073)	(0.0052)	(0.0077)	(0.0095)
9	-0.0078	1.0894	0.0815	0.0337
	(0.0078)	(0.0055)	(0.0081)	(0.0101)
10 Winners	0.0297	1.2523	0.3592	-0.1753
	(0.0103)	(0.0074)	(0.0183)	(0.0135)
10-1 Momentum strategy	0.0811	0.0739	-0.0286	-0.1165
	(0.0189)	(0.0127)	(0.0183)	(0.0230)

Joint test for $\alpha_{LR,i} = 0$, $i = 1, ..., 10$.
Wald statistic $W_0 = 91.0$, and p-value = 0.0000.

mid- to late 1990s. At the end of the sample the conditional HML loading is still particularly high at over 1.5. Despite this very pronounced time variation, conditional Fama and French factor loadings still cannot completely price the momentum portfolios.

Figure 6: Conditional Fama and French (1993) loadings of the momentum strategy

Notes. The figure shows monthly estimates of conditional Fama and French (1993) factor loadings of the momentum strategy, which goes long the tenth past return decile portfolio and short the first past return decile portfolio. National Bureau of Economic Research designated recession periods are shaded in horizontal bars.



6 Conclusion

We develop a new nonparametric methodology for estimating conditional factor models. We derive asymptotic distributions for conditional alphas and factor loadings at any given time and also for long-run alphas and factor loadings averaged over time. We also develop a test for the null hypothesis that the conditional alphas and factor loadings are constant over time. The tests can be run for a single asset and also jointly across a system of assets. In the special case of no time variation in the conditional alphas and factor loadings and homoskedasticity, our tests reduce to the well-known Gibbons, Ross, and Shanken (1989) test.

We apply our tests to decile portfolios sorted by book-to-market ratios and past returns. We find significant variation in factor loadings, but overwhelming evidence that a conditional CAPM and a conditional version of the Fama and French (1993) model cannot account for the value premium or the momentum effect. Joint tests for whether long-run alphas are equal to zero in the presence of time-varying factor loadings decisively reject for both the conditional CAPM and Fama and French models. We also find that conditional market betas for a book-to-market strategy exhibit little covariation with market risk premia. Consistent with the book-to-market and momentum portfolios rejecting the conditional models, accounting for time-varying factor loadings only slightly reduces the OLS alphas from the unconditional CAPM and Fama and French regressions which assume constant betas.

Our tests are easy to implement, powerful, and can be estimated asset-by-asset, just as in the traditional classical Gibbons, Ross, and Shanken (1989) test. There are many further empirical applications of the tests to other conditional factor models and other sets of portfolios, especially situations where betas are dynamic, such as many derivative trading, hedge fund returns, and time-varying leverage strategies. Theoretically, the tests can also be extended to incorporate adaptive estimators to take account the bias at the endpoints of the sample. Such estimators can also be adapted to yield estimates of future conditional alphas or factor loadings that do not use forward-looking information.

Appendix A Technical assumptions

Our theoretical results are derived under the assumption that the true data-generating process is given by Eqs. (13) and (15). Throughout the Appendix all assumptions and arguments are stated conditionally on the realizations of $\alpha(t)$, $\beta(t)$, $\mu_F(t)$, $\Lambda_{FF}(t)$, and $\Sigma(t)$. We assume that we have observed data at $-cT \leq t \leq (1+c)T$ for some fixed $c > 0$ chosen so that the end-point bias is negligible to avoid any boundary issues and to keep the notation simple.

Let $C^2[0,1]$ denote the space of twice continuously differentiable functions on the interval, $[0,T]$. We impose the following assumptions:

A.1 There exists $B, L < \infty$ and $\nu > 1$ such that $|K(u)| \leq B\|u\|^{-\nu}$ for $\|u\| \geq L$; either $K(u) = 0$ for $\|u\| > L$ and $|K(u) - K(u')| \leq B\|u - u'\|$; or $K(u)$ is differentiable with $|\partial K(u)/\partial u| \leq B$ and $|\partial K(u)/\partial u| \leq B\|u\|^{-\nu}$ for $\|u\| \geq L$; $\int_{\mathbb{R}} K(z)\,dz = 1$, $\int_{\mathbb{R}} zK(z)\,dz = 0$ and $\kappa_2 := \int_{\mathbb{R}} z^2 K(z)\,dz < \infty$.

A.2 The realizations $\alpha(t)$, $\beta(t)$, $\mu_F(t)$, $\Lambda_{FF}(t)$, and $\Sigma(t)$ all lie in $C^2[0,T]$. Furthermore, $\Lambda_{FF}(t)$ is positive definite for any $t \in [0,T]$.

A.3 Conditional on the realizations in A.2, the errors z_i and u_i are i.i.d. with $E[z_i] = 0$, $E[u_i] = 0$, $E[z_i z_i'] = I_M$ and $E[u_i u_i'] = I_J$.

A.4 The processes $\alpha(t)$, $\beta(t)$, $\Lambda_{FF}(t)$, and $\Sigma(t)$ are stationary, ergodic, and bounded.

A.5 The bandwidth satisfies $\Delta^{-1}(Th)^4 \to 0$, $\Delta/(Th)^2 \to 0$, and $\Delta^{1-\epsilon}/(Th)^{7/4} \to 0$ for some $\epsilon > 0$.

Most standard kernels, including the Gaussian and the uniform one, satisfy A.1. The smoothness conditions in A.2 rule out jumps in the coefficients; Theorem 1 and 4 remain valid at all points where no jumps have occurred, and we conjecture that Theorems 2 and 3 remain valid with a finite jump activity because this will have a minor impact as we smooth over the whole time interval. One could exchange $C^2[0,T]$ for the space of Lipschitz functions of order $\lambda > 0$, $\|f(t) - f(t')\| \leq C|t-t|^{\lambda}$ with all the theoretical results remaining correct after suitably adjustment of the requirements imposed on the bandwidth h. The requirement that $\Lambda_{FF}(t) > 0$ is a local version of the rank condition known from OLS. The i.i.d. assumption in A.3 can be replaced by mixing or martingale conditions. We however maintain the i.i.d. assumption for simplicity because the proofs otherwise would become more involved. Since all the moment conditions have to hold conditionally on the realizations, A.3 rules out leverage effect. We conjecture that leverage effects can be accommodated by employing arguments similar to the ones used in the realized volatility literature, see, e.g., Foster and Nelson (1996). Assumption A.4 is imposed when we analyze the long-run alpha and beta estimators to ensure that the long-run alphas and betas are well-defined as the limits of sample averages. The conditions on the bandwidth in A.5 are needed for the estimators of the long-run alphas and betas and when testing the null of constant alphas or betas. The assumption entails undersmoothing (relative to the point optimal choice of h in terms of MSE) in the computation of the long-run estimates and the test statistics.

Appendix B Proofs

Proof of Theorem 1. Define $\bar{h} = hT$ such that $\hat{\beta}(t)$ can be rewritten as

$$\hat{\beta}(t) = \left[\Delta^2 \sum_{i=1}^{n} K_{\bar{h}}(t_i - t) f_i f_i' - \Delta \bar{f}(t) \bar{f}(t)'\right]^{-1} \left[\Delta^2 \sum_{t=1}^{n} K_{\bar{h}}(t_i - t) f_i R_i' - \Delta \bar{f}(t) \bar{R}(t)'\right], \tag{32}$$

where $\bar{f}(t) = \Delta \sum_{i=1}^{n} K_{\bar{h}}(t_i - t) f_i$ and $\bar{R}(t) = \Delta \sum_{i=1}^{n} K_{\bar{h}}(t_i - t) R_i$. Because

$$f_i = \Delta F_i / \Delta = \mu_F(t_i) + \frac{1}{\sqrt{\Delta}} \Lambda_{FF}^{1/2}(t_i) u_i, \tag{33}$$

we obtain that

$$\bar{f}(t) = \Delta \sum_{i=1}^{n} K_{\bar{h}}(t_i - t) \mu_F(t_i) + \sqrt{\Delta} \sum_{i=1}^{n} K_{\bar{h}}(t_i - t) \Lambda_{FF}^{1/2}(t_i) u_i =: \bar{f}_1(t) + \bar{f}_2(t). \tag{34}$$

Using standard results for Riemann sums and kernel estimators,

$$\bar{f}_1(t) = \Delta \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,\mu_F(t_i) = \int_0^T K_{\bar{h}}(s-t)\,\mu_F(s)\,ds + O(\Delta) = \mu_F(t) + O\left(\bar{h}^2\right) + O(\Delta). \quad (35)$$

The second term has mean zero while its variance satisfies

$$\operatorname{var}\left(\bar{f}_2(\tau)\right) = \Delta \sum_{i=1}^{n} K_{\bar{h}}^2(t_i - t)\,\Lambda_{FF}(t_i) = \frac{1}{\bar{h}}\left\{\kappa_2 \Lambda_{FF}(t) + O\left(\bar{h}^2\right) + O(\Delta)\right\}, \quad (36)$$

where we employ the same arguments as in the analysis of the first term. A similar analysis can be carried out for $\bar{R}(t)$ and we conclude that

$$\bar{f}(t) = \mu_F(t) + O_P(\bar{h}^2) + O_P(1/\sqrt{\bar{h}}), \quad \bar{R}(t) = \alpha(t) + \beta(t)'\,\mu_F(t) + O_P(\bar{h}^2) + O_P(1/\sqrt{\bar{h}}). \quad (37)$$

Similarly,

$$\begin{aligned}\Delta^2 \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,f_i f_i' &= \Delta \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,\Lambda_{FF}^{1/2}(t_i)\,u_i u_i' \Lambda_{FF}^{1/2}(t_i) + \Delta^2 \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,\mu_F(t_i)\,\mu_F(t_i)' \\ &\quad + \Delta^2 \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,\mu_F(t_i)\,u_i' \Lambda_{FF}^{1/2}(t_i) \\ &= \Lambda_{FF}(t) + \Delta \mu_F(t)\,\mu_F(t)' + O_P\left(\bar{h}^2\right) + O_P(\sqrt{\Delta/\bar{h}}), \quad (38)\end{aligned}$$

and

$$\begin{aligned}\Delta^2 \sum_{t=1}^{n} K_{\bar{h}}(t_i - t)\,f_i R_i' &= \Delta^2 \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,f_i f_i' \beta(t_i) + \Delta^2 \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,f_i \alpha(t_i)' \\ &\quad + \Delta^{3/2} \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,f_i z_i' \Sigma^{1/2}(t_i) \\ &= \Lambda_{FF}(t)\,\beta(t) + \Delta \mu_F(t)\,\alpha(t)' + O_P\left(\bar{h}^2\right) + \sqrt{\Delta/\bar{h}}\,U_n(t), \quad (39)\end{aligned}$$

where, by a Central Limit Theorem (CLT) for martingales (see, e.g., Brown, 1971),

$$U_n(t) := \Delta \sqrt{\bar{h}} \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,f_i z_i' \Sigma^{1/2}(t_i) \xrightarrow{d} N\left(0, \kappa_2 \Lambda_{FF}(t) \otimes \Sigma(t)\right). \quad (40)$$

Collecting the results of Eqs. (37)-(39), we obtain

$$\begin{aligned}&\sqrt{\Delta^{-1}\bar{h}}\left\{\Delta^2 \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,f_i R_i' - \Delta \bar{f}(t)\,\bar{R}(t)' - \Lambda_{FF}(t)\,\beta(t)\right\} \\ &= U_n(t) + o_P(1) \xrightarrow{d} N\left(0, \kappa_2 \Lambda_{FF}(t) \otimes \Sigma(t)\right), \quad (41)\end{aligned}$$

and

$$\Delta^2 \sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,f_i f_i' - \Delta \bar{f}(t)\,\bar{f}(t)' = \Lambda_{FF}(t) + o_P(1). \quad (42)$$

This yields the claimed result. ■

Proof of Theorem 2. The proof proceeds along the same lines as in Kristensen (2010, Proof of Theorem 4) and so we sketch only the arguments. First consider $\hat{\beta}_{\mathrm{LR}}$. As demonstrated in the proof of Theorem 1, with $\bar{h} = hT$,

$$\begin{aligned}\hat{\beta}(t) &\simeq \left[\sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,\Delta F_i \Delta F_i'\right]^{-1} \left[\sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,\Delta F_i \Delta F_i' \beta(t_i)\right] \\ &\quad + \sqrt{\Delta}\left[\sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,\Delta F_i \Delta F_i'\right]^{-1} \left[\sum_{i=1}^{n} K_{\bar{h}}(t_i - t)\,\Delta F_i z_i' \Sigma^{1/2}(t_i)\right], \quad (43)\end{aligned}$$

where, uniformly over $t \in [0, T]$,

$$\sum_{i=1}^{n} K_{\bar{h}}\left(t_i - t\right) \Delta F_i \Delta F_i' = \Lambda_{FF}\left(t\right) + O_P\left(\bar{h}^2\right) + O_P\left(\sqrt{\Delta/\bar{h}}\right). \tag{44}$$

Thus,

$$\begin{aligned}
\hat{\beta}_{\mathrm{LR}} &\simeq \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{n}\Lambda_{FF}^{-1}\left(t_j\right)K_{\bar{h}}\left(t_i - t_j\right)\right\}\Delta F_i \Delta F_i' \beta\left(t_i\right) \\
&\quad + \frac{\sqrt{\Delta}}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{n}\Lambda_{FF}^{-1}\left(t_j\right)K_{\bar{h}}\left(t_i - t_j\right)\right\}\Delta F_i z_i' \Sigma^{1/2}\left(t_i\right) \\
&\simeq \frac{1}{n\Delta}\sum_{i=1}^{n}\Lambda_{FF}^{-1}\left(t_i\right)\Delta F_i \Delta F_i' \beta\left(t_i\right) + \frac{1}{n\sqrt{\Delta}}\sum_{i=1}^{n}\Lambda_{FF}^{-1}\left(t_i\right)\Delta F_i z_i' \Sigma^{1/2}\left(t_i\right) \\
&\equiv B_1 + B_2.
\end{aligned} \tag{45}$$

Here and in the following, $\simeq$ is used to denote that the left and right hand side are identical up to some higher-order term which is asymptotically negiglible under our assumptions. Because $\Delta F_i \Delta F_i'/\Delta \simeq \Lambda_{FF}\left(t_i\right)$, $B_1 \simeq n^{-1}\sum_{i=1}^{n}\beta\left(t_i\right) \simeq T^{-1}\int_0^T \beta\left(s\right)ds$. The second term is a martingale with quadratic variation

$$\begin{aligned}
\left\langle B_2 \right\rangle &= \frac{1}{n^2\Delta}\sum_{i=1}^{n}\Lambda_{FF}^{-1}\left(t_i\right)\Delta F_i \Delta F_i' \Lambda_{FF}^{-1}\left(t_i\right) \otimes \Sigma\left(t_i\right) \simeq \frac{1}{n^2}\sum_{i=1}^{n}\Lambda_{FF}^{-1}\left(t_i\right) \otimes \Sigma\left(t_i\right) \\
&\simeq \frac{1}{nT}\int_0^T \Lambda_{FF}^{-1}\left(s\right) \otimes \Sigma\left(s\right)ds.
\end{aligned} \tag{46}$$

The result for the LR betas now follows by the CLT for martingales together with the Law of Large Numbers (LLN) for stationary and ergodic processes.

Next, consider $\hat{\alpha}_{\mathrm{LR}}$: With $\bar{f}\left(t\right) = \sum_{i=1}^{n} K_{\bar{h}}\left(t_i - t\right) f_i$ and $\bar{R}\left(t\right) = \sum_{i=1}^{n} K_{\bar{h}}\left(t_i - t\right) R_i$,

$$\hat{\alpha}_{\mathrm{LR}} = \frac{1}{n}\sum_{j=1}^{n}\hat{\alpha}\left(t_j\right) = \frac{1}{n}\sum_{j=1}^{n}\frac{\bar{R}\left(t_j\right) - \hat{\beta}\left(t_j\right)\bar{f}\left(t_j\right)}{\sum_{j=1}^{n} K_{\bar{h}}\left(t_i - t_j\right)}, \tag{47}$$

where uniformly over $t \in [0, T]$, $\Delta\sum_{i=1}^{n} K_{\bar{h}}\left(t_i - t\right) \simeq 1$ and $\hat{\beta}\left(t\right) \simeq \beta\left(t\right)$. Thus,

$$\begin{aligned}
\hat{\alpha}_{\mathrm{LR}} &\simeq \frac{\Delta}{n}\sum_{j=1}^{n}\left\{\bar{R}\left(t_j\right) - \beta\left(t_j\right)\bar{f}\left(t_j\right)\right\} \\
&= \frac{1}{n\Delta}\sum_{i=1}^{n}\left\{\Delta\sum_{j=1}^{n} K_{\bar{h}}\left(t_i - t_j\right)\right\}\Delta s_i - \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{n} K_{\bar{h}}\left(t_i - t_j\right)\beta\left(t_j\right)\right\}\Delta F_i \\
&\simeq \frac{1}{n\Delta}\sum_{i=1}^{n}\left\{\Delta s_i - \beta\left(t_i\right)\Delta F_i\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\alpha\left(t_i\right) + \frac{1}{n\sqrt{\Delta}}\sum_{i=1}^{n}\Sigma^{1/2}\left(t_i\right)z_i \\
&\equiv A_1 + A_2.
\end{aligned} \tag{48}$$

By the same arguments as for $\hat{\beta}_{\mathrm{LR}}$, $A_1 \simeq T^{-1}\int_0^T \alpha\left(s\right)ds$, while A_2 is a martingale with quadratic variation

$$\left\langle A_2 \right\rangle = \frac{1}{n^2\Delta}\sum_{i=1}^{n}\Sigma\left(t_i\right) \simeq \frac{1}{T^2}\int_0^T \Sigma\left(s\right)ds. \tag{49}$$

The result for the LR alphas now follows by the CLT for martingales and the LLN for stationary and ergodic processes. ■

Proof of Theorem 3. The proof follows along the same lines as the proof of Kristensen (2011, Theorem 3.3) and so we only sketch it. We suppress the index k in the following, and first consider the test statistic for constant alphas, $W(\alpha)$. Let $\alpha(t) = \alpha$ denote the true, constant alpha realization, and for simplicity suppose that $F_t := 0$ such that the data-generating process under $H(\alpha)$ is $\Delta s_i = \alpha\Delta + \sqrt{\Delta}\sigma(t_i) z_i$. The extension to $F_t \neq 0$ is straightforward but tedious. Using that $\Delta \sum_{j=1}^{n} K_{\bar{h}}(t_j - t) \simeq 1$ uniformly over $t \in [0, T]$, the nonparametric estimators of α can therefore be written as

$$\hat{\alpha}(t) = \frac{\sum_{j=1}^{n} K_{\bar{h}}(t_j - t)\Delta s_j}{\Delta \sum_{j=1}^{n} K_{\bar{h}}(t_j - t)} = \alpha + \frac{\sum_{j=1}^{n} K_{\bar{h}}(t_j - t)\sigma(t_j) z_j}{\sqrt{\Delta}\sum_{j=1}^{n} K_{\bar{h}}(t_j - t)} \simeq \alpha + \sqrt{\Delta}\sum_{j=1}^{n} K_{\bar{h}}(t_j - t)\sigma(t_j) z_j, \quad (50)$$

where $\bar{h} = hT$. Because $\hat{\alpha}_{\mathrm{LR}}$ is $\sqrt{T}$-consistent, we are allowed to set $\hat{\alpha}_{\mathrm{LR}} = \alpha$. Also, by the same arguments as in Kristensen (2011, Proof of Theorem 3.3), we can treat $\sigma(t)$ as known. Thus,

$$\begin{aligned}
W(\alpha) &\simeq \frac{1}{n}\sum_{i=1}^{n} \sigma^{-2}(t_i)\left[\hat{\alpha}(t) - \alpha\right]^2 \\
&\simeq \frac{1}{n}\sum_{i=1}^{n} \sigma^{-2}(t_i)\left[\sqrt{\Delta}\sum_{j=1}^{n} K_{\bar{h}}(t_j - t)\sigma(t_j) z_j\right]^2 \\
&\simeq \frac{\Delta}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \sigma^{-2}(t_i) K_{\bar{h}}(t_j - t_i) K_{\bar{h}}(t_k - t_i)\sigma(t_j)\sigma(t_k) z_j z_k \qquad (51)\\
&= \frac{1}{n}\sum_{j=1}^{n}\left\{\Delta\sum_{i=1}^{n} K_{\bar{h}}^2(t_j - t_i)\sigma^{-2}(t_i)\right\}\sigma^2(t_j) z_j^2 \\
&\quad + \frac{1}{n}\sum_{j\neq k}\left\{\Delta\sum_{i=1}^{n} K_{\bar{h}}(t_j - t_i) K_{\bar{h}}(t_k - t_i)\sigma^{-2}(t_i)\right\}\sigma(t_j)\sigma(t_k) z_j z_k,
\end{aligned}$$

where, uniformly over $t_j, t_k \in [0, T]$,

$$\begin{aligned}
\Delta\sum_{i=1}^{n} K_{\bar{h}}(t_j - t_i) K_{\bar{h}}(t_k - t_i)\sigma^{-2}(t_i) &\simeq \int_0^T K_{\bar{h}}(t_j - s) K_{\bar{h}}(t_k - s)\sigma^{-2}(s)\, ds \\
&\simeq (K * K)_{\bar{h}}(t_j - t_k)\sigma^{-2}(t_j). \qquad (52)
\end{aligned}$$

Thus, with $\kappa_2 = (K * K)(0)$,

$$W(\alpha) \simeq \frac{\kappa_2}{n\bar{h}}\sum_{j=1}^{n} z_j^2 + \frac{1}{n}\sum_{j\neq k}\phi_{j,k} \simeq \frac{\kappa_2}{\bar{h}} + \frac{1}{n}\sum_{j\neq k}\phi_{j,k}, \qquad (53)$$

where $\phi_{j,k} := (K * K)_{\bar{h}}(t_j - t_k)\sigma^{-1}(t_j)\sigma(t_k) z_j z_k$. We recognize $\sum_{j\neq k}\phi_{j,k}$ as a degenerate U-statistic. Because

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{j\neq k}\phi_{j,k}\Big) &= 2\sum_{j\neq k}(K * K)_{\bar{h}}^2(t_j - t_k)\sigma^{-2}(t_j)\sigma^2(t_k) E\left[z_j^2\right] E\left[z_k^2\right] \\
&= 2\sum_{j\neq k}(K * K)_{\bar{h}}^2(t_j - t_k)\sigma^{-2}(t_j)\sigma^2(t_k) \qquad (54)\\
&\simeq \frac{2}{\Delta^2}\int_0^T\int_0^T (K * K)_{\bar{h}}^2(s - t)\sigma^{-2}(s)\sigma^2(t)\, ds dt \\
&\simeq \frac{2}{\Delta^2\bar{h}}\int (K * K)^2(z)\, dz,
\end{aligned}$$

it follows by standard arguments for degenerate U-statistics that $\sqrt{\bar{h}}\Delta\sum_{j\neq k}\phi_{j,k}\to^{d} N\left(0,2\int\left(K*K\right)^{2}(z)\,dz\right)$. In total, with $m(\alpha)=\kappa_{2}/\bar{h}$ and $v^{2}(\alpha)=2\int\left(K*K\right)^{2}(z)\,dz/\left(T^{2}\bar{h}\right)$, $\left[W(\alpha)-m(\alpha)\right]/v(\alpha)\xrightarrow{d}N(0,1)$. This shows the result.

Next, consider $W(\beta)$. We focus on the simple case in which $\alpha(t)=0$ such that, under $H(\beta)$, $\Delta s_{i}=\beta'\Delta F_{i}+\sqrt{\Delta}\sigma(t_{i})\,z_{i}$. The nonparametric estimators of β can therefore be written as

$$\hat{\beta}(t)\simeq\beta+\Lambda_{FF}^{-1}(t)\sqrt{\Delta}\sum_{j=1}^{n}K_{\bar{h}}(t_{j}-t)\,\Delta F_{j}\sigma(t_{j})\,z_{j},\tag{55}$$

and we obtain, using that $\Delta F_{j}\Delta F_{j}'\simeq\Delta\Lambda_{FF}^{-1}(t_{j})$,

$$\begin{aligned}
W(\beta) &\simeq \frac{1}{n}\sum_{i=1}^{n}\sigma^{-2}(t_{i})\left[\hat{\beta}(t_{i})-\beta\right]'\Lambda_{FF}(t_{i})\left[\hat{\beta}(t_{i})-\beta\right]\\
&\simeq \frac{1}{n}\sum_{j=1}^{n}\sigma^{2}(t_{j})\,z_{j}^{2}\Delta F_{j}'\left\{\Delta\sum_{i=1}^{n}\Lambda_{FF}^{-1}(t_{i})\,\sigma^{-2}(t_{i})\,K_{\bar{h}}^{2}(t_{j}-t_{i})\right\}\Delta F_{j}\\
&\quad+\frac{1}{n}\sum_{j\neq k}^{n}\sigma(t_{j})\,z_{j}\Delta F_{j}'\left\{\Delta\sum_{i=1}^{n}\Lambda_{FF}^{-1}(t_{i})\,\sigma^{-2}(t_{i})\,K_{\bar{h}}(t_{k}-t_{i})\,K_{\bar{h}}(t_{j}-t_{i})\right\}\Delta F_{k}\sigma(t_{k})\,z_{k} \qquad (56)\\
&\simeq \frac{\kappa_{2}}{n\bar{h}}\sum_{j=1}^{n}z_{j}^{2}\Delta F_{j}'\Lambda_{FF}^{-1}(t_{j})\,\Delta F_{j}\\
&\quad+\frac{1}{n}\sum_{j\neq k}^{n}\sigma^{-1}(t_{j})\,z_{j}\Delta F_{j}'\Lambda_{FF}^{-1}(t_{j})\,(K*K)_{\bar{h}}(t_{j}-t_{k})\,\Delta F_{k}\sigma(t_{k})\,z_{k}\\
&\simeq \frac{\kappa_{2}J\Delta}{\bar{h}}+\frac{1}{n}\sum_{j\neq k}^{n}\phi_{j,k},
\end{aligned}$$

where $\phi_{j,k}=(K*K)_{h}(t_{j}-t_{k})\,\Delta F_{j}'\Lambda_{FF}^{-1}(t_{j})\,\Delta F_{k}\sigma^{-1}(t_{j})\,\sigma(t_{k})\,z_{j}z_{k}$. $\sum_{j\neq k}^{n}\phi_{j,k}$ is a degenerate U-stastistic satisfying

$$\begin{aligned}
\operatorname{Var}\Big(\sum_{j\neq k}\phi_{j,k}\Big) &= 2\Delta^{2}\sum_{j\neq k}(K*K)_{\bar{h}}^{2}(t_{j}-t_{k})\,\Lambda_{FF}^{-1}(t_{j})\,\Lambda_{FF}(t_{k})\,\sigma^{-2}(t_{j})\,\sigma^{2}(t_{k}) \qquad (57)\\
&\simeq \frac{2J}{\bar{h}}\int(K*K)^{2}(z)\,dz,
\end{aligned}$$

and so $\sqrt{\bar{h}}\sum_{j\neq k}^{n}\phi_{j,k}\to^{d}N\left(0,2J\int\left(K*K\right)^{2}(z)\,dz\right)$. In conclusion, with $m(\beta)=\kappa_{2}J\Delta/\bar{h}$ and $v^{2}(\beta)=2J\int\left(K*K\right)^{2}(z)\,dz/\left(n^{2}\bar{h}\right)$, $\left[W(\beta)-m(\beta)\right]/v(\beta)\xrightarrow{d}N(0,1)$ as desired. ■

Appendix C Conditional alphas

In this section we develop an asymptotic theory for the conditional alpha estimators under the following additional assumption.

A.6 The realizations satisfy $\alpha(t)=a(t/T)$, $\beta(t)=b(t/T)$, $\mu_{F}(t)=m_{F}(t/T)$, $\Lambda_{FF}(t)=L_{FF}(t/T)$, and $\Sigma(t)=S(t/T)$ for some functions $a(\tau)$, $b(\tau)$, $m_{F}(\tau)$, $L_{FF}(\tau)$, and $S(\tau)$ that all lie in $C^{2}[0,1]$.

The normalization imposed in A.6 is similar to the one used in the structural change (or break points) literature. These models are originally developed by, among others, Andrews (1993) and Bai and Perron (1998). Bekaert, Harvey, and Lumsdaine (2002), Paye and Timmermann (2006), and Lettau and Van Nieuwerburgh (2007), among others, apply these models in finance. In testing for structural breaks with one fixed model, the number of observations before the break remains fixed as the sample size increases, $n\to\infty$, and so identification of the break point is not possible. Instead, the literature normalizes time t by sample size n such that the number of observations both before and after the break increases as $n \to \infty$. This allows the break point and the parameters describing the coefficients before the break to be identified as n increases. The same time normalization is also used in nonparametric methods for detecting structural change as initally proposed in Robinson (1989) and further extended by Cai (2007) and Kristensen (2011). The structural break specifications, in which the coefficients take on different (constant) values before and after a break point, are special cases of our more general model, which allows for a (finite) number of jumps (see Appendix A).

The same idea of employing a sequence of models to construct valid asymptotic inference is also used in the literature on local-to-unity tests introduced by Chan and Wei (1987) and Phillips (1987). The local-to-unity models employ a sequence of models indexed by sample size to ensure that the same asymptotics apply for both stationary and non-stationary cases. Analogously, our sequence of models scales the underlying parameters by time span to obtain well-defined asymptotic distributions.

With the time normalization in A.6, the bias of the conditional alpha estimator changes from the one stated in Eq. (19) and now becomes $\mathrm{E}[\hat{\alpha}(t)] \simeq \alpha(t) + h^2\alpha^{(2)}(t)$ while the variance in Eq. (19) remains unchanged. Thus, we can now choose the bandwidth such that the bias vanishes faster than the variance component, and so obtain the following formal result.

Theorem 4 *Assume that assumption A.1–A.3 and A.6 hold and the bandwidth is chosen such that $Th \to \infty$ and $Th^5 \to 0$. Then, for any $t \in [0, T]$,*

$$\sqrt{Th}\left\{\hat{\alpha}(t) - \alpha(t)\right\} \sim N(0, \kappa_2 \Sigma(t)) \quad \textit{in large samples.} \tag{58}$$

Moreover, the conditional estimators are asymptotically independent across any set of distinct time points.

Proof. With $\tau := t/T$ and $\tau_i := t_i/T$, we can write the estimator as $\hat{\alpha}(t) = \hat{a}(\tau T)$ where

$$\hat{a}(\tau) = \frac{\sum_{i=1}^n K_h(\tau_i - \tau)\left\{R_i - \hat{b}(\tau)' f_i\right\}}{\sum_{i=1}^n K_h(\tau_i - \tau)}, \tag{59}$$

and $\hat{b}(\tau) = \hat{\beta}(\tau T)$. Given the normalizations in A.6, we obtain by the same arguments used in the proof of Theorem 1 that $\hat{b}(\tau) = b(\tau) + O_P\left(h^2\right) + O_P\left(1/\sqrt{nh}\right)$ such that $R_i - \hat{b}(\tau_i)' f_i \simeq a(\tau_i) + S^{1/2}(\tau_i) z_i/\sqrt{\Delta}$. This in turn implies that

$$\sqrt{Th}(\hat{a}(\tau) - a(\tau)) = O_P\left(\sqrt{Th^5}\right) + \frac{\sqrt{h}}{\sqrt{n}} \sum_{t=1}^n K_h(\tau_i - \tau) S^{1/2}(\tau_i) z_i, \tag{60}$$

where, as $Th \to \infty$,

$$\frac{\sqrt{h}}{\sqrt{n}} \sum_{t=1}^n K_{\bar{h}}(\tau_i - \tau) S^{1/2}(\tau_i) z_i \xrightarrow{d} N(0, \kappa_2 S(\tau)). \tag{61}$$

■

The crucial assumption in Theorem 4 is the time normalization in A.6, without which we cannot estimate conditional alphas as discussed in Subsection 2.5. Given the large literature which identifies the constant terms in rolling-window regression estimates as conditional alphas, Theorem 4 can be used to identify these as estimates of conditional alphas only with the time-normalization assumption in Eq. (21).

Using Theorem 4, we can test the hypothesis that $\alpha(t) = 0$ jointly across M stocks for a given value of $t \in [0, T]$ by the Wald statistic

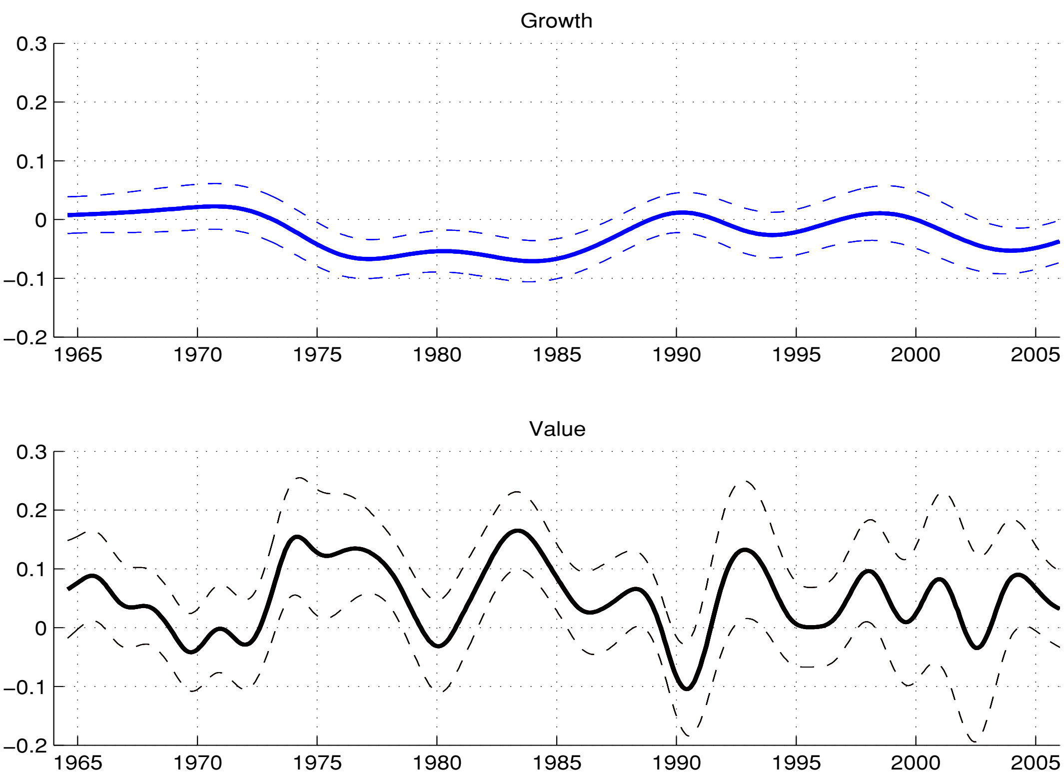
$$W(t) \equiv Th\hat{\alpha}(t)'\left[\kappa_2 \hat{\Sigma}(t)\right]^{-1} \hat{\alpha}(t) \sim \chi^2_M, \tag{62}$$

in large samples, where $\hat{\Sigma}(t)$ is given in Eq. (17). Given independence across distinct time points, we can also construct tests across any given finite set of time points. However, this test is not able to detect all departures from the null, because we test only for departures at a finite number of time points. To test the conditional alphas being equal to zero uniformly over time, we advocate using the test for constancy of the conditional alphas in Section 2.7.

In Fig. 7 we plot conditional alphas from the growth and value portfolios. Again we emphasize that without the time normalization A.6, the estimates of the conditional alphas cannot consistently estimate the true conditional alphas. In Fig. 7 the conditional alphas of both growth and value stocks have fairly wide standard errors, which often encompass zero. These results are similar to Ang and Chen (2007) who cannot reject that conditional alphas of value stocks is equal to zero over the post-1926 sample. Conditional alphas of growth stocks are significantly negative during 1975-1985 and reach a low of -7.09% in 1984. Growth stock conditional alphas are again significantly negative from 2003 to the end of our sample. The conditional alphas of value stocks are much more variable than the conditional alphas of growth stocks, but their standard errors are wider and so we cannot reject that the conditional alphas of value stocks are equal to zero except for the mid-1970s, the early 1980s, and the early 1990s. During the mid-1970s and the early 1980s, estimates of the conditional alphas of value stocks reach approximately 15%. During 1991, value stock conditional alphas decline to below -10%. Interestingly, the poor performance of value stocks during the late 1990s does not correspond to negative conditional alphas for value stocks during this time.

Figure 7: Conditional Alphas of Growth and Value Portfolios

Notes. The figure shows monthly estimates of conditional conditional alphas from a conditional CAPM of the first and tenth decile book-to-market portfolios (growth and value, respectively). We plot 95% confidence bands in dashed lines. The conditional alphas are annualized by multiplying by 252.



The contrast between the wide standard errors for the conditional alphas in Fig. 7 compared to the tight confidence bands for the long-run alphas in Table 2 is due to the fact that the conditional alpha estimators converge at the nonparametric rate $\sqrt{Th}$, which is less than the classical rate $\sqrt{T}$, and thus the conditional standard error bands are wider. This is exactly what Fig. 7 shows and what Ang and Chen (2007) pick up in an alternative parametric procedure.

Appendix D Gibbons, Ross, and Shanken (1989) as a special case

First, we derive the asymptotic distribution of the Gibbons, Ross, and Shanken (1989) estimators (GRS estimators) within the setting of the continuous-time diffusion model. The GRS estimators, which we denote

$\tilde{\gamma}_{\rm LR} = (\tilde{\alpha}_{\rm LR}, \tilde{\beta}_{\rm LR})$, are standard least squares estimator of the form $\tilde{\gamma}_{\rm LR} = \left[\sum_{i=1}^{n} X_i X_i'\right]^{-1} \left[\sum_{i=1}^{n} X_i R_i'\right]$. Under assumptions A.1-A.2 and A.5,

$$\begin{aligned} \tilde{\beta}_{\rm LR} &\simeq \left[\frac{1}{T}\sum_{i=1}^{n} \Delta F_i \Delta F_i'\right]^{-1} \left[\frac{1}{T}\sum_{i=1}^{n} \Delta F_i \Delta F_i' \beta(t_i)\right] + \left[\frac{1}{T}\sum_{i=1}^{n} \Delta F_i \Delta F_i'\right]^{-1} \left[\frac{\Delta^{1/2}}{T}\sum_{i=1}^{n} \Delta F_i z_i' \Sigma^{1/2}(t_i)\right] \\ &\simeq \bar{\beta}_{\rm LR} + {\rm E}\left[\Lambda_{FF}(t)\right]^{-1} \frac{\Delta^{1/2}}{T}\sum_{i=1}^{n} \Delta F_i z_i' \Sigma^{1/2}(t_i) \end{aligned} \tag{63}$$

where $\sqrt{n}\frac{\Delta^{1/2}}{T}\sum_{i=1}^{n} \Delta F_i z_i' \Sigma^{1/2}(t_i) \xrightarrow{d} N\left(0, {\rm E}\left[\Lambda_{FF}(t) \otimes \Sigma(t)\right]\right)$ and

$$\bar{\beta}_{\rm LR} = {\rm E}\left[\Lambda_{FF}(t)\right]^{-1} {\rm E}\left[\Lambda_{FF}(t)\beta(t)\right] \tag{64}$$

Thus, $\sqrt{n}(\tilde{\beta}_{\rm LR} - \bar{\beta}_{\rm LR}) \to^d N\left(0, {\rm E}\left[\Lambda_{FF}(t)\right]^{-1} {\rm E}\left[\Lambda_{FF}^2(t) \otimes \Sigma(t)\right] {\rm E}\left[\Lambda_{FF}(t)\right]^{-1}\right)$. Next,

$$\begin{aligned} \tilde{\alpha}_{\rm LR} &\simeq \frac{1}{T}\sum_{i=1}^{n} \Delta s_i - \hat{\beta}_{\rm LR}' \frac{1}{T}\sum_{t=1}^{n} \Delta F_i \simeq \frac{1}{T}\sum_{i=1}^{n} \left\{\Delta s_i - \bar{\beta}_{\rm LR}' \Delta F_i\right\} \\ &= \frac{1}{T}\sum_{i=1}^{n} \left\{\alpha(t_i)\Delta + \left[\beta(t_i) - \bar{\beta}_{\rm LR}\right]' \Delta F_i\right\} + \frac{\sqrt{\Delta}}{T}\sum_{i=1}^{n} \Sigma^{1/2}(t_i) z_i \\ &\simeq \bar{\alpha}_{\rm LR} + \frac{\sqrt{\Delta}}{T}\sum_{i=1}^{n} \Sigma^{1/2}(t_i) z_i, \end{aligned} \tag{65}$$

where $\frac{\sqrt{\Delta}}{\sqrt{T}}\sum_{i=1}^{n} \Sigma^{1/2}(t_i) z_i \xrightarrow{d} N\left(0, {\rm E}\left[\Sigma(t)\right]\right)$ and

$$\bar{\alpha}_{\rm LR} = {\rm E}\left[\alpha(t)\right] + {\rm E}\left[\left[\beta(t) - \bar{\beta}_{\rm LR}\right]' \mu_F(t)\right]. \tag{66}$$

We conclude that$\sqrt{T}(\tilde{\alpha}_{\rm LR} - \bar{\alpha}_{\rm LR}) \xrightarrow{d} N\left(0, {\rm E}\left[\Sigma(t)\right]\right)$.

From the above expressions, we see that the GRS estimator $\tilde{\alpha}_{\rm LR}$ of the long-run alphas in general is inconsistent because it is centered around $\bar{\alpha}_{\rm LR} \neq {\rm E}\left[\alpha(t)\right]$. It is only consistent if the factors are uncorrelated with the loadings such that $\mu_F(t) = \mu$ and $\Lambda_{FF}(t) = \Lambda_{FF}$ are constant, in which case $\bar{\beta}_{\rm LR} = \beta_{\rm LR}$ and $\bar{\alpha}_{\rm LR} = \alpha_{\rm LR}$.

Finally, in the case of constant alphas and betas and homoskedastic errors, $\Sigma(s) = \Sigma$, the variance of our proposed estimator of $\gamma_{\rm LR}$ is identical to the one of the GRS estimator.

Appendix E Two-sided versus one-sided filters

We focus on the estimator of $\beta(t)$; the same arguments apply to the alpha estimator. When using a two-sided symmetric kernel where $\mu_1 = \int K(z) z dz = 0$ and $\mu_2 = \int K(z) z^2 dz < \infty$, the finite-sample variance is, with $\bar{h} = h/T$,

$$\text{var}\left(\hat{\beta}_k(t)\right) \simeq \frac{\Delta}{\bar{h}} v_k(t;\beta) \quad \text{with} \quad v_k(t;\beta) = \kappa_2 \Lambda_{FF}^{-1}(t)\,\sigma_{kk}^2(t), \tag{67}$$

while the bias is given by

$$\text{Bias}\left(\hat{\beta}_k(t)\right) \simeq \bar{h}^2 b_k^{\rm sym}(t) \quad \text{with} \quad b_k^{\rm sym}(t) = \frac{\mu_2}{2}\beta_k^{(2)}(t), \tag{68}$$

where we assume that $\beta_k(t)$ is twice differentiable with second order derivative $\beta_k^{(2)}(t)$. In this case the bias is of order $O\left(h^2\right)$. When a one-sided kernel is used, the variance remains unchanged, but because $\mu_1 = \int K(z) z dz \neq 0$ the bias now takes the form

$$\text{Bias}\left(\hat{\beta}_k(t)\right) \simeq h b_k^{\rm one}(t) \quad \text{with} \quad b_k^{\rm one}(t) = \mu_1 \beta_k^{(1)}(t). \tag{69}$$

The bias is in this case of order $O(h)$ and is, therefore, larger compared with the two-sided kernel estimator.

As a consequence, for the symmetric kernel the optimal global bandwidth minimizing the integrated MSE, $IMSE = \frac{1}{T}\int_0^T \mathrm{E}[\|\hat{\gamma}_{j,\tau} - \gamma_{j,\tau}\|^2]d\tau$, is given by

$$h_k^* = \left(\frac{V_k}{B_k^{\text{sym}}}\right)^{1/5} n^{-1/5}, \tag{70}$$

where $B_k^{\text{sym}} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}(b_k^{\text{sym}}(t_i))^2$ and $V_k = \lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} v(t_i)$ are the integrated versions of the time-varying (squared) bias and variance components. With this bandwidth choice, $\sqrt{IMSE}$ is of order $O\left(n^{-2/5}\right)$. If instead a one-sided kernel is used, the optimal bandwidth is

$$h_k^* = \left(\frac{V_k}{B_k^{\text{one}}}\right)^{1/3} n^{-1/3}, \tag{71}$$

where $B_k^{\text{one}} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}(b_k^{\text{one}}(t_i))^2$, with the corresponding $\sqrt{IMSE}$ being of order $O\left(n^{-1/3}\right)$. Thus, the symmetric kernel estimator's RMSE is generally smaller and substantially smaller if n is large. The two exceptions from this general result are if one wishes to estimate alphas and betas at time $t = 0$ and $t = T$. In these cases, the symmetric kernel suffers from boundary bias while a forward- and backward-looking kernel estimator, respectively, remain asymptotically unbiased. We avoid this case in our empirical work by omitting the first and last years in our sample when estimating conditional alphas and betas.

Appendix F Bandwidth choice for long-run estimators

We sketch the outline of the derivation of an optimal bandwidth for estimating the integrated or long-run betas in a discrete-time setting.[10] We follow the same strategy as in Cattaneo, Crump, and Jansson (2010) and Ichimura and Linton (2005), among others. With $\hat{\beta}_{\text{LR},k}$ denoting the long-run estimators of the alphas and betas for the kth asset, first note that by a third-order Taylor expansion with respect to $m(t) = \Lambda_{FF}(t)\beta(t)$ and $\Lambda_{FF}(t)$,

$$\hat{\beta}_{\text{LR},k} - \beta_{\text{LR},k} = U_{1,n} + U_{2,n} + R_n, \tag{72}$$

where $R_n = O\left(\sup_{1\le i\le n}||\hat{m}_k(t_i) - m_k(t_i)||^3\right) + O\left(\sup_{1\le i\le n}||\hat{\Lambda}_{FF}(t_i) - \Lambda_{FF}(t_i)||^3\right)$ and

$$U_{1,n} = \Delta\sum_{i=1}^{n}\left\{\Lambda_{FF}^{-1}(t_i)[\hat{m}_k(t_i) - m_k(t_i)] - \Lambda_{FF}^{-1}(t_i)\left[\hat{\Lambda}_{FF}(t_i) - \Lambda_{FF}(t_i)\right]\beta_k(t_i)\right\}; \tag{73}$$

$$\begin{aligned} U_{2,n} &= \Delta\sum_{i=1}^{n}\Lambda_{FF}^{-1}(t_i)\left[\hat{\Lambda}_{FF}(t_i) - \Lambda_{FF}(t_i)\right]\Lambda_{FF}^{-1}(t_i)\left[\hat{\Lambda}_{FF}(t_i) - \Lambda_{FF}(t_i)\right]\beta_k(t_i) \\ &\quad -\Delta\sum_{i=1}^{n}\Lambda_{FF}^{-1}(t_i)\left[\hat{\Lambda}_{FF}(t_i) - \Lambda_{FF}(t_i)\right]\Lambda_{FF}^{-1}(t_i)[\hat{m}_k(t_i) - m_k(t_i)]. \end{aligned} \tag{74}$$

Thus, our estimator is (approximately) the sum of a second, and third order U-statistic. We proceed to compute the mean and variance of each of these to obtain an MSE expansion of the estimator as a function of the bandwidth h.

To compute the variance of $U_{1,n}$, define $\phi(W_i, W_j) = a(W_i, W_j) + a(W_i, W_j)$, where $W_i = (\Delta F_i, u_i)$,

$$a(W_i, W_j) = K_{i,j}\Lambda_{FF}^{-1}(t_i)\Delta F_j\varepsilon_j' + K_{i,j}\Lambda_{FF}^{-1}(t_i)\Delta F_j\Delta F_j'[\beta_k(t_j) - \beta_k(t_i)], \tag{75}$$

and $\varepsilon_{k,i} \equiv \sigma_{kk}(t_i)z_i$. Observe that

$$\mathrm{E}\left[\phi(w, W_j)\phi(w, W_j)'\right] = \mathrm{E}\left[a(w, W_j)a(w, W_j)'\right] + \mathrm{E}\left[a(W_j, w)a(W_j, w)'\right] + 2\mathrm{E}\left[a(w, W_j)a(W_j, w)'\right], \tag{76}$$

[10] We wish to thank Matias Cattaneo for helping us with this part.

Here,

$$\begin{aligned}
\mathrm{E}\left[a\left(w,W_j\right)a\left(w,W_j\right)'\right] &= \Delta\sum_{j=1}^{n}K_{i,j}^{2}\Lambda_{FF}^{-1}\left(t_i\right)\sigma_{kk}^{2}\left(t_j\right)\Lambda_{FF}\left(t_j\right)\Lambda^{-1}\left(t_i\right) \\
&\quad +\Delta\sum_{j=1}^{n}K_{i,j}^{2}\Lambda_{FF}^{-1}\left(t_i\right)\Lambda_{FF}\left(t_j\right)\left[\beta_k\left(t_j\right)-\beta_k\left(t_i\right)\right]\left[\beta_k\left(t_j\right)-\beta_k\left(t_i\right)\right]'\Lambda_{FF}\left(t_j\right)\Lambda_{FF}^{-1}\left(t_i\right) \\
&\simeq \frac{1}{h}\times q_1\left(w\right),
\end{aligned} \tag{77}$$

where $q_1\left(w\right)=\kappa_2\Lambda_{FF}^{-1}\left(t_i\right)\sigma_{kk}^{2}\left(t_i\right)$. Similarly,

$$\begin{aligned}
\mathrm{E}\left[a\left(W_t,w\right)a\left(W_t,w\right)'\right] &= \Delta\sum_{j=1}^{n}K_{i,j}^{2}\Lambda_{FF}^{-1}\left(t_j\right)xx'e_k^{2}\Lambda_{FF}^{-1}\left(t_j\right) \\
&\quad +\Delta\sum_{j=1}^{n}K_{i,j}^{2}\Lambda_{FF}^{-1}\left(t_i\right)xe\left[\beta_k\left(t_j\right)-\beta_k\left(t_i\right)\right]'xx'\Lambda_{FF}^{-1}\left(t_j\right) \\
&\quad +\frac{1}{n}\sum_{j=1}^{n}K_{i,j}^{2}\Lambda_{FF}^{-1}\left(t_j\right)xx'\left[\beta_k\left(t_j\right)-\beta_k\left(t_i\right)\right]\left[\beta_k\left(t_j\right)-\beta_k\left(t_i\right)\right]'xx'\Lambda_{FF}^{-1}\left(t_j\right) \\
&\simeq \frac{1}{h}\times q_2\left(w\right),
\end{aligned} \tag{78}$$

where $q_1\left(w\right)=\kappa_2\Lambda_{FF}^{-1}\left(t_i\right)xx'e_k^{2}\Lambda_{FF}^{-1}\left(t_i\right)$, while the cross-product term is of smaller order. Employing the same arguments as in Powell and Stoker (1996), it therefore holds that $\mathrm{var}[U_{1,n}]\simeq n^{-1}V_{\mathrm{LR},kk}+\left(n^{2}h\right)^{-1}\times\Sigma_{\mathrm{LR},kk}$, where

$$\Sigma_{\mathrm{LR},kk}=\mathrm{E}\left[q_1\left(W_t\right)\right]+\mathrm{E}\left[q_2\left(W_t\right)\right]=2\kappa_2\frac{1}{n}\sum_{i=1}^{n}\Lambda_{FF}^{-1}\left(t_i\right)\sigma_{kk}^{2}\left(t_i\right)=2\kappa_2\frac{1}{T}\int_0^T\Lambda_{FF}^{-1}\left(t\right)\sigma_{kk}^{2}\left(t\right)dt, \tag{79}$$

while $\mathrm{var}\left(U_{2,n}\right)$ is of higher order and so can be ignored. In total,

$$MSE(\hat{\gamma}_{\mathrm{LR},k})\simeq\left\|B_{\mathrm{LR},k}^{(1)}\times h^{2}+B_{\mathrm{LR},k}^{(2)}\times\frac{1}{nh}\right\|^{2}+\frac{1}{n}\times V_{\mathrm{LR},kk}+\frac{1}{n^{2}h}\times\Sigma_{\mathrm{LR},kk}. \tag{80}$$

When minimizing this expression with respect to h, we can ignore the two last terms in the above expression since they are of higher order (this can be established using uniform convergence results of, e.g., Kristensen, 2009), and so the optimal bandwidth minimizing the squared component is

$$h_{\mathrm{LR}}^{*}=\begin{cases}\left[-B_{\mathrm{LR},k}^{(1)\prime}B_{\mathrm{LR},k}^{(2)}/||B_{\mathrm{LR},k}^{(1)}||^{2}\right]^{1/3}\times n^{-1/3}, & B_{\mathrm{LR},k}^{(1)\prime}B_{\mathrm{LR},k}^{(2)}<0\\ \left[\frac{1}{2}B_{\mathrm{LR},k}^{(1)\prime}B_{\mathrm{LR},k}^{(2)}/||B_{\mathrm{LR},k}^{(1)}||^{2}\right]^{1/3}\times n^{-1/3} & B_{\mathrm{LR},k}^{(1)\prime}B_{\mathrm{LR},k}^{(2)}>0\end{cases}, \tag{81}$$

because in general $B_{\mathrm{LR},k}^{(1)}\neq-B_{\mathrm{LR},k}^{(2)}$.

References

Aït-Sahalia, Y., 1996. Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527–560.

Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. Journal of Finance 53, 499–547.

Aït-Sahalia, Y., Brandt, M., 2007. Nonparametric consumption and portfolio choice with option-implied state prices. Unpublished working paper. Duke University, Durham, NC.

Andersen, T., Bollerselv, T., Diebold, F., Wu, J., 2006. Realized beta: Persistence and predictability. In: Fomby, T. and Terrell, D. (Eds.), Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series in Honor of R. F. Engle and C. W. J. Granger, Volume B, pp. 1-40. Elsevier, Amsterdam, Netherlands.

Andrews, D., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Ang, A., Chen, J., 2007. CAPM over the long run: 1926-2001. Journal of Empirical Finance 14, 1–40.

Ang, A., Chen, J., Xing, Y., 2006. Downside risk. Review of Financial Studies 19, 1191–1239.

Ang, A., Kristensen, D., 2011. Testing conditional factor models. Discussion Paper No. 01/2011-030, Tilburg University, Network for Studies on Pensions, Aging and Retirement. Tilburg, Netherlands.

Bai, J., Perron, P., 1998. Testing for and estimation of multiple structural changes. Econometrica 66, 47–79.

Bandi, F., 2002. Short-term interest rate dynamics: A spatial approach. Journal of Financial Economics 65, 73–110.

Bandi, F., Phillips, P., 2003. Fully nonparametric estimation of scalar diffusion models. Econometrica 71, 241–283.

Barndorff-Nielsen, O., Hansen, P., Lunde, A., Shephard, N., 2009. Multivariate realized kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Journal of Econometrics 162, 149–169.

Bansal, R., Viswanathan, S., 1993. No arbitrage and arbitrage pricing: A new approach. Journal of Finance 48, 1231-1262.

Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. Review of Financial Studies 13, 1–42.

Boguth, O., Carlson, M., Fisher, A., Simutin, M., 2011. Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. Journal of Financial Economics 102, 363–389.

Brandt, M., 1999. Estimating portfolio and consumption choice: A conditional Euler equations approach. Journal of Finance 54, 1609-1645.

Brown, B.M., 1971. Martingale central limit theorems. Annals of Mathematical Statistics 42, 59–66.

Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136, 163188.

Campbell, J., Cochrane, J., 1999. By force of habit. Journal of Political Economy 107, 105–251.

Cattaneo, M., Crump, R., Jansson, M., 2010. Generalized jackknife estimators of weighted average derivatives. Unpublished working paper, University of Michigan, Ann Arbor, MI.

Chan, N., Wei, C., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Annals of Statistics 15, 1050–1063.

Christopherson, J., Ferson, W., Glassman, D., 1998. Conditioning manager alpha on economic information: Another look at the persistence of performance. Review of Financial Studies 3, 32–43.

Diggle, P., Hutchinson, M., 1989. On spline smoothing with correlated errors. Australian Journal of Statistics 31, 166–82.

Dimson, E., 1979. Risk measurement when shares are subject to infrequent trading. Journal of Financial Economics 7, 197–226.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E., French, K., 1997. Industry costs of equity. Journal of Financial Economics 43, 153–193.

Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607–636.

Fan, J., Zhang, C., Zhang, W., 2001. Generalized likelihood ratio statistics and Wilks phenomenon. Annals of Statistics 29, 153–193.

Ferson, W., Harvey, C., 1991. The variation of economic risk premiums. Journal of Political Economy 99, 385–415.

Ferson, W., Schadt, R., 1996. Measuring fund strategy and performance in changing economic conditions. Journal of Finance 51, 425–462.

Ferson, W., Qian, M., 2004. Conditional evaluation performance: revisited. Research Foundation of the CFA Institute, Charlotteville, VA.

Foster, D., Nelson, D., 1996. Continuous record asymptotics for rolling sample variance estimators. Econometrica 64, 139–174.

French, K., Schwert, G., Stambaugh, R., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Ghysels, E., 1998. On stable factor structures in the pricing of risk: Do time varying betas help or hurt? Journal of Finance 53, 549–574.

Gibbons, M. R., Ross, S., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Gomes, J., Kogan, L., Zhang, L., 2003. Equilibrium cross section of returns. Journal of Political Economy 111, 693–732.

Härdle, W., Hall, P., Marron, J., 1988. How far are automatically chosen regression smoothing parameters from their optimum? Journal of the American Statistical Association 83, 86–101.

Hart, J., 1991. Kernel regression estimation with time series errors. Journal of the Royal Statistical Society, Series B (Methodological) 53, 173–187.

Harvey, C., 2001. The specification of conditional expectations. Journal of Empirical Finance 8, 573–638.

Hayashi, T., Yoshida, N., 2005. On covariance estimation of non-synchronously observed diffusion Processes. Bernoulli 11, 359–379.

Ichimura, H., Linton, O., 2005. Asymptotic expansions for some semiparametric program evaluation estimators. In: Andrews, D., Stock, J., (Eds.), Identification and Inference for Econometric Models: A Festschrift for Tom Rothenberg. Cambridge University Press, Cambridge, United Kingdom.

Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross section of expected returns. Journal of Finance 51, 3–53.

Jostova, G., Philipov, A., 2005. Bayesian analysis of stochastic betas. Journal of Financial and Quantitative Analysis 40, 747–778.

Kanaya, S., Kristensen, D., 2010. Estimation of stochastic volatility models by nonparametric filtering. Research Paper No. 2010–67, Center for Research in Econometric Analysis of Time Series, University of Aarhus, Denmark.

Kristensen, D., 2009. Uniform convergence rates of kernel estimators with heterogeneous, dependent data. Econometric Theory 25, 1433–1445.

Kristensen, D., 2010. Nonparametric filtering of the realised spot volatility: A kernel-based approach. Econometric Theory 26, 60–93.

Kristensen, D., 2011. Nonparametric detection and estimation of structural change. Research paper No. 2011–13, Center for Research in Econometric Analysis of Time Series, University of Aarhus, Denmark.

Lettau, M., Ludvigson, S., 2001a. Consumption, aggregate wealth, and expected stock seturns. Journal of Finance 56, 815–849.

Lettau, M., Ludvigson, S., 2001b. Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time varying. Journal of Political Economy 109, 1238–1287.

Lettau, M., Van Nieuwerburgh, S., 2007. Reconciling the return predictability evidence. Review of Financial Studies 21, 1607–1652.

Lewellen, J., Nagel, S., 2006. The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics 82, 289–314.

Li, Y., Yang, L., 2011. Testing conditional factor models: A nonparametric approach. Forthcoming in Journal of Empirical Finance.

Loughran, T. 1997. Book-to-market across firm size, exchange, and seasonality: Is there an effect? Journal of Financial and Quantitative Analysis 32, 249–268.

Mamaysky, H., Spiegel M., Zhang, H., 2008. Estimating the dynamics of mutual fund alphas and betas. Review of Financial Studies 21, 233–264.

Merton, R., 1980. On estimating the expected return on the market. Journal of Financial Economics 8, 323–361.

Mykland, P., Zhang, L., 2006. ANOVA for diffusions and Ito processes. Annals of Statistics 34, 1931–1963.

Newey, W., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics. Vol. IV, pp. 2111–2245. Elsevier, Amsterdam, Netherlands.

Opsomer, J., Wang, Y., Yang, Y., 2001. Nonparametric regression with correlated errors. Statistical Science 16, 134–153.

Paye, B., Timmermann, A., 2006. Instability of return prediction models. Journal of Empirical Finance 13, 274–315.

Petkova, R., Zhang, L., 2005. Is value riskier than growth? Journal of Financial Economics 78, 187–202.

Phillips, P., 1987. Towards an unified asymptotic theory for autoregression. Biometrika 74, 535–547.

Powell, J., Stock, J., Stoker, T., 1989. Semiparametric estimation of index coefficients. Econometrica 57, 1403–1430.

Powell, J., Stoker, T., 1996. Optimal bandwidth choice for density-weighted averages. Journal of Econometrics 75, 291–316.

Robinson, P., 1989. Nonparametric estimation of time-varying parameters. In: Hackl, P. (Ed.), Statistical Analysis and Forecasting of Economic Structural Changes, pp. 253–264. Springer-Verlag, Berlin, Germany.

Ruppert, D., Sheather, S., Wand, M., 1995. An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association 90, 1257–1270.

Scholes, M., Williams, J., 1977. Estimating betas from non-synchronous data. Journal of Financial Economics 5, 309–328.

Shanken, J., 1990. Intertemporal asset pricing: An empirical investigation. Journal of Econometrics 45, 99–120.

Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. Journal of Finance 52, 1973–2002.

Wang, K., 2003. Asset pricing with conditioning information: A new test. Journal of Finance 58, 161–196.

Zhang, L., 2005. The value premium. Journal of Finance 60, 67–103.