Netspar
Network for Studies on Pensions, Aging and Retirement

Netspar DISCUSSION PAPERS

Andrew Ang and Dennis Kristensen

Testing Conditional Factor Models

Discussion Paper 01/2011-030

Testing Conditional Factor Models*

Andrew Ang†
Columbia University and NBER

Dennis Kristensen‡
Columbia University and CREATES

This Version: 21 January, 2011

Keywords: Nonparametric estimator, time-varying beta, conditional alpha, book-to-market premium, value and momentum.

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

*We thank Pengcheng Wan for research assistance and Kenneth French for providing data. We thank an anonymous referee, Matias Cattaneo, Will Goetzmann, Jonathan Lewellen, Serena Ng, Jay Shanken, Masahiro Watanabe, and Guofu Zhou for helpful comments and suggestions, and seminar participants at Columbia University, Dartmouth University, Georgetown University, Emory University, Federal Reserve Board, MIT, Oxford-Man Institute, Princeton University, SUNY-Albany, Washington University in St. Louis, Yale University, University of Montréal, the American Finance Association 2010, the Banff International Research Station Conference on Semiparametric and Nonparametric Methods in Econometrics 2009, the Econometric Society Australasian meeting 2009, the Humboldt-Copenhagen Conference on Recent Developments in Financial Econometrics 2009, the NBER Summer Institute 2010, and the NYU Five Star conference 2009. Ang acknowledges funding from Netspar. Kristensen acknowledges funding from the Danish National Research Foundation (through CREATES) and the NSF (grant no. SES-0961596).

†Email: aa610@columbia.edu

‡Email: dk2313@columbia.edu

Testing Conditional Factor Models

Abstract

We develop a methodology for estimating time-varying alphas and factor loadings based on nonparametric techniques. We test whether conditional alphas and long-run alphas, which are averages of conditional alphas, are equal to zero and derive test statistics for the constancy of factor loadings. The tests can be performed 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, like the widely used Gibbons, Ross, and Shanken (1989) test, assume that the factor loadings are constant. However, there is overwhelming evidence that factor loadings, especially for the standard 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; Ang and Chen, 2007). Time-varying factor loadings can distort the 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 that tests for the significance of conditional alphas in the presence of time-varying betas. Our factor model is formulated in discrete time. Conditional on the realized alphas and betas, our factor model 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 originally developed by Robinson (1989).[1] Thus, our estimators are highly robust since they impose very weak restrictions on the dynamics of the alphas and betas. Our methodology derives the joint distribution of the estimators, both at each moment in time and their long-run distributions across time.

Our tests are straightforward to apply, powerful, and involve no more than running a series of kernel-weighted OLS regressions for each asset. The tests can be applied to a single asset or jointly across a system of assets. With appropriate technical conditions, we derive a joint asymptotic distribution for the conditional alphas and betas at any point in time. In addition, we construct a test statistic for long-run alphas and betas, which averages the conditional alphas or factor loadings across time, both for a single portfolio and for the multi-asset case. We also derive a test for constancy of the conditional alphas or factor loadings. Interestingly, while the conditional nonparametric estimators converge at slower rates than maximum likelihood

[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.

estimators, we show that tests involving average or long-run conditional alphas converge at the same rate as classical estimators. Consequently, in the special case where betas are constant and there is no heteroskedasticity, our tests for whether the long-run alphas equal zero are asymptotically equivalent to Gibbons, Ross, and Shanken (1989).

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, and most importantly, we provide a general distribution theory for conditional and long-run estimators which the earlier literature did not derive. For example, Lewellen and Nagel's (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 leads to consistent estimates. Lewellen and Nagel further test whether the average conditional alpha is equal to zero using a Fama and MacBeth (1973) procedure. Since 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 bandwidths.

Second, by using a nonparametric kernel to estimate time-varying betas we are able to use all the data efficiently. The nonparametric kernel allows us to estimate conditional alphas and betas at any moment in time. Naturally, our methodology allows for any valid kernel and so nests the one-sided, equal-weighted weighted filters used by French, Schwert, and Stambaugh (1987), Andersen et al. (2006), Lewellen and Nagel (2006), and others, as special cases. All of these studies use truncated, backward-looking windows to estimate second moments.[3]

[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.

[3] Foster and Nelson (1996) derive optimal two-sided filters to estimate covariance matrices under the null of a GARCH data generating process. Foster and Nelson's exponentially declining weights can be replicated by special choice kernel weights (see Kristensen, 2010a, Section 6). 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 GARCH (see for example, Bekaert and Wu, 2000) or a stochastic volatility model (see for example, Jostova and

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 without any curse of dimensionality. 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 (2009). Li and Yang also use nonparametric methods to estimate conditional parameters and formulate a test statistic based on average conditional alphas. However, they 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 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 of estimating time-varying alphas and betas of a conditional factor model. We develop tests of long-run alphas and factor loadings and tests of constancy of the conditional alphas and betas. Section 3 discusses our data. In Sections 4 and 5 we investigate tests of conditional CAPM and Fama-French models on the book-to-market and momentum portfolios. Section 6 concludes. We relegate all technical proofs to the appendix.

2 Statistical Methodology

2.1 Model

Let $R_t = (R_{1,t}, ..., R_{M,t})'$ denote the observed vector of excess returns of M assets at discrete time points $t = 1, 2, ..., n$. We wish to explain the returns through a set of J common tradeable factors, $f_t = (f_{1,t}, ..., f_{J,t})'$ which are also observed at $t = 1, 2, ..., n$. We consider the following

Philipov, 2005; Ang and Chen, 2007).

discrete-time conditional factor model to explain the returns of stock i at time t:

$$R_{i,t} = \alpha_{i,t} + \beta'_{i,t} f_t + \sigma_{ii,t} z_{i,t}, \tag{1}$$

or in matrix notation:

$$R_t = \alpha_t + \beta'_t f_t + \Omega_t^{1/2} z_t, \tag{2}$$

where $\alpha_t = (\alpha_{1,t}, ..., \alpha_{M,t})' \in \mathbb{R}^M$ is the vector of conditional alphas across stocks $i = 1, ..., M$ and $\beta_t = (\beta_{1,t}, ..., \beta_{M,t})' \in \mathbb{R}^{J \times M}$ is the corresponding matrix of conditional betas. The vector $z_t \in \mathbb{R}^M$ contains the errors $z_t = (z_{1,t}, ..., z_{M,t})'$ and the matrix $\Omega_t \in \mathbb{R}^{M \times M}$ contains the covariances, $\Omega_t = [\sigma^2_{ijt}]_{i,j}$.

We collect the alphas and betas in a matrix of time-varying coefficients $\gamma_t = (\alpha_t \ \beta'_t)' \in \mathbb{R}^{(J+1) \times M}$. By defining $X_t = (1 \ f'_t)' \in \mathbb{R}^{(J+1)}$, we can write the model for the M stock returns more compactly as

$$R_t = \gamma'_t X_t + \Omega_t^{1/2} z_t. \tag{3}$$

The conditional covariance of the errors, $\Omega_t \in \mathbb{R}^{M \times M}$, allows for both heteroskedasticity and time-varying cross-sectional correlation.

We assume the error terms satisfy

$$\mathrm{E}\left[z_t | \gamma_s, \Omega_s, X_s, \ 1 \leq s \leq n\right] = 0 \quad \text{and} \quad \mathrm{E}\left[z_t z'_t | \gamma_t, \Omega_t, X_t, \ 1 \leq s \leq n\right] = I_M, \tag{4}$$

where I_M denotes the M-dimensional identity matrix. Equation (4) is the identifying assumption of the model and rules out non-zero correlations between betas and errors and non-zero correlations between factors and errors. This orthogonality assumption is an extension of standard OLS which specifies that errors and factors are orthogonal.[4] Importantly, our framework allows the alphas and betas to be 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 the time-series estimates of the realized conditional alphas, α_t, and the conditional factor loadings, β_t, along with their standard errors. Under the null of a factor

[4] The strict factor structure rules out leverage effects and other nonlinear relationships between asset and factor returns which Boguth et al. (2010) argue may 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. Simulation results show that our long-run alpha estimators perform well under mild misspecification of modestly correlated betas and error terms. We leave these extensions to further research. Appendix A details further technical assumptions.

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, β_t, with factors, f_t, is zero, then the unconditional pricing errors of a conditional factor model are 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 loadings (see Jagannathan and Wang, 1996; Lewellen and Nagel, 2006).[5]

Given the realized alphas and betas at each point in time, we define the long-run alphas and betas for asset i, $i = 1, ..., M$, as

$$\begin{aligned} \alpha_{\mathrm{LR},i} &\equiv \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \alpha_{i,t} \in \mathbb{R}, \\ \beta_{\mathrm{LR},i} &\equiv \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \beta_{i,t} \in \mathbb{R}^J, \end{aligned} \tag{5}$$

We use the terminology "long run" (LR) to distinguish the conditional alpha at a point in time, $\alpha_{i,t}$, from the conditional alpha averaged over the sample, $\alpha_{\mathrm{LR},i}$. When the factors are correlated with the betas, the long-run alphas are potentially very different from OLS alphas. We test the hypothesis that the long-run alphas are jointly equal to zero across M assets:

$$H_0 : \alpha_{\mathrm{LR},i} = 0, \quad i = 1, ..., M. \tag{6}$$

In a setting with constant factor loadings and constant alphas, Gibbons, Ross, and Shanken (1989) develop a test of the null H_0. Our methodology can be considered to be the conditional version of the Gibbons-Ross-Shanken test when both conditional alphas and betas potentially vary over time.

2.2 Nonparametric Estimation

We treat the sequence $\gamma_t \equiv (\alpha_t \ \beta_t)$, $t = 1, ..., n$, as a collection of n parameter vectors by conditioning on the particular sample paths of alphas and betas that generated the data. Without any further restrictions on the model, we cannot identify these parameters since there are as many parameters as observations in data. Following Robinson (1989) and others, we identify γ_t by imposing a smoothness condition: for some (possibly random) functions $\alpha : [0, 1] \mapsto \mathbb{R}^M$ and $\beta : [0, 1] \mapsto \mathbb{R}^{J\times M}$, the sequences of alphas and betas satisfy:

$$\alpha_t = \alpha\left(t/n\right) \quad \text{and} \quad \beta_t = \beta\left(t/n\right). \tag{7}$$

[5] See Appendix F for further details.

We use α and β to denote both the actual sequences and the underlying functions generating them.

The restriction in equation (7) changes the estimation problem from estimating n parameter vectors to a curve-fitting problem. Without any restrictions on the alphas and betas, only one observation, (R_t, f_t), carries information about α_t and β_t. In contrast under equation (7), observations in the vicinity of t carry more and more information about α_t and β_t as the sample size grows, $n \to \infty$, which allows us to estimate them consistently.

Equation (7) implies that we no longer sample from a single model. That is, $\alpha_t = \alpha\left(t/n\right)$ and $\beta_t = \beta\left(t/n\right)$ depend on sample size n and we have a sequence of models as $n \to \infty$. At first glance, this seems non-standard but it is a commonly used tool in econometrics to construct estimators and asymptotic distributions. Appendix B discusses these related literatures. Consistency of an estimator of α_t or β_t requires that the amount of data increases, but merely increasing the length of the time series, n, does not necessarily improve estimation of α_t or β_t at time t; consistency and the construction of an asymptotic distribution requires that the amount of *local* information increases as $n \to \infty$. The smoothness condition (7) accomplishes this.

Our analysis of the model and estimators is done conditional on the particular realization of alphas and betas that generated the data. That is, our analysis assumes the following conditional relation between the observations and the parameters of interest:

$$E\left[R_t|\gamma(\cdot), \Omega(\cdot)\right] = \alpha_t + \beta_t' E\left[f_t|\gamma(\cdot), \Omega(\cdot)\right], \tag{8}$$

where $\gamma(\cdot)$ and $\Omega(\cdot)$ can be thought of as the full sample paths of the alphas, betas, and covariance matrix. Equation (8) identifies this sample path from data, given the particular realization of alphas and betas in our data. In particular,

$$\gamma_t = \Lambda_t^{-1} E\left[X_t R_t'|\gamma(\cdot), \Omega(\cdot)\right], \tag{9}$$

where Λ_t denotes the conditional second moment of the regressors:

$$\Lambda_t := \mathrm{E}\left[X_t X_t'|\gamma(\cdot), \Omega(\cdot)\right]. \tag{10}$$

The time variation in Λ_t reflects potential correlation between factors and betas. If there is zero correlation (and the factors are stationary), then $\Lambda_t = \Lambda$ is constant over time, but in general Λ_t varies over time. One advantage of conducting the analysis conditional on the sample realization is that we can tailor our estimates of the particular realization of alphas and betas.[6]

[6] An alternative approach taken in the analysis of parametric conditional factor models focuses on unconditional

2.3 Conditional Estimators

Given observations of returns and factors at $t = 1, ..., n$, we propose the following local least-squares estimators of the functions $\alpha_i(\tau)$ and $\beta_i(\tau)$ for asset i in equations (1) and (7) at any (normalized) point in time $\tau \in (0, 1)$:

$$(\hat{\alpha}_i(\tau), \hat{\beta}_i(\tau)')' = \arg\min_{(\alpha,\beta)} \sum_{t=1}^{n} K_{h_i}(t/n - \tau)(R_{i,t} - \alpha - \beta' f_t)^2, \tag{11}$$

where $K_h(z) = K(z/h)/h$ with $K(\cdot)$ being a kernel and $h > 0$ a bandwidth. The optimal estimators in equation (11) are simply kernel-weighted least squares and it is easily seen that

$$(\hat{\alpha}_i(\tau), \hat{\beta}_i(\tau)')' = \left[\sum_{t=1}^{n} K_{h_i}(t/n - \tau) X_t X_t'\right]^{-1} \left[\sum_{t=1}^{n} K_{h_i}(t/n - \tau) X_t R_{i,t}'\right]. \tag{12}$$

Given the function estimates, $\hat{\alpha}_i(\tau)$ and $\hat{\beta}_i(\tau)$, natural estimates of the corresponding sequences are

$$\hat{\alpha}_{i,t} = \hat{\alpha}_i(t/n) \quad \text{and} \quad \hat{\beta}_{i,t} = \hat{\beta}_i(t/n).$$

The proposed estimators are sample analogues to equation (9) giving weights to the individual observations according to how close in time they are to the time point of interest, τ. 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 et al. (2006) and Lewellen and Nagel (2006), among others. In common with other non-parametric estimation methods, as long as the kernel is symmetric, the most important choice is not so much the shape of the kernel that matters but the bandwidth interval, h_i.

The bandwidth, h_i, controls the time window used in the estimation for the ith stock, and as such effectively controls how many observations are used to compute the estimated coefficients $\hat{\alpha}_i(\tau)$ and $\hat{\beta}_i(\tau)$ at time τ. A small bandwidth means only observations close to τ are weighted

---

estimates of equation (8):

$$\mathrm{E}[R_t] = \alpha_u + \beta_u' f_u + \mathrm{cov}(\beta_t, \bar{f}_t),$$

where $\bar{f}_t = \mathrm{E}[|\gamma(\cdot), \Omega(\cdot)]$, $f_u = \mathrm{E}[f_t]$, $\alpha_u = \mathrm{E}[\alpha_t]$ and $\beta_u = \mathrm{E}[\beta_t]$. Our interest lies in pinning down the particular realization of the alphas and betas in our data sample, so this unconditional approach is not very informative as it averages over all possible realizations.

and used 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 towards zero at a suitable rate in order for any finite-sample biases and variances to vanish. We discuss the bandwidth choice in Section 2.7.

We run the kernel regression (11) separately stock by stock for $i = 1, ..., M$. 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,

$$\left(\hat{\alpha}\left(\tau\right), \hat{\beta}\left(\tau\right)'\right)' = \left[\sum_{t=1}^{n} K_h\left(t/n - \tau\right) X_t X_t'\right]^{-1} \left[\sum_{t=1}^{n} K_h\left(t/n - \tau\right) X_t R_t'\right].$$

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 may differ widely across stocks and each stock may 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, in order to avoid cumbersome notation, we present the asymptotic results for the estimators $\hat{\alpha}\left(\tau\right)$ and $\hat{\beta}\left(\tau\right)$ assuming one common bandwidth, h, across all stocks. The asymptotic results are identical in the case with bandwidths under the assumption that all the bandwidths converge at the same rate as $n \to \infty$.

To develop the asymptotic theory of the proposed estimators, we assume that the conditional second moments of regressors and errors satisfy conditions analogous to equation (7):

$$\Lambda_t = \Lambda(t/n) \qquad \Omega_t = \Omega(t/n), \tag{13}$$

for some (possibly random) functions $\Lambda(\cdot)$ and $\Omega(\cdot)$. We now state a result regarding the asymptotic properties of the local least-squares estimator of conditional alphas and betas:

Theorem 1 *Assume that (A.1)-(A.4) given in Appendix A hold and the bandwidth is chosen such that $nh \to \infty$ and $nh^5 \to 0$. Then, for any $\tau \in [0, 1]$, $\hat{\gamma}\left(\tau\right) = \left(\hat{\alpha}\left(\tau\right), \hat{\beta}\left(\tau\right)'\right)'$ satisfies*

$$\sqrt{nh}(\hat{\gamma}\left(\tau\right) - \gamma\left(\tau\right)) \xrightarrow{d} N\left(0, \kappa_2 \Lambda^{-1}\left(\tau\right) \otimes \Omega\left(\tau\right)\right), \tag{14}$$

where $\gamma(\cdot)$ and $\Omega(\cdot)$ denote the full trajectories of the alphas, betas, and error variances, respectively, $X_t = (1, f_t')'$, and $\kappa_2 = \int K^2\left(z\right) dz$ (= 0.2821 for the normal kernel).

Furthermore, the conditional estimators are asymptotically independent across any set of distinct points $\tau_1 \neq ... \neq \tau_m$.

As would be expected, our local OLS estimators have a similar asymptotic distribution to standard ("global") OLS. Simple estimators of the two terms appearing in the asymptotic variance in equation (14) are obtained as follows:

$$\hat{\Lambda}(\tau) = \frac{1}{n}\sum_{t=1}^{n} K_h(t/n - \tau) X_t X_t' \quad \text{and} \quad \hat{\Omega}(\tau) = \frac{1}{n}\sum_{t=1}^{n} K_h(t/n - \tau)\hat{\varepsilon}_t \hat{\varepsilon}_t', \tag{15}$$

where $\hat{\varepsilon}_t = R_t - \hat{\alpha}_t - \hat{\beta}_t' f_t$ are the fitted residuals. Due to the asymptotic independence across different values of τ, pointwise confidence bands can easily be computed.

Using Theorem 1, we test the hypothesis that $\alpha(\tau) = 0$ jointly across M stocks for a given value of $\tau \in (0, 1)$ by the following statistic:

$$W(\tau) = \hat{\alpha}(\tau)' \hat{V}_{\alpha\alpha}^{-1}(\tau)\hat{\alpha}(\tau) \xrightarrow{d} \chi_M^2, \tag{16}$$

where $\hat{V}_{\alpha\alpha}(\tau)$ consist of the first $M \times M$ components of $\hat{V}(\tau) = \kappa_2 \hat{\Lambda}^{-1}(\tau) \otimes \hat{\Omega}(\tau) / (nh)$.

2.4 Comments on Theorem 1

Theorem 1 holds for a discrete-time domain, which is the setting of the vast majority of empirical studies estimating conditional alphas, or alphas computed on rolling subsample regressions.[7] Our procedure develops a formal procedure to estimate both time-varying alphas and betas without needing to assume a specific parametric model for the dynamics of γ_t. The alphas and betas can take on any sample path in the data, subject to the (weak) restrictions in Appendix A, including non-stationary and discontinuous cases, and time-varying dependence of conditional betas and factors. Without assumption (7) where the alphas and betas are scaled by sample size, and consequently the assumption of a sequence of models, there is no way to construct a suitable asymptotic distribution in the nonparametric setting. Our methodology applies in the special case where the data are generated from standard time-series models, like random walk or AR(1) processes. In particular, Appendix E shows that our estimator is able to track the trajectory of a random walk, and by extension any ARMA model, and by assuming a sequence of models, we are able to conduct valid asymptotic inference in a non-stationary setting.

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

It is important not to confuse our sequence of discrete-time models as high-frequency sampling in continuous-time models. In continuous time, the conditional variances of the factors and errors shrink as we sample at higher frequencies, while in our setting they do not vanish as $n \to \infty$. Thus, our theoretical results in Theorem 1 are not directly applicable to continuous-time models. However, our estimators can be adapted to a continuous-time setting following Kristensen (2010a), as we show in Appendix D. In continuous time, the alpha (drift) and the beta (diffusion term) estimators no longer converge with the same rate. With infill asymptotics, the conditional variance of the observed factors changes with n and this causes the estimators for α_t and β_t to converge at different rates, as pointed out by Bandi and Phillips (2003). In contrast, the conditional alphas and betas converge at the same rates in our discrete-time setting because the moments are scaled with n in Theorem 1.[8]

In Theorem 1, the rate of convergence is $\sqrt{nh}$ which is the standard rate of convergence for a nonparametric estimator. This is slower than the classical convergence rate of $\sqrt{n}$ since $h \to 0$. However, below, we show that a test for the average alpha across the sample equal to zero converges at the $\sqrt{n}$ rate. A major advantage of our procedure in contrast to most other nonparametric procedures is that our estimators do not suffer from the curse of dimensionality. Since we only smooth over the time variable t, increasing the number of regressors, J, or the number of stocks, M, do not affect the performance of the estimator. A further advantage is that the point estimates $\hat{\alpha}_i(\tau)$ and $\hat{\beta}_i(\tau)$ can be estimated stock by stock, making the procedure easy to implement. This is similar to the classical Gibbons, Ross, and Shanken (1989) test where the alphas and betas are also separately estimated asset by asset.

A closing comment is that bias at end points is a well-known issue for kernel estimators. When a symmetric kernel is used, our proposed estimator suffers from excess bias when τ $(= t/n)$ is close to either 0 or 1; that is at the beginning and end of our sample. In particular, the estimator is asymptotically biased when evaluated at the end points,

$$\mathrm{E}\left[\hat{\gamma}(0)\right] \to \frac{1}{2}\gamma(0) \quad \text{and} \quad \mathrm{E}\left[\hat{\gamma}(1)\right] \to \frac{1}{2}\gamma(1) \quad \text{as } h \to 0.$$

This can be handled in a number of different ways. The first and easiest way, which is also the procedure we follow in the empirical work, is to simply refrain from reporting estimates

[8] Without assuming the analogous smoothness condition (7) in the continuous-time setting, nonparametric estimation of the instantaneous drift is not possible as Merton (1980), Bandi and Phillips (2003), and Kristensen (2010a), among others, point out. Nevertheless, consistency and construction of an asymptotic distribution for long-run alpha estimators is possible and follows Theorem 2. Note that inference for conditional betas is possible without assumption (7) as instantaneous diffusion estimators are identified only with infill asymptotics.

close to the two boundaries: All our theoretical results are established under the assumption that our sample has been observed across (normalized) time points $\tau = t/n \in [-a, 1+a]$ for some $a > 0$ and we then only estimate $\gamma(\tau)$ for $\tau \in [0,1]$. In the empirical work, we do not report the time-varying alphas and betas during the first and last year of our post-1963 sample. Second, as demonstrated in Kristensen (2010a) in a continuous-time setting, adaptive estimators which control for the boundary bias could be used. Two such estimators are boundary kernels and locally linear estimators. The former involves exchanging the fixed kernel K for another adaptive kernel which adjusts to how close we are to the boundary, while the latter uses a local linear approximation of α_t and β_t instead of a local constant one. Usage of these kernels does not affect the asymptotic distributions in Theorem 1 or the asymptotic distributions we derive for long-run alphas and betas in Section 2.5.

2.5 Tests for Long-Run Alphas and Betas

To test the null of whether the long-run alphas are equal to zero (H_0 in equation (6)), we construct an estimator of the long-run alphas in equation (5) from the estimators of the conditional alphas, α_t, and the conditional betas, β_t. A natural way to estimate the long-run alphas and betas for stock i is to simply plug the pointwise kernel estimators into the expressions found in equation (5):

$$\hat{\alpha}_{\mathrm{LR},i} = \frac{1}{n}\sum_{t=1}^{n} \hat{\alpha}_{i,t} \quad \text{and} \quad \hat{\beta}_{\mathrm{LR},i} = \frac{1}{n}\sum_{t=1}^{n} \hat{\beta}_{i,t},$$

where $\hat{\alpha}_{i,t} = \hat{\alpha}_i(t/n)$ and $\hat{\beta}_{i,t} = \hat{\beta}_i(t/n)$. The following theorem states the joint distribution of $\hat{\gamma}_{\mathrm{LR}} = (\hat{\alpha}_{\mathrm{LR}}, \hat{\beta}'_{\mathrm{LR}})' \in \mathbb{R}^{(J+1)\times M}$ where $\hat{\alpha}_{\mathrm{LR}} = (\hat{\alpha}_{\mathrm{LR},1}, ..., \hat{\alpha}_{\mathrm{LR},M})' \in \mathbb{R}^M$ 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.6) given in the Appendix hold. Then,*

$$\sqrt{n}(\hat{\gamma}_{\mathrm{LR}} - \gamma_{\mathrm{LR}}) \xrightarrow{d} N(0, V_{\mathrm{LR}}), \tag{17}$$

where

$$V_{\mathrm{LR}} \equiv \int_0^1 \Lambda^{-1}(\tau) \otimes \Omega(\tau)\, d\tau = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \Lambda_t^{-1} \otimes \Omega_t. \tag{18}$$

In particular,

$$\sqrt{n}(\hat{\alpha}_{\mathrm{LR}} - \alpha_{\mathrm{LR}}) \xrightarrow{d} N(0, V_{\mathrm{LR},\alpha\alpha}),$$

where $V_{\mathrm{LR},\alpha\alpha}$ are the first $M \times M$ components of V_{LR}:

$$V_{\mathrm{LR},\alpha\alpha} = \int_0^1 \Lambda_{\alpha\alpha}^{-1}(\tau) \otimes \Omega(\tau)\, d\tau = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \Lambda_{\alpha\alpha,t}^{-1} \otimes \Omega_t.$$

with

$$\Lambda_{\alpha\alpha,t} = 1 - E[f_t]' E[f_t f_t']^{-1} E[f_t].$$

The asymptotic variance can be consistently estimated by

$$\hat{V}_{\mathrm{LR}} = \frac{1}{n} \sum_{t=1}^{n} \hat{\Lambda}_t^{-1} \otimes \hat{\Omega}_t, \tag{19}$$

where $\hat{\Lambda}_t = \hat{\Lambda}(t/n)$ and $\hat{\Omega}_t = \hat{\Omega}(t/n)$ are given in equation (15).

The long-run estimators converge at the standard parametric rate $\sqrt{n}$, despite the fact that they are based on preliminary estimators $\hat{\gamma}_\tau$ that converge at the slower, nonparametric rate $\sqrt{nh}$. 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 equation (12). This occurs in other semiparametric estimators involving integrals of kernel estimators (see, for example, Newey and McFadden, 1994, Section 8; Powell, Stock, and Stoker, 1989).[9]

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

$$W_{\mathrm{LR}} = n\hat{\alpha}_{\mathrm{LR}}' \hat{V}_{\mathrm{LR},\alpha\alpha}^{-1} \hat{\alpha}_{\mathrm{LR}} \in \mathbb{R}_+, \tag{20}$$

where $\hat{V}_{\mathrm{LR},\alpha\alpha}$ is an estimator of the variance of $\hat{\alpha}_{\mathrm{LR}}$,

$$\hat{V}_{\mathrm{LR},\alpha\alpha} = \frac{1}{n} \sum_{t=1}^{n} \hat{\Lambda}_{\alpha\alpha,t}^{-1} \otimes \hat{\Omega}_t,$$

with

$$\hat{\Lambda}_{\alpha\alpha,t}^{-1} = 1 - \hat{E}[f_t]' \hat{E}[f_t f_t']^{-1} \hat{E}[f_t],$$

[9] Two closely related two-step semiparametric estimators are the Newey and McFadden (1994) nonparametric estimator of consumer surplus and Powell, Stock, and Stoker's (1989) semiparametric estimation of index coefficients. These can be written as integrals over a ratio of two kernel estimators. Our long-run alpha and beta estimators are similar because the first-step estimators $\hat{\gamma}(\tau) = (\hat{\alpha}(\tau), \hat{\beta}(\tau)')'$ in equation (12) are ratios of kernel functions and the long-run alpha and betas are integrals of these kernel estimators.

and

$$\hat{E}\left[f_t\right] = \frac{1}{n}\sum_{s=1}^{n} K_h\left(s/n - t/n\right) f_s \quad \text{and} \quad \hat{E}\left[f_t f_t'\right] = \frac{1}{n}\sum_{s=1}^{n} K_h\left(s/n - t/n\right) f_s f_s'.$$

As a direct consequence of Theorem 2, we obtain

$$W_{\mathrm{LR}} \xrightarrow{d} \chi^2_M. \tag{21}$$

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

A special case of our model is when the factor loadings are constant with $\beta_t = \beta \in \mathbb{R}^{J\times M}$ for all t. Under the null that beta is indeed constant, $\beta_t = \beta$, and with no heteroskedasticity, $\Omega_t = \Omega$ for all t, the asymptotic distribution of $\sqrt{n}(\hat{\alpha}_{\mathrm{LR}} - \alpha_{\mathrm{LR}})$ is identical to the standard Gibbons, Ross, and Shanken (1989) test. This is shown in Appendix F. 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-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 alphas.

2.6 Tests for Constancy of Alphas and Betas

We wish to test for constancy of a subset of the time-varying parameters of stock i, $\gamma_{i,t} = (\alpha_{i,t}, \beta_{i,t}')' \in \mathbb{R}^{J+1}$. We split up the set of regressors, $X_t = (1, f_t')'$ and coefficients, $\gamma_{i,t}$, into two components (after possibly rearranging the regressors): $\gamma_{i1,t} \in \mathbb{R}^m$, which is the set of coefficients we wish to test for constancy with $X_{1,t} \in \mathbb{R}^m$ the associated regressors, and $\gamma_{i2,t} \in \mathbb{R}^{J+1-m}$ the remaining coefficients with $X_{2,t} \in \mathbb{R}^{J+1-m}$ the remaining regressors, respectively. Using this notation we can rewrite our model as:

$$R_{i,t} = \gamma_{i1,t}' X_{1,t} + \gamma_{i2,t}' X_{2,t} + \sigma_{ii,t} z_{i,t}. \tag{22}$$

We consider the following hypothesis:

$$H_1 : \gamma_{i1,t} = \gamma_{i1} \text{ for all } 1 \leq t \leq n.$$

Under the null hypothesis, m of the $J+1$ coefficients are constant whereas under the alternative hypothesis all $J+1$ coefficients vary through time. Our hypothesis covers both the situation of constant alphas,[10]

$$H_1' : \alpha_{it} = \alpha_i \in \mathbb{R} \quad \text{with} \quad X_{1,t} = 1,\, X_{2,t} = f_t,\, \gamma_{i1,t} = \alpha_{i,t},\, \gamma_{i2,t} = \beta_{i,t},$$

and constant betas,

$$H_1'' : \beta_{i,t} = \beta_i \in \mathbb{R}^J \quad \text{with} \quad X_{1,t} = f_t,\, X_{2,t} = 1,\, \gamma_{i1,t} = \beta_{i,t},\, \gamma_{i2,t} = \alpha_{i,t}.$$

Under H_1, we obtain an estimator of the constant parameter vector γ_1 by using local profiling. First, we treat γ_{i1} as known and estimate $\gamma_{i2}(\tau)$ by

$$\hat{\gamma}_{i2}(\tau) = \arg\min_{\gamma_{i2}} \sum_{t=1}^{n} K_h(t/n - \tau) \left[R_{i,t} - \gamma_{i1}' X_{1,t} - \gamma_{i2}' X_{2,t}\right]^2 = \tilde{m}_{R_i,t}(\tau) - \tilde{m}_1(\tau)\gamma_{i1}, \quad (23)$$

where

$$\begin{aligned}
\hat{m}_{R_i}(\tau) &= \left[\sum_{t=1}^{n} K_h(t/n - \tau) X_{2,t} X_{2,t}'\right]^{-1} \left[\sum_{t=1}^{n} K_h(t/n - \tau) X_{2,t} R_{i,t}'\right] \in \mathbb{R}^{J+1-m} \\
\hat{m}_1(\tau) &= \left[\sum_{t=1}^{n} K_h(t/n - \tau) X_{2,t} X_{2,t}'\right]^{-1} \left[\sum_{t=1}^{n} K_h(t/n - \tau) X_{2,t} X_{1,t}'\right] \in \mathbb{R}^{(J+1-m)\times m}.
\end{aligned}$$

The corresponding sequences $\hat{m}_{R_i,t} := \hat{m}_{R_i}(t/n)$ and $\hat{m}_{1,t} := \hat{m}_1(t/n)$ are estimators of

$$m_{R_i,t} = E\left[X_{2,t} X_{2,t}'\right]^{-1} E\left[X_{2,t} R_{i,t}'\right], \quad m_{1,t} = E\left[X_{2,t} X_{2,t}'\right]^{-1} E\left[X_{2,t} X_{1,t}'\right]. \quad (24)$$

In the second stage, we obtain an estimator of the constant component γ_{i1}. We do this by substituting the conditional estimator $\hat{\gamma}_{i2,t} = \hat{\gamma}_{i2}(\tau/n)$ into the weighted least-squares criterion $Q(\gamma_{i1})$ given by:

$$Q(\gamma_{i1}) = \sum_{t=1}^{n} \hat{\sigma}_{ii,t}^{-2} \left[R_{i,t} - \gamma_{i1}' X_{1,t} - \hat{\gamma}_{i2,t}' X_{2,t}\right]^2 = \sum_{i=1}^{n} \hat{\sigma}_{ii,t}^{-2} \left[\hat{R}_{i,t} - \gamma_{i1}' \hat{X}_{1,t}\right]^2,$$

where $\hat{X}_{1,t} = X_{1,t} - \hat{m}_{1,t}' X_{2,t} \in \mathbb{R}^m$ and $\hat{R}_{i,t} = R_{i,t} - \hat{m}_{R_i,t}' X_{2,t} \in \mathbb{R}$, and $\hat{\sigma}_{ii,t}^2$ is an estimate of the unrestricted conditional time t variance of stock i given in equation (15). We then choose our estimator to minimize $Q(\gamma_{i1})$, which is again a simple least-squares problem with solution:

$$\tilde{\gamma}_{i1} = \left[\sum_{t=1}^{n} \hat{\sigma}_{ii,t}^{-2} \hat{X}_{1,t} \hat{X}_{1,t}'\right]^{-1} \left[\sum_{t=1}^{n} \hat{\sigma}_{ii,t}^{-2} \hat{X}_{1,t} \hat{R}_{i,t}\right]. \quad (25)$$

[10] To test $H_1 : \alpha_{j,t} = 0$ for all $1 \le t \le n$ simply set $\gamma_{j1} = \hat{\gamma}_{j1} = 0$.

The above estimator similar to to the residual-based estimator of Robinson (1988) and the local linear profile estimator of Fan and Huang (2005). Fan and Huang demonstrate that in a cross-sectional framework with homoskedastic errors, the estimator of γ_{i1} is semiparametric efficient. Substituting equation (25) back into equation (23), the following estimator of the nonparametric component appears, $\tilde{\gamma}_{i2,t} = \hat{m}_{R_i,t} - \hat{m}_{1,t}\tilde{\gamma}'_{i1}$.

Once the restricted estimators have been computed, we test H_1 by comparing the unrestricted and restricted model fit through a F-statistic. Introducing the rescaled errors under the full model and under H_1 respectively as,

$$\hat{z}_{i,t} = \hat{\sigma}_{ii,t}^{-1}\hat{\varepsilon}_{i,t} \quad \text{and} \quad \tilde{z}_{i,t} = \hat{\sigma}_{ii,t}^{-1}\tilde{\varepsilon}_{i,t},$$

where

$$\hat{\varepsilon}_{i,t} = R_{i,t} - \hat{\gamma}'_{i1,t}X_{1,t} - \hat{\gamma}'_{i2,t}X_{2,t}$$

are the residuals under the alternative and

$$\tilde{\varepsilon}_{i,t} = R_{i,t} - \tilde{\gamma}'_{i1}X_{i,t} - \tilde{\gamma}'_{i2,t}X_{2,t}$$

are the residuals under the null, we can compute the sums of (rescaled) squared residuals by:

$$SSR_i = \sum_{t=1}^{n} \hat{z}_{i,t}^2 \quad \text{and} \quad SSR_{i,1} = \sum_{t=1}^{n} \tilde{z}_{it}^2.$$

The F-statistic then takes the following form:

$$F_i = \frac{n}{2}\frac{SSR_{i,1} - SSR_i}{SSR_i}. \tag{26}$$

The proposed test statistic is related to the generalized likelihood-ratio test statistics advocated in Fan, Zhang and Zhang (2001). To state the asymptotic distribution, we introduce the following notation: For a random sequence $F = F_n$, we write $F \overset{as}{\sim} \chi_b^2/r$ for a sequence $b = b_n \to \infty$ and a constant r if $(rF - b)/\sqrt{2b} \xrightarrow{d} N(0,1)$, cf. Fan, Zhang and Zhang (2001, p. 156).

Theorem 3 *Assume that assumptions (A.1)-(A.6) given in the Appendix hold. Under H_1:*

$$\sqrt{n}(\tilde{\gamma}_{i1} - \gamma_{i1}) \xrightarrow{d} N\left(0, \Sigma_{ii}^{-1}\right), \tag{27}$$

where, with $A_t := X_{1,t} - m'_{1,t}X_{2,t}$,

$$\Sigma_{ii} = \lim_{n\to\infty} \sum_{t=1}^{n} \sigma_{ii,t}^{-2} A_t A'_t. \tag{28}$$

The F-statistic satisfies

$$F_i \overset{as}{\sim} \chi^2_{q\mu}/q,$$

where

$$q = \frac{K(0) - 1/2\kappa_2}{\int \left[K(z) - 1/2\,(K * K)(z)\right]^2 dz} \quad \textit{and} \quad \mu = \frac{2m}{h}\left[K(0) - 1/2\kappa_2\right].$$

For Gaussian kernels, $q = 2.5375$ and $\mu = 2mc/h$ where $c = 0.7737$.

An important point is that the limiting distribution of the test statistic under the null is nuisance parameter free since μ and q only depend on the bandwidth (h), the number of parameter restrictions (m), and the chosen kernel (K). All of these are known, which facilitates the implementation of the test. Moreover, Fan, Zhang and Zhang (2001) demonstrate in a cross-sectional setting that test statistics of the form of F_i are, in general, asymptotically optimal and can even be adaptively optimal. The above test procedure can easily be adapted to construct joint tests of parameter constancy across multiple stocks. For example, to test for joint parameter constancy jointly across all stocks, simply set $R_{i,t} = R_t$ in the above expressions.

2.7 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. Previous authors using high frequency data to estimate covariances or betas, such as Andersen et al. (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 towards zero as the sample size grows, $h \equiv h_n \to 0$ as $n \to \infty$ in order to remove any biases of the estimator.[11] However, a given sample requires a particular choice of h, which we now discuss.

[11] For very finely sampled data, especially intra-day data, non-synchronous trading may induce bias. There is a large literature on methods to handle non-synchronous 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 where $f_t = R_{m,t}$ is the market return. As an ad-hoc adjustment for non-synchronous 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 (2009). More recently, there has been a growing literature on how to adjust for non-synchronous effects in the estimation of realized volatility. Again, these can be carried over to our setting. For example, it is possible to adapt the methods proposed in, for example, Hayashi and Yoshida (2005) or Barndorff-Nielsen et al. (2009) to adjust for the biases due to non-synchronous observations.

We advocate using two-sided symmetric kernels because, in general, the bias from two-sided symmetric kernels is lower than for one-sided filters. In our data where n is over 10,000 daily observations, the improvement in the integrated root mean squared error (RMSE) using a Gaussian filter over a backward-looking uniform filter can be quite 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 G.

There are two bandwidth selection issues unique to our estimators that we now discuss, which are separate bandwidth choices for the conditional and long-run estimators. We choose one bandwidth for the point estimates of conditional alphas and betas and a different bandwidth for the long-run alphas and betas. 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.[12] Section 2.7.1 discusses bandwidth choices for the conditional estimates of alphas and betas while Section 2.7.2 treats the problem of specifying the bandwidth for the long-run alpha and beta estimators.

2.7.1 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) 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 point in time may be more useful; the following procedure can be adapted to this purpose by replacing all sample averages by subsample ones in the expressions below.

For a symmetric kernel with $\mu_2 := \int K(z) z^2 dz$, the optimal global bandwidth that minimizes the (integrated) MSE for stock i is

$$h_i^* = \left(\frac{V_i}{B_i^2}\right)^{1/5} n^{-1/5}, \tag{29}$$

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.

[12] We conducted simulation studies showing that the proposed methods work well in practice.

where $V_i = \int_0^1 v_i(s)\,ds = \lim_{n\to\infty} n^{-1}\sum_{t=1}^n v_{i,t}$ and $B_i^2 = \int_0^1 b_i^2(s)\,ds = \lim_{n\to\infty} n^{-1}\sum_{t=1}^n b_{i,t}^2$ are the integrated time-varying variance and squared-bias components with the pointwise variance and bias given by

$$v_{i,t} = \kappa_2 \Lambda_t^{-1} \otimes \sigma_{ii,t}^2 \quad \text{and} \quad b_{i,t} = \mu_2 \gamma_{i,t}^{(2)}.$$

Here, $\sigma_{ii,t}^2$ is the diagonal component of $\Omega_t = [\sigma_{ij,t}^2]_{i,j}$, while $\gamma_{i,t}^{(2)} = \gamma_i^{(2)}(t/n)$ with $\gamma_i^{(2)}(\cdot)$ denoting the second order derivative of $\gamma_i(\tau)$. Ideally, we would compute v_i and b_i in order to obtain the optimal bandwidth given in equation (29). However, these depend on unknown components, Λ_t, $\gamma_{i,t}$, and $\sigma_{ii,t}^2$. In order to implement the bandwidth choice we specify a two-step method to provide preliminary estimates of these unknown quantities.[13]

1. Treat $\Lambda_t = \Lambda$ and $\sigma_{ii,t}^2 = \sigma^2$ as constant, and $\gamma_{i,t} = a_{i0} + a_{i1}t/n + ... + a_{ip}(t/n)^p$ as a polynomial. We obtain parametric least-squares estimates $\hat{\Lambda}$, $\hat{\sigma}_{ii}^2$ and $\hat{\gamma}_{i,t} = \hat{a}_{i0} + \hat{a}_{i1}t/n + ... + \hat{a}_{ip}(t/n)^p$. Compute for each stock ($i = 1, ..., M$)

$$\hat{V}_{i,1} = \kappa_2 \hat{\Lambda}^{-1} \otimes \hat{\sigma}_{ii}^2 \quad \text{and} \quad \hat{B}_{i,1}^2 = \mu_2^2 \frac{1}{n}\sum_{t=1}^n (\hat{\gamma}_{i,t}^{(2)})^2,$$

   where $\hat{\gamma}_{i,t}^{(2)} = 2\hat{a}_{i2} + 6\hat{a}_{i3}(t/n) + ... + p(p-1)\hat{a}_{ip}(t/n)^{p-2}$. Then, using these estimates we compute the first-pass bandwidth

$$\hat{h}_{i,1} = \left[\frac{\hat{V}_{i,1}}{\hat{B}_{i,1}^2}\right]^{1/5} \times n^{-1/5}. \tag{30}$$

2. Given $\hat{h}_{i,1}$, compute the kernel estimators $\hat{\gamma}_{i,t} = \hat{\Lambda}_t^{-1} n^{-1}\sum_{s=1}^n K_{\hat{h}_{i,1}}(s/n - t/n) X_s R_{i,s}'$, where $\hat{\Lambda}_t$ and $\hat{\Omega}_t$ are computed as in equation (15) with $h = \hat{h}_{i,1}$.

$$\hat{V}_{i,2} = \kappa_2 \frac{1}{n}\sum_{t=1}^n \hat{\Lambda}_t^{-1} \otimes \hat{\sigma}_{ii,t}^2 \quad \text{and} \quad \hat{B}_{i,2}^2 = \mu_2^2 \frac{1}{n}\sum_{t=1}^n (\hat{\gamma}_{i,t}^{(2)})^2,$$

   where $\hat{\gamma}_{i,t}^{(2)} = \hat{\gamma}_i^{(2)}(t/n)$ with $\hat{\gamma}_i^{(2)}(\tau)$ being the second derivative of the kernel estimator with respect to τ. These are in turn used to obtain the second-pass bandwidth:

$$\hat{h}_{i,2} = \left[\frac{\hat{V}_{i,2}}{\hat{B}_{i,2}^2}\right]^{1/5} \times n^{-1/5}. \tag{31}$$

[13] 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 bandwidth in equation (31).

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), and 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.[14]

One could alternatively use (generalized) cross-validation (GCV) 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 3-5 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 GCV procedures is likely due to a number of factors. First, it is well-known that cross-validated bandwidths may 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 GCV procedures are used on time-series data as found in various studies (Diggle and Hutchinson, 1989; Hart, 1991; Opsomer, Wang, and Yang, 2001).

2.7.2 Bandwidth for Long-Run Estimators

To estimate the long-run alphas and betas we re-estimate the conditional coefficients by under-smoothing relative to the bandwidth in equation (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 the conditional alphas and betas will result in over-smoothing.

Ideally, we would choose an optimal long-run bandwidth to minimize the mean-squared error $\mathrm{E}\left[||\hat{\gamma}_{\mathrm{LR},i} - \gamma_{\mathrm{LR},i}||^2\right]$, which we derive in Appendix H. As demonstrated there, the bandwidth used for the long-run estimators should be chosen to be of order $h_{\mathrm{LR},i} = O\left(n^{-1/3}\right)$, when the alphas and betas are twice continuously differentiable functions of time. Thus, the optimal bandwidth for the long-run estimates is required to shrink at a faster rate than the one used for pointwise estimates where the optimal rate is $n^{-1/5}$ as we saw above.

[14] The order of the polynomial is an initial belief on the underlying smoothness of the process; it does not imply that a polynomial of this order fits the estimated conditional parameters.

In our empirical work, we select the bandwidth for the long-run alphas and betas by first computing the optimal second-pass conditional bandwidth $\hat{h}_{i,2}$ in equation (31) and then scaling this down by setting

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

3 Data

In our empirical work, we consider two specifications of conditional factor models: a conditional CAPM where there is a single factor which is the market excess return and a conditional version of the Fama and French (1993) model where the three factors are the market excess return, MKT, and two zero-cost mimicking portfolios, which are 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.[15] The book-to-market portfolios are rebalanced annually at the end of June while the momentum portfolios are rebalanced every month sorting on prior returns from over the past two to twelve months. We use the Fama and French (1993) factors, MKT, SMB, and HML as explanatory factors. All our data is at the daily frequency from July 1963 to December 2007. 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 end-point 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 summary statistics of our factors. We report annualized means and standard deviations. The market premium is 5.32% compared to 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

[15] These are available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

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 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

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 $\hat{h}_{i,2}$ in equation (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 there is significant time variation in beta 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).

We estimate conditional alphas and betas at the end of each month, and for these monthly estimates compute their standard deviations over the sample in the columns labeled "Stdev of Conditional Estimates." Below, we further characterize the time variation of these monthly conditional estimates. The standard deviation of book-to-market conditional alphas is small, at 0.035. In contrast, conditional betas of the book-to-market strategy have much larger time variation with 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 10 book-to-market portfolios are equal to zero using the Wald test of equation (20). 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.

Figure 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 to a long-run alpha of 6.81%. Thus accounting for time-varying betas has reduced the OLS alpha by approximately only 1.1%.

4.2 Time Variation of Conditional Alphas and Betas

In this section, we characterize the time variation of conditional alphas and betas from the one-factor market model. We begin by testing for constant conditional alphas or 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. Note that 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.

Figure 2 charts the annualized estimates of conditional alphas and betas for the growth (decile 1) and value (decile 10) portfolios at a monthly frequency. We plot 95% confidence bands in dashed lines. In Panel A 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.

The contrast between the wide standard errors for the conditional alphas in Panel A of Figure 2 compared to the tight confidence bands for the long-run alphas in Table 2 is due to the following reason. Conditional alphas at a point in time are hard to estimate as only observations close to that point in time provide useful information. In our framework, the conditional estimators converge at the nonparametric rate $\sqrt{nh}$, which is less than the classical rate $\sqrt{n}$ and thus the conditional standard error bands are quite wide. This is exactly what Figure 2 shows and what Ang and Chen (2007) pick up in an alternative parametric procedure.

In comparison, the long-run estimators converge at the standard rate $\sqrt{n}$ causing the long-run alphas to have much tighter standard error bounds than the conditional alphas. The tests for constancy of the conditional estimators also converge at rate $\sqrt{n}$. Intuitively, the long-run estimators exploit the information in the full conditional time series: while the standard errors for a given time point τ are wide, the long-run and constancy tests recognize and exploit the information from all τ. Note that Theorem 1 shows that the conditional alphas at different points in time are asymptotically uncorrelated. Intuitively, as averaging occurs over the whole sample, the uncorrelated errors in individual point estimates diversify away as the sample size increases.

Panel B of Figure 2 plots conditional betas of the growth and value deciles. Conditional factor loadings are estimated relatively precisely with tight 95% confidence bands. Growth betas are largely constant around 1.2, except after 2000 where 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 one 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.

In Figure 3, we plot conditional alphas and betas of the book-to-market strategy. Since the conditional alphas and betas of growth stocks are fairly flat, almost all of the time variation of the conditional alphas and betas of the book-to-market strategy is driven by the conditional alphas and betas of the decile 10 value stocks. Figure 3 also overlays estimates of conditional alphas and betas from a backward-looking, flat 12-month filter. Similar filters are employed by Andersen et al. (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.[16] However, the

[16] The standard error bands of the uniform filters (not shown) are much larger than the standard error bands of

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 to 0.206 for the optimal two-sided conditional beta estimates. This implies that Lewellen and Nagel's (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.

Several authors like Jagannathan and Wang (1996) and Lettau and Ludvigson (2001b) 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; 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.

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 NBER recessions conditional betas also go the wrong way and tend to be lower. The industrial production, term spread, Lettau and Ludvigson's (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 in-

the optimal estimates.

struments 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, there is little evidence that betas of value stocks are higher during times when the market risk premium is high.

4.3 Tests of the Conditional Fama-French (1993) Model

We now investigate alphas and betas 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-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-French model. The conditional Fama-French model over-compensates 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, there is 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 and HML loadings in the conditional Fama-French model.

We plot the conditional factor loadings in Figure 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 upwards, particularly during the late 1990s bull market. This is a period where value stocks performed poorly and the high SMB loadings translate into larger negative conditional Fama-French alphas during this time. After 2000, the SMB loadings decrease to end the sample around 0.25.

Figure 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 actually decrease for the book-to-market strategy. One may 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 precisely what the book-to-market strategy does. However, the breakpoints of the HML factor are quite 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 10% extremes covary quite 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-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-French model cannot price the momentum portfolios.

Table 7 shows that there is no pattern 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-French loadings in Figure 5 which graphs the long-run alphas of the momentum portfolios 1-10 and the long-run alpha of the momentum strategy (portfolio 11 on the graph). 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. Figure 5 shows that the reduction of the absolute values of OLS alphas compared to the long-run conditional alphas is particularly large for both the extreme loser and winner deciles.

Figure 6 shows that all the Fama-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-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-French betas are 0.386, 0.584, and 0.658 for MKT, SMB, and HML, respectively. Figure 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 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-French factor loadings still cannot completely price the momentum portfolios.

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 point in 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) statistics.

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-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-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 tests like Gibbons, Ross, and Shanken (1989). 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. While our empirical work refrains from reporting conditional estimates close to the beginning and

end of the sample and so does not suffer from this bias, boundary kernels and locally linear estimators can be used to provide conditional estimates at the endpoints. 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 by specifying the following sequence of (vector) models:

$$R_{n,t} = \gamma'_{n,t} X_{n,t} + \Omega_{n,t}^{1/2} z_t, \quad i = 1, ..., n,$$

where $\gamma_{n,t} = \left(\alpha_{n,t}, \beta'_{n,t}\right)'$ and $X_{n,t} = \left(1, f'_{n,t}\right)'$. We here allow for the factors to depend on sample size. We assume that we have observed data at $-an \leq t \leq (1+a)\, n$ for some fixed $a > 0$ to avoid any boundary issues and keep the notation simple. We introduce the conditional second moment of the regressors,

$$\Lambda_{n,t} = \mathrm{E}\left[X_{n,t} X'_{n,t} | \gamma, \Omega\right],$$

where $\gamma = \{\gamma_1, \gamma_2 ...\}$ and $\Omega = \{\Omega_1, \Omega_2, ...\}$. Throughout the appendix all arguments are made conditional on the sequences γ and Ω. For notational convenience, we suppress this dependence in the following and so, for example, write $E\left[X_{n,t} X'_{n,t}\right]$ for $E\left[X_{n,t} X'_{n,t} | \gamma, \Omega\right]$.

Let $C^r[0,1]$ denote the space of r times continuously differentiable functions on the unit interval, $[0,1]$. We impose the following assumptions conditional on γ and Ω:

A.1 The sequences $\gamma_{n,t}$, $\Lambda_{n,t}$, and Ω_t satisfy:

$$\gamma_{n,t} = \gamma(t/n) + o(1), \quad \Lambda_{n,t} = \Lambda(t/n) + o(1), \quad \Omega_{n,t} = \Omega(t/n) + o(1),$$

for some functions $\gamma : [0,1] \mapsto R^{(J+1)\times M}$, $\Lambda : [0,1] \mapsto R^{(J+1)\times(J+1)}$ and $\Omega : [0,1] \mapsto R^{M\times M}$ which all lie in $C^r[0,1]$ for some $r \geq 2$. Furthermore, $\Lambda(\tau)$ and $\Omega(\tau)$ are positive definite for any $\tau \in [0,1]$.

A.2 The following moment conditions hold: $\sup_{n\geq 1} \sup_{t\leq n} E\left[\|X_{n,t}\|^s\right] < \infty$ and $\sup_{n\geq 1} \sup_{t\leq n} E\left[\|z_t\|^s\right] < \infty$ for some $s > 8$. The sequence $\{R_{n,t}, X_{n,t}, z_t\}$, $i = 1, ..., n$, is β-mixing where the mixing coefficients are bounded, $\beta_n(i) \leq \beta(i)$, with the bounding sequence $\beta(i)$ satisfying $\beta(i) = O\left(i^{-b}\right)$ for some $b > 2(s-1)/(s-2)$.

A.3 $E[z_t | X_{n,t}] = 0$ and $E[z_s z'_t | X_{n,t}] = I_M$ if $s = t$ and zero otherwise for all $1 \leq s, t \leq n$, $n \geq 1$.

A.4 The kernel K satisfies: There exists $B, L < \infty$ such that either (i) $K(u) = 0$ for $\|u\| > L$ and $|K(u) - K(u')| \leq B\|u - u'\|$, or (ii) $K(u)$ is differentiable with $|\partial K(u)/\partial u| \leq B$ and, for some $\nu > 1$, $|\partial K(u)/\partial u| \leq B\|u\|^{-\nu}$ for $\|u\| \geq L$. Also, $|K(u)| \leq B\|u\|^{-\nu}$ for $\|u\| \geq L$, $\int_{\mathbb{R}} K(z)\, dz = 1$, $\int_{\mathbb{R}} z^i K(z)\, dz = 0$, $i = 1, ..., r-1$, and $\int_{\mathbb{R}} |z|^r K(z)\, dz < \infty$.

A.5 The covariance matrix Σ defined in equation (28) is non-singular.

A.6 The bandwidth satisfies $nh^{2r} \to 0$, $\log^2(n) / \left(nh^2\right) \to 0$ and $1/\left(n^{1-\epsilon} h^{7/4}\right) \to 0$ for some $\epsilon > 0$.

The smoothness conditions in (A.1) rule out jumps in the coefficients; Theorem 1 remains valid at all points where no jumps has occurred, and we conjecture that Theorems 2 and 3 remain valid with a finite jump activity since this will have a minor impact as we smooth over the whole time interval. The mixing and moment conditions in (A.2) are satisfied by most standard time-series models allowing, for example, f_t to follow an ARMA model. The requirement that eighth moments exist can be weakened to fourth moments in Theorem 1, but for simplicity we maintain this assumption throughout. Assumption (A.3) is a stronger version of equation (4) since the conditional moment restrictions on z_{t_i} are required to hold conditional on all past and present values of γ and Ω. In particular, it rules out leverage effects; 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). The assumption (A.4) imposed on the kernel K are satisfied by most kernels including the Gaussian and the uniform kernel. For $r > 2$, the requirement that the first $r-1$ moments are zero does however not hold for these two standard kernels. This condition is only needed for the semiparametric estimation and in practice the impact of using such so-called higher-order kernels is negiglible. The requirement in (A.5) that $\Sigma > 0$ is an identification condition used to identify the constant component in γ_t under H_1; this is similar to the condition imposed in Robinson (1988). The conditions on the bandwidth in (A.6) are needed for the estimators of the long-run alphas and betas and under the null of constant alphas or betas, and entails undersmoothing in the kernel estimation.

B Related Statistical Models

The assumption that the parameters, $\gamma(t/n) = (\alpha(t/n) \ \beta(t/n))$, depend on sample size, n, is also made in several statistical models. For example, the literature on structural change (or break points) employs the same assumption that normalizes time, t, by the sample size, n. 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 normalize time t by sample size n such that the number of observations before 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. As an example for the simplest case with a single break point, the unstable regression parameters (in our case γ_t) satisfy:

$$\gamma_{n,t} = \bar{\gamma}_1 \mathbb{I}\{t/n \leq \tau_0\} + \bar{\gamma}_2 \mathbb{I}\{t/n > \tau_0\}, \tag{B.1}$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function. In equation (B.1), $\tau_0 \in (0,1)$ is the proportion of data observed before the break point, and the values of γ_t before and after the break point are denoted by $\bar{\gamma}_1$ and $\bar{\gamma}_2$, respectively. This is a special case of our general framework where the function $\gamma(\cdot) = (\alpha(\cdot), \beta(\cdot))$ has been chosen as a step function. Lin and Teräsvirta (1994) take a similar approach to equation (B.1), except they replace the indicator functions in the structural break model by cumulative density functions:

$$\gamma_{n,t} = a_1 F(t/n; \gamma) + a_2 [1 - F(t/n; \gamma)],$$

so that discontinuous jumps are replaced by smooth transitions. This is also a special case of our general model.

Whereas the break point literature takes a parametric approach to the estimation of unstable parameter paths, a feature of our approach is that it is nonparametric. Thus, our model falls into a large statistics literature on nonparametric estimation of regression models with varying coefficients (see, for example, Fan and Zhang, 2008, for an overview). However, this literature generally focuses on i.i.d. models where an independent regressor is responsible for changes in the coefficients. In contrast to most of this literature, our regressor is a function of time, rather than a function of another independent variable. Our work builds on the work of Robinson (1989), further extended by Cai (2007) and Kristensen (2010b), who originally proposed to use kernel methods to estimate unstable discrete-time models with the coefficients restricted to satisfy equation (7).

The same idea of employing a sequence of models to construct a valid asymptotic distribution is also used in the large 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, n:

$$y_t = \rho_n y_{t-1} + u_t, \quad t = 1, ..., n, \quad \rho_n = 1 - c/n, \tag{B.2}$$

where the autoregressive parameter ρ_n is a function of the sample size, n. Obviously the true autoregressive coefficient in population is fixed and does not change as the sample size changes; however, the local-to-unity theory specifies a sequence of models where the parameter varies according to the sample size. This is done as a statistical device to ensure that the same asymptotics apply for both stationary and non-stationary cases. Analogously, specifying a sequence of discrete-time models by scaling the underlying parameters by sample size also allows us to construct well-defined asymptotic distributions for our estimators.

A final point is that the restriction in equation (7) should not be confused with high-frequency sampling of continuous-time models. Our (sequence of) model(s) and our analysis are firmly rooted in a discrete-time setting. In particular, our sequence of models does not have a well-defined continuous-time limit as $n \to \infty$. Equation (7) a statistical tool that allows us to bypass the issue of identifying n parameters from n observations.

There is, however, some connection between our discrete-time estimators and the literature on estimating realized volatility in a continuous-time setting (see the summary by Andersen et al., 2003). These studies have concentrated on estimating variances, but recently Andersen et al. (2006) estimate realized quarterly-frequency betas of 25 Dow Jones stocks from daily data. Andersen et al.'s estimator is similar to Lewellen and Nagel (2006) and uses only a backward-looking filter with constant weights. Within our framework, Andersen et al. (2006) and Lewellen and Nagel (2006) are estimating integrated or averaged betas, $\bar{\beta}_{\Delta,t} = \int_{t-\Delta}^{t} \beta_s ds$, where $\Delta > 0$ is the window over which they compute their OLS estimators, say over a month. Integrated betas implicitly ignore the variation of beta within each window as they are the average beta across the time period of the window. Note that by choosing a flat kernel and the bandwidth, $h > 0$, to match the given time window, $\Delta > 0$, our proposed estimators are numerically identical to the realized beta estimators. But, while the realized betas develop an asymptotic theory for a fixed window width, our theoretical results establish results where the time window shrinks with sample size. This allows us to recover the spot alphas and betas. Moreover, we develop bandwidth (time window) selection rules and so optimally select the window width.

Given that realized betas in finite samples are a special case of our more general estimator, one would expect that our methodology can be adapted to a continuous-time setting. In Appendix D, we demonstrate that this is the case and adapt our estimators to a continuous-time setting which allows for estimation of instantaneous alphas and betas. The nonparametric identification and estimation of the alphas in the continuous-time setting relies on the introduction of a sequence of continuous-time models similar to the sequence of models used to analyze our estimators in a discrete-time setting. Without this stastical tool we cannot identify the alphas in continuous time as is well known from Merton (1980).

C Proofs

Proof of Theorem 1. See Kristensen (2010b). ■

Proof of Theorem 2. We write $K_{st} = K_h(s/n - t/n)$. Define $\hat{\Lambda}_t = n^{-1}\sum_{s=1}^n K_{st}X_sX_s'$ and $\hat{m}_t = n^{-1}\sum_{s=1}^n K_{st}X_sR_s'$ such that

$$\hat{\gamma}_t - \gamma_t = \hat{\Lambda}_t^{-1}\hat{m}_t - \Lambda_t^{-1}m_t,$$

where $m_t = \mathrm{E}[X_tR_t'] = \Lambda_t\gamma_t$. By a second-order Taylor expansion of the right hand side,

$$\hat{\gamma}_t - \gamma_t = \Lambda_t^{-1}[\hat{m}_t - m_t] - \Lambda_t^{-1}\left[\hat{\Lambda}_t - \Lambda_t\right]\gamma_t + O\left(||\hat{m}_t - m_t||^2\right) + O\left(||\hat{\Lambda}_t - \Lambda_t||^2\right). \tag{C.1}$$

By Kristensen (2009, Theorem 1), we obtain that uniformly over $1 \leq t \leq n$,

$$\begin{aligned}\hat{m}_t &= m_t + O_P(h^r) + O_P\left(\sqrt{\log(n)/(nh)}\right)\\ \hat{\Lambda}_t &= \Lambda_t + O_P(h^r) + O_P\left(\sqrt{\log(n)/(nh)}\right),\end{aligned} \tag{C.2}$$

such that the two remainder terms in equation (C.1) are both $o_P(1/\sqrt{n})$ given the conditions imposed on h in (A.6).

Define $W_t = (z_t, X_t, u_t)$ with $u_t = t/n \in [0,1]$; we can treat u_t as i.i.d. uniformly distributed random variables that are independent of (z_t, X_t). We introduce the function $a(W_s, W_t)$ given by

$$\begin{aligned}a(W_s,W_t) &\equiv \Lambda_s^{-1}[K_{st}X_tR_t' - m_s] - \Lambda_s^{-1}[K_{st}X_tX_t' - \Lambda_s]\gamma_s\\ &= K_{st}\Lambda_s^{-1}X_t[R_t' - X_t'\gamma_s]\\ &= K_{st}\Lambda_s^{-1}X_t[X_t'\gamma_t - X_t'\gamma_s + \varepsilon_t]\\ &= K_{st}\Lambda_s^{-1}X_t\varepsilon_t' + K_{st}\Lambda_s^{-1}X_tX_t'[\gamma_t - \gamma_s],\end{aligned} \tag{C.3}$$

where $\varepsilon_t \equiv \Omega_t^{1/2}z_t$. Using equations (C.1) and (C.2), we can write:

$$\begin{aligned}\hat{\gamma}_{\mathrm{LR}} - \gamma_{\mathrm{LR}} &= \frac{1}{n}\sum_{s=1}^n\left\{\Lambda_s^{-1}[\hat{m}_s - m_s] - \Lambda_s^{-1}\left[\hat{\Lambda}_s - \Lambda_s\right]\gamma_s\right\} + o_P(1/\sqrt{n})\\ &= \frac{1}{n^2}\sum_{s=1}^n\sum_{t=1}^n\left\{\Lambda_s^{-1}[K_{st}X_tR_t' - m_s] - \Lambda_s^{-1}[K_{st}X_tX_t' - \Lambda_s]\gamma_s\right\} + o_P(1/\sqrt{n})\\ &= \frac{1}{n^2}\sum_{s=1}^n\sum_{t=1}^n a(W_s,W_t) + o_P(1/\sqrt{n}).\end{aligned} \tag{C.4}$$

Defining

$$\phi(W_s,W_t) = a(W_s,W_t) + a(W_t,W_s), \tag{C.5}$$

we may write

$$\frac{1}{n^2}\sum_{s=1}^n\sum_{t=1}^n a(W_s,W_t) = \frac{n-1}{n}U_n + \frac{1}{n^2}\sum_{s=1}^n\phi(W_s,W_s), \tag{C.6}$$

where $U_n := \sum_{s<t}\phi(W_s,W_t)/[n(n-1)]$. Here, $\sum_{s=1}^n\phi(W_s,W_s)/n^2 = O_P(1/n)$, while, by the Hoeffding decomposition, $U_n = 2\sum_{s=1}^n\bar{\phi}(W_s)/n + r_n$, where r_n is the remainder term, and $\bar{\phi}(w)$ is the projection function. This is given by

$$\bar{\phi}(w) = \mathrm{E}[\phi(w,W_t)] = \mathrm{E}[a(w,W_t)] + \mathrm{E}[a(W_s,w)], \tag{C.7}$$

where, with $w = (z, x, \tau)$,

$$\begin{aligned} \mathrm{E}\left[a\left(w, W_t\right)\right] &= \int_0^1 K_h\left(\tau - t\right) \Lambda^{-1}\left(t\right) dt \times xz' \Omega^{1/2}\left(\tau\right) \\ &\quad + \int_0^1 K_h\left(\tau - t\right) \Lambda^{-1}\left(t\right) xx' \left[\gamma\left(s\right) - \gamma\left(\tau\right)\right] dt \\ &= \Lambda^{-1}\left(\tau\right) xz' \Omega^{1/2}\left(\tau\right) + \Lambda^{-1}\left(\tau\right) xx' \gamma^{(r)}\left(\tau\right) \times h^r + o\left(h^r\right) \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}\left[a\left(W_s, w\right)\right] &= \int_0^1 K_h\left(s - \tau\right) \left[\gamma\left(s\right) - \gamma\left(\tau\right)\right] ds \\ &= \gamma^{(r)}\left(\tau\right) \times h^r + o\left(h^r\right), \end{aligned}$$

with $\gamma^{(r)}\left(\tau\right)$ denoting the rth order derivative of $\gamma\left(\tau\right)$. In total,

$$\bar{\phi}\left(W_s\right) = \Lambda_s^{-1} \times X_s z_s' \Omega_s^{1/2} + \Lambda_s^{-1} X_s X_s' \gamma_s^{(r)} \times h^r + \gamma_s^{(r)} \times h^r + o\left(h^r\right). \tag{C.8}$$

By Denker and Keller (1983, Proposition 2), the remainder term of the decomposition satisfies $r_n = O_P\left(n^{-1+\epsilon/2} M_\delta\right)$ for any $\epsilon > 0$, where $M_\delta \equiv \sup_{s,t} E\left[\left|\phi\left(W_s, W_t\right)\right|^{2+\delta}\right]^{1/(2+\delta)}$. The moment M_δ is bounded as follows:

$$\begin{aligned} M_\delta^{2+\delta} &\equiv \sup_{s,t} \mathrm{E}\left[\left|\phi\left(Z_s, Z_t\right)\right|^{2+\delta}\right] \\ &\leq 2 \sup_{s,t} \mathrm{E}\left[\left|a\left(Z_s, Z_t\right)\right|^{2+\delta}\right] \\ &\leq 2 \int_0^1 \int_0^1 \left|K_h\left(s - t\right)\right|^{2+\delta} \left\|\Lambda^{-1}\left(s\right)\right\|^{2+\delta} \mathrm{E}\left[\left\|X_{[sn]}\right\|^{2+\delta}\right] \mathrm{E}\left[\left\|\varepsilon_{[tn]}\right\|^{2+\delta}\right] dsdt \\ &\quad + 2 \int_0^1 \int_0^1 \left|K_h\left(s - t\right)\right|^{2+\delta} \left\|\Lambda^{-1}\left(s\right)\right\|^{2+\delta} \mathrm{E}\left[\left\|X_{[tn]}\right\|^{2+\delta}\right] \left|\gamma\left(t\right) - \gamma\left(s\right)\right|^{2+\delta} dsdt \\ &\leq \frac{C}{h^{1+\delta}} \int_0^1 \left\|\Lambda^{-1}\left(t\right)\right\|^{2+\delta} \mathrm{E}\left[\left\|X_{[tn]}\right\|^{2+\delta}\right] \mathrm{E}\left[\left\|\varepsilon_{[tn]}\right\|^{2+\delta}\right] dt + O\left(1\right) \\ &= O\left(h^{-(1+\delta)}\right). \end{aligned}$$

Choosing $\delta = 6$, we obtain $\sqrt{n}\Delta_n = O_P\left(n^{(-1+\epsilon)/2} h^{-7/8}\right)$. In total,

$$\begin{aligned} \sqrt{n}(\hat{\gamma}_{\mathrm{LR}} - \gamma_{\mathrm{LR}}) &= \frac{1}{\sqrt{n}} \sum_{s=1}^n \Lambda_s^{-1} X_s \varepsilon_s + O_P\left(\sqrt{n} h^r\right) \\ &\quad + O_P\left(\log\left(n\right) / \left(\sqrt{n} h\right)\right) + O_P\left(1 / \left(n^{(1-\epsilon)/2} h^{7/8}\right)\right), \end{aligned} \tag{C.9}$$

where, applying standard CLT results for heterogenous mixing sequences, see for example Wooldridge and White (1988),

$$\sum_{s=1}^n \Lambda_s^{-1} X_s \varepsilon_s / \sqrt{n} \xrightarrow{d} N\left(0, V_{\mathrm{LR}}\right). \tag{C.10}$$

■

Proof of Theorem 3. See Kristensen (2010b). ■

D Continuous-Time Factor Models

Consider the following stochastic differential equation for $s_t = \log\left(S_t\right)$ containing M log-prices,

$$ds_t = \alpha_t dt + \beta_t' dF_t + \Sigma_t^{1/2} dW_t, \tag{D.1}$$

where F_t are J factors and W_t is a M-dimensional Brownian motion. Suppose we have observed s_t and F_t over the time span $[0, T]$ at n discrete time points, $0 \leq t_0 < t_1 < ... < t_n \leq T$. For simplicity, these are assumed equi-distant such that $\Delta = t_i - t_{i-1}$ is constant. This is the ANOVA model considered in Andersen et al. (2006) and Mykland and Zhang (2006), where estimators of the integrated factor loadings, $\int_0^T \beta_s ds$, are developed. We wish to obtain estimates of the spot alphas and betas which can be interpreted as the realized instantaneous drift and (co-)volatility term of (s_t, F_t). As is well known, one can nonparametrically estimate the spot volatility in diffusion models without any further restrictions, while the drift is unidentified (see, for example, Merton, 1980; Bandi and Phillips, 2003; Kristensen, 2010a). In this appendix, we show how to estimate the alpha and beta processes by adapting our discrete-time estimators to the continuous-time setting.

To solve the issue of non-identification of the instantaneous alpha, we proceed as in the discrete-time setting and impose the following additional assumption on the alpha and beta processes:

$$\alpha_t = \alpha\left(t/T\right) \quad \text{and} \quad \beta_t = \beta\left(t/T\right), \tag{D.2}$$

for some functions $\alpha\left(\cdot\right)$ and $\beta\left(\cdot\right)$. It should be noted that while we normalize by number of observations, n, in the discrete-time setting, we now normalize by the time span T. If we set $\Delta = 1$, then these two normalizations are equivalent.

The motivation behind assumption (D.2) is the same as in the discrete-time setting: by introducing a sequence of models indexed by the time span T, then as $T \to \infty$, data in the neighborhood of a given normalized point in time, say τ, carries information about $\alpha\left(\tau\right)$ and $\beta\left(\tau\right)$. This has no economic interpretation beyond maintaining a factor model structure throughout the sequence of models and is a convenient statistical tool that allows us to develop asymptotic inference for our estimators of the spot alphas (see Appendix B for similar statistical approaches). Without introducing this sequence of models, it would not possible to identify the $\alpha(\cdot)$ process nonparametrically. Note that, in contrast to the discrete-time setting, we actually only need to impose $\alpha_t = \alpha\left(t/T\right)$ to ensure identification. In contrast, $\beta_t = \beta(\cdot)$ is identified without any further restrictions, and so we could remove the restriction that $\beta_t = \beta\left(t/T\right)$. The long-run alphas and long-run betas are also identified without imposing $\alpha_t = \alpha\left(t/T\right)$. However, in this appendix we choose to maintain the assumption (D.2) to make clear the connection between the discrete- and continuous-time setting.

To facilitate the development and analysis of the alphas and betas in the diffusion setting, we introduce a discretized version of the continuous-time model,

$$\Delta s_{t_i} = \alpha_{t_i}\Delta + \beta_{t_i}'\Delta F_{t_i} + \Sigma_{t_i}^{1/2}\sqrt{\Delta}z_{t_i}, \tag{D.3}$$

where

$$\Delta s_{t_i} = s_{t_i} - s_{t_{i-1}}, \quad \Delta F_{t_i} := F_{t_i} - F_{t_{i-1}},$$

and $z_{t_i} \sim$ i.i.d. $(0, I_M)$. In the following we treat (D.3) as the true, data-generating model. The extension to treat (D.1) as the true model requires some extra effort to ensure that the discrete-time model (D.3) is an asymptotically valid approximation of (D.1). The analysis would involve additional error terms due to the approximations $\mu_{t_i}\Delta \approx \int_{t_{i-1}}^{t_i} \mu_s ds$ and $\Sigma_{t_i}^{1/2}\sqrt{\Delta}z_{t_i} \approx \int_{t_{i-1}}^{t_i} \Sigma_s^{1/2} dW_s$, that would need to vanish sufficiently fast as $\Delta \to 0$ to obtain asymptotic results. This could be done along the lines of, for example, Bandi and Phillips (2003); by applying their arguments, we expect that the following results can be extended to the actual continuous-time model instead of its discretized version.

Defining

$$R_i := \Delta s_{t_i}/\Delta, \quad f_i := \Delta F_{t_i}/\Delta, \quad \Omega_i := \Sigma_{t_i}/\Delta, \tag{D.4}$$

we can rewrite the discretized diffusion model as in equation (2). Given that the discretized model can be written in the form of equation (2), we propose the following adapted version of our estimators:

$$\left(\hat{\alpha}\left(\tau\right), \hat{\beta}\left(\tau\right)'\right)' = \left[\sum_{i=1}^{n} K_h\left(t_i/T - \tau\right) X_i X_i'\right]^{-1} \left[\sum_{t=1}^{n} K_h\left(t_i/T - \tau\right) X_i R_i'\right], \tag{D.5}$$

for any normalized point in time $\tau \in [0, 1]$, where $X_i = (1, f_i)$, f_i, and R_i are defined in equation (D.4). The only difference between the discrete-time and continuous-time versions of the filtered alphas and betas is that we normalize the time points by the time span, T, instead of the number of observations, n.

While the proposed estimators of the alphas and betas in a continuous-time setting look very similar to the ones developed in the discrete-time setting of the main text, the analysis of the continuous-time estimators are different. First, we rely on combined infill ($\Delta \to 0$) and long span ($T \to \infty$) asymptotics in the analysis, while in the discrete time model we only used long-span arguments. The reason is that the continuous-time setting must

have a well-defined limit, while the sequence of models in the discrete-time setting are always specified at discrete time points.

Second, the asymptotic behaviour of the variance component is different due to the infill asymptotics. As $n = T/\Delta \to \infty$ in (the discretized version of) the continuous-time model, $\Omega_i := \Sigma_{t_i}/\Delta \to 0$ as we sample more frequently. In contrast, in the discrete-time model (2), the variance is constant asymptotically as we only rely on long-span asymptotics. As such, the asymptotic analysis and properties of the estimators proposed for the difffusion model are closely related to the kernel-regression type estimators proposed in Bandi and Phillips (2003), Kanaya and Kristensen (2010), and Kristensen (2010a). 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 (2010a) only consider estimation of univariate stochastic volatility models. In contrast, we model time-inhomogenous, multivariate processes where the rescaled time parameter $\tau_i := t_i/T$ is used as kernel regressor. Nevertheless, the arguments employed in the analysis of our estimators are similar to the ones used in these papers.

In particular, there is a similarity between the identification assumption (D.2) and the assumption of recurrence of Bandi and Phillips (2003). In Bandi and Phillips, 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. In our case, by assuming that $\alpha_t = \alpha(t/T)$, the spot alpha becomes a function of the recurrent process $Z_t = t/T \in [0,1]$; to be more specific, Z_{t_i}, $i = 1, ..., n$, can be thought of as i.i.d. draws from the uniform distribution on $[0,1]$. Thus, with this assumption, we accomplish the same increase in local information around a given point $\tau \in [0,1]$ in each successive model as $T \to \infty$. Note that in contrast the un-normalized time, $Z_t = t$, is not a recurrent process, and so without the introduction of normalized time, we would not be able to identify α_t.

We only supply an informal analysis of the estimators for the discretized diffusion model; a complete analysis is outside the scope of this paper. In addition to equation (D.3), we assume that the factors solve a discretized diffusion model,

$$\Delta F_{t_i} = \mu_{F,t_i}\Delta + \Lambda_{FF,t_i}^{1/2}\sqrt{\Delta}u_{t_i}, \tag{D.6}$$

where $u_{t_i} \sim \text{i.i.d}(0, I_J)$, and the processes μ_t and $\Lambda_{FF,t}$ satisfy

$$\mu_{F,t} = \mu_F(t/T) + o(t/T), \quad \Lambda_{FF,t} = \Lambda_{FF}(t/T) + o(t/T), \tag{D.7}$$

where $\mu(\cdot)$ and $\Lambda_{FF}(\cdot)$ are r times differentiable (possibly random) functions. We claim that the following two results hold for our estimators in a diffusion framework under suitable regularity conditions:

Claim 4 *Under suitable regularity conditions, the spot alpha and beta estimators satisfy:*

1. *As $\Delta^{-1}h \to \infty$ and $\Delta^{-1}h^{2r+1} \to 0$,*

$$\sqrt{\Delta^{-1}h}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} N\left(0, \kappa_2\Lambda_{FF}^{-1}(\tau) \otimes \Sigma(\tau)\right). \tag{D.8}$$

2. *As $Th \to \infty$ and $Th^{2r+1} \to 0$,*

$$\sqrt{Th}(\hat{\alpha}(\tau) - \hat{\alpha}(\tau)) \xrightarrow{d} N(0, \kappa_2\Sigma(\tau)). \tag{D.9}$$

For the first result, we require that $\Delta^{-1}h \to \infty$ while in the second part we require $Th \to \infty$. This is due to the fact that $\text{var}(\hat{\beta}(\tau)) = O\left(1/\left(\Delta^{-1}h\right)\right)$ while $\text{var}(\hat{\alpha}(\tau)) = O(1/(Th))$. This is similar to the results found in Bandi and Phillips (2003); their drift estimator converges with rate $\sqrt{Th}$ while their diffusion estimator converges with rate $\sqrt{nh}$.

Informal proof of equation (D.8): First, note that $\hat{\beta}(\tau)$ can be rewritten as

$$\hat{\beta}(\tau) = \left[\sum_{i=1}^{n} K_h(t_i/T - \tau)\left(f_i - \bar{f}(\tau)\right)\left(f_i - \bar{f}(\tau)\right)'\right]^{-1}\left[\sum_{t=1}^{n} K_h(t_i/T - \tau)\left(f_i - \bar{f}(\tau)\right)\left(R_i - \bar{R}(\tau)\right)\right],$$

where $\bar{f}(\tau) = \sum_{i=1}^{n} K_h(t_i/T - \tau) f_i/n$ and $\bar{R}(\tau) = \sum_{i=1}^{n} K_h(t_i/T - \tau) R_i/n$. Since

$$f_i := \Delta F_{t_i}/\Delta = \mu_{F,t_i} + \frac{1}{\sqrt{\Delta}}\Lambda_{FF,t_i}u_{t_i},$$

we obtain that

$$\bar{f}(\tau) = \frac{1}{n}\sum_{i=1}^{n} K_h(t_i/T - \tau)\,\mu_F(t_i/T) + \frac{1}{n\sqrt{\Delta}}\sum_{i=1}^{n} K_h(t_i/T - \tau)\,\Lambda_{FF}(t_i/T)\,u_{t_i} := \bar{f}_1(\tau) + \bar{f}_2(\tau).$$

Using the fact that a Rieman integral can be approximated by its corresponding average and standard results for kernel estimators,

$$\bar{f}_1(\tau) = \int_0^1 K_h(s-\tau)\,\mu_F(s)\,ds + O(1/n) = \mu_F(s) + O\left(h^2\right) + O(1/n).$$

The second term has mean zero and variance

$$\text{var}\left(\bar{f}_2(\tau)\right) = \frac{1}{n\Delta} \times \frac{1}{n}\sum_{i=1}^{n} K_h(t_i/T-\tau)^2\,\Lambda_{FF}^2(t_i/T) = \frac{1}{n\Delta} \times \left\{\frac{1}{h}\kappa_2 \Lambda_{FF}^2(t_i/T)\right\},$$

where we employed the same arguments as in the analysis of the first term. In total,

$$\bar{f}(\tau) = \mu_F(\tau) + O_P(h^r) + O_P\left(\frac{1}{\sqrt{Th}}\right). \tag{D.10}$$

By similar arguments,

$$\bar{R}(\tau) = \alpha(\tau) + \beta(\tau)'\,\mu_F(\tau) + O_P(h^r) + O_P\left(\frac{1}{\sqrt{Th}}\right). \tag{D.11}$$

Noting that

$$f_i f_i' = \mu_{F,t_i}\mu_{F,t_i}' + \frac{1}{\Delta}\Lambda_{FF,t_i}^{1/2} u_{t_i} u_{t_i}' \Lambda_{FF,t_i}^{1/2},$$

we obtain

$$\begin{aligned}
\frac{\Delta}{n}\sum_{i=1}^{n} K_h(t_i/T-\tau)\,f_i f_i' &= \frac{\Delta}{n}\sum_{i=1}^{n} K_h(t_i/T-\tau)\,\mu_{F,t_i}\mu_{F,t_i}' + \frac{1}{n}\sum_{i=1}^{n} K_h(t_i/T-\tau)\,\Lambda_{FF,t_i}^{1/2} u_{t_i} u_{t_i}' \Lambda_{FF,t_i}^{1/2} \\
&= \Delta\mu_F(\tau)\,\mu_F(\tau)' + \Lambda_{FF}(\tau) + +O_P(h^r) + O_P\left(\frac{1}{\sqrt{nh}}\right),
\end{aligned} \tag{D.12}$$

and

$$\begin{aligned}
\frac{\Delta}{n}\sum_{t=1}^{n} K_h(t_i/T-\tau)\,f_i R_i' &= \frac{\Delta}{n}\sum_{t=1}^{n} K_h(t_i/T-\tau)\,\alpha_{t_i} + \frac{\Delta}{n}\sum_{t=1}^{n} K_h(t_i/T-\tau)\,f_{t_i} f_{t_i}'\beta_{t_i} \\
&\quad + \frac{\sqrt{\Delta}}{n}\sum_{t=1}^{n} K_h(t_i/T-\tau)\,f_{t_i} z_{t_i}'\Sigma_{t_i}^{1/2} \\
&= \Delta\alpha(\tau) + \Lambda_{FF}(\tau)\,\beta(\tau) + \sqrt{\frac{\Delta}{h}}U_n(\tau) + O_P(h^r)
\end{aligned} \tag{D.13}$$

where, by a CLT for i.i.d. heterogenous arrays (e.g. Wooldridge and White, 1988),

$$\frac{\sqrt{h\Delta}}{\sqrt{n}}\sum_{t=1}^{n} K_h(t_i/T-\tau)\,f_{t_i} z_{t_i}'\Sigma_{t_i}^{1/2} \xrightarrow{d} N\left(0, \kappa_2\Lambda_{FF}(\tau)\otimes\Sigma(\tau)\right).$$

Collecting the results of equations (D.10)-(D.13), we obtain equation (D.8).

Informal proof of equation (D.9): The alpha estimator can be rewritten as

$$\hat{\alpha}(\tau) = \bar{R}(\tau) - \hat{\beta}(\tau)\,\bar{f}(\tau),$$

where $\bar{R}(\tau)$ and $\bar{f}(\tau)$ are defined above. By the same arguments used in the proof of equation (D.8), it easily follows that

$$\sqrt{Th}(\hat{\alpha}(\tau) - \alpha(\tau)) = O_P\left(\sqrt{Th^{2r+1}}\right) + \frac{\sqrt{h}}{\sqrt{n}}\sum_{t=1}^{n} K_h(t_i/T - \tau)\,\Sigma_{t_i}^{1/2} z_{t_i}, \tag{D.14}$$

where

$$\frac{\sqrt{h}}{\sqrt{n}}\sum_{t=1}^{n} K_h(t_i/T - \tau)\,\Sigma_{t_i}^{1/2} z_{t_i} \xrightarrow{d} N(0, \kappa_2 \Sigma(\tau)).$$

E Performance of Our Estimator under a Random Walk

We analyze the behavior of our nonparametric estimator under the assumption that the true DGP of α_t is a random walk. We show informally that our estimator consistently estimates the trajectory of α_t in this setting. We argue that the analysis can be extended to allow alphas and betas following other standard time-series models such as ARMA models.

For simplicity, we impose the hypothesis that $\beta_t = 0$ such that our conditional factor model collapses to:

$$R_t = \alpha_t + \varepsilon_t. \tag{E.1}$$

The restriction $\beta_t = 0$ is merely made to keep the notation simple and is not necessary for the following arguments to go through. Under this restriction, the nonparametric estimator of α_t is given by $\hat{\alpha}_t = \hat{\alpha}(t/n)$ where

$$\hat{\alpha}(\tau) = \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right) R_t, \quad \tau \in [0,1].$$

Suppose that in fact that the alpha sequence, $\{\alpha_t : t = 1, ..., n\}$, that generates the observed returns, $\{R_t : t = 1, ..., n\}$, arrives from a random walk model:

$$\alpha_t = \alpha_{t-1} + z_t, \tag{E.2}$$

where z_t is i.i.d. $\left(0, \sigma_z^2\right)$. Also, assume that ε_t are i.i.d.$\left(0, \sigma_\varepsilon^2\right)$ and independent of z_t.

To analyze the properties of our estimator under the assumption of equation (E.2), first write

$$\begin{aligned}
\hat{\alpha}(\tau) &= \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right)\alpha_t + \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right)\varepsilon_t \\
&= \sqrt{n}\left\{\frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right)\frac{1}{\sqrt{n}}\alpha_t\right\} + \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right)\varepsilon_t.
\end{aligned} \tag{E.3}$$

The second term in equation (E.3) satisfies

$$\frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right)\varepsilon_t \xrightarrow{p} 0, \quad nh \to \infty, \tag{E.4}$$

by standard arguments for kernel estimators. To analyze the first term in equation (E.3), note that, under the random walk hypothesis, the normalized version of α_t, $\alpha_t/\sqrt{n}$, can be approximated by a Brownian motion. Formally, the following sequence of random functions

$$\bar{\alpha}_n(\tau) := \frac{1}{\sqrt{n}}\alpha_{[\tau n]},$$

satisfies

$$\bar{\alpha}_n(\tau) = \bar{\alpha}(\tau) + o_P(1), \tag{E.5}$$

where $\tau \mapsto \bar{\alpha}(\tau)$ is a Brownian motion. Our methodology is conditional on the sample path of the parameters in data. Thus, we condition on $\{z_t\}$ so we can treat $\bar{\alpha}(\tau)$ as a *deterministic continuous function*, and so obtain as $nh \to \infty$ and $h \to 0$,

$$\begin{aligned}
\frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right)\frac{1}{\sqrt{n}}\alpha_t &= \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right)\bar{\alpha}_n(t/n) \\
&= \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right)\{\bar{\alpha}(t/n) + o(1)\} \\
&= \bar{\alpha}(\tau) + o(1).
\end{aligned} \tag{E.6}$$

Plugging equations (E.4) and (E.6) into equation (E.3), we obtain the following informal expansion of the estimator $\hat{\alpha}(\tau)$:

$$\hat{\alpha}(\tau) \approx \sqrt{n}\bar{\alpha}(\tau) \approx \sqrt{n}\bar{\alpha}_n(\tau). \tag{E.7}$$

In particular,

$$\hat{\alpha}_t = \hat{\alpha}(t/n) \approx \sqrt{n}\bar{\alpha}_n(t/n) = \alpha_t.$$

This shows that our estimator is able to track the trajectory of α_t, which follows a random walk, over time.

The above argument is heuristic and is not valid at a formal level since the limit $\alpha_{[\tau n]} = \sqrt{n}\bar{\alpha}_n(\tau)$, explodes almost surely as $n \to \infty$. Thus, one cannot establish equation (E.7) formally or establish any limiting distribution theory.

Instead, to formalize the above asymptotic approximation and to obtain asymptotic results as those stated in the paper, we need to redefine the DGP in a way such that the object of interest does not explode as the sample size grows. One way of doing this is to redefine α_t in equation (E.2) by writing equivalently

$$\alpha_t = \alpha_{t-1} + \frac{1}{\sqrt{n}}\bar{z}_t, \tag{E.8}$$

where $\bar{z}_t$ is i.i.d. $\left(0, \bar{\sigma}_z^2\right)$. For a given sample size n, the two DGPs, equations (E.2) and (E.8), can be made equivalent by choosing $\bar{\sigma}_z^2 = n\sigma_z^2$. Thus, we cannot distinguish between whether data is generated by equations (E.2) or (E.8) for a given sample size. But with the random walk model in equation (E.8), we can now conduct a formal statistical analysis and establish the theoretical results in the paper. Note that with equation (E.8), we assume a sequence of DGPs which depend on the sample size, n. In particular, $\alpha_n(\tau) := \alpha_{[\tau n]} \xrightarrow{d} \alpha(\tau)$ where $\alpha(\tau)$ is a Brownian motion.

The above argument can be extended to any time-series model for α_t that can be approximated by a continuous-time process. For example, (the normalized version of) an AR(1) process can be approximated by a Ornstein-Uhlenbeck process (see, e.g. Cumberland and Sykes, 1982) and our estimator also works for such DGPs.

F Gibbons, Ross and, Shanken (1989) as a Special Case

First, we derive the asymptotic distribution of the Gibbons, Ross, and Shanken (1989) [GRS] estimators within our setting. The GRS estimator, which we denote $\tilde{\gamma}_{\mathrm{LR}} = (\tilde{\alpha}_{\mathrm{LR}}, \tilde{\beta}_{\mathrm{LR}})$, is a standard least squares estimator of the form

$$\begin{aligned}
\tilde{\gamma}_{\mathrm{LR}} &= \left[\sum_{t=1}^{n} X_t X_t'\right]^{-1}\left[\sum_{t=1}^{n} X_t R_t'\right] \\
&= \left[\sum_{t=1}^{n} X_t X_t'\right]^{-1}\sum_{t=1}^{n} X_t X_t'\gamma_t + \left[\sum_{t=0}^{n} X_t X_t'\right]^{-1}\sum_{t=1}^{n} X_t \varepsilon_t' \\
&= \bar{\gamma}_{\mathrm{LR}} + U_n,
\end{aligned} \tag{F.1}$$

where, under assumptions (A.1)-(A.5),

$$\bar{\gamma}_{\mathrm{LR}} = \left(\int_0^1 \Lambda(s)\, ds\right)^{-1}\int_0^1 \Lambda(s)\,\gamma(s)\, ds,$$

and

$$\sqrt{n}U_n \xrightarrow{d} N\left(0, \left(\int_0^1 \Lambda(s)\, ds\right)^{-1}\left(\int_0^1 \Lambda(s) \otimes \Omega(s)\, ds\right)\left(\int_0^1 \Lambda(s)\, ds\right)^{-1}\right).$$

To separately investigate the performance of $\tilde{\alpha}_{\mathrm{LR}}$, we write

$$\Lambda_t = \begin{bmatrix} 1 & \mu_{f,t} \\ \mu_{f,t} & \Lambda_{ff,t} \end{bmatrix}, \tag{F.2}$$

where $\mu_{f,t} = \mathrm{E}[f_t|\gamma, \Omega] = \mu_f(t/n) + o(1)$ and $\Lambda_{ff,t} = \mathrm{E}[f_t f_t'|\gamma, \Omega] = \Lambda_{ff}(t/n) + o(1)$. We also define $V_{f,t} \equiv \mathrm{var}(f_t|\gamma, \Omega)$ which satisfies $V_{f,t} = V_f(t/n)$, where $V_f(\tau) = \Lambda_{ff}(\tau) - \mu(\tau)\mu(\tau)'$. Thus, $\bar{\gamma}_{\mathrm{LR}} = \left(\bar{\alpha}_{\mathrm{LR}}, \bar{\beta}_{\mathrm{LR}}\right)'$

can be written as

$$\begin{aligned}\bar{\beta}_{\mathrm{LR}} &= \left[\int_0^1 V_f(\tau)\,ds\right]^{-1}\int_0^1 \beta(s)\,V_f(s)\,ds,\\ \bar{\alpha}_{\mathrm{LR}} &= \int_0^1 \alpha(s)\,ds+\int_0^1\left[\beta(s)-\bar{\beta}_{\mathrm{LR}}\right]'\mu(s)\,ds.\end{aligned} \tag{F.3}$$

From the above expressions, we see that the GRS estimator $\tilde{\alpha}_{\mathrm{LR}}$ of the long-run alphas in general is inconsistent since it is centered around $\bar{\alpha}_{\mathrm{LR}} \neq \int_0^1 \alpha(s)\,ds$. It is only consistent if the factors are uncorrelated with the loadings such that $\mu_t = \mu$ and $\Lambda_{ff,t} = \Lambda_{ff}$ are constant, in which case $\bar{\beta}_{\mathrm{LR}} = \beta_{\mathrm{LR}}$ and $\bar{\alpha}_{\mathrm{LR}} = \alpha_{\mathrm{LR}}$.

Finally, we note that in the case of constant alphas and betas and homoskedastic errors, $\Omega_s = \Omega$, the variance of our proposed estimator of γ_{LR} is identical to the one of the GRS estimator.

G Two-Sided versus One-Sided Filters

Kristensen (2010b) shows that with a two-sided symmetric kernel where $\mu_1 = \int K(z)\,zdz = 0$ and $\mu_2 = \int K(z)\,z^2dz < \infty$, the finite-sample variance is

$$\mathrm{var}\left(\hat{\gamma}_i(\tau)\right) = \frac{1}{nh}v_i(\tau) + o\left(1/(nh)\right) \quad \text{with} \quad v_t(\tau) = \kappa_2\Lambda^{-1}(\tau)\,\Omega_{ii}(\tau), \tag{G.1}$$

while the bias is given by

$$\mathrm{Bias}\left(\hat{\gamma}_i(\tau)\right) = h^2 b_i^{\mathrm{sym}}(\tau) + o\left(h^2\right) \quad \text{with} \quad b_i^{\mathrm{sym}}(\tau) = \frac{\mu_2}{2}\gamma_i^{(2)}(\tau), \tag{G.2}$$

where we have assumed that $\gamma_i(\tau)$ is twice differentiable with second order derivative $\gamma_i^{(2)}(\tau)$. 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 since $\mu_1 \neq 0$ the bias now takes the form

$$\mathrm{Bias}\left(\hat{\gamma}_i(\tau)\right) = hb_i^{\mathrm{one}}(\tau) + o(h) \quad \text{with} \quad b_i^{\mathrm{one}}(\tau) = \mu_1\gamma_i^{(1)}(\tau). \tag{G.3}$$

The bias is in this case of order $O(h)$ and is therefore larger compared to when a two-sided kernel is employed.

As a consequence, for the symmetric kernel the optimal global bandwidth minimizing the integrated MSE,

$$IMSE = \int_0^1 \mathrm{E}[\|\hat{\gamma}_{j,\tau} - \gamma_{j,\tau}\|^2]d\tau,$$

is given by:

$$h_j^* = \left(\frac{V_i}{B_i^2}\right)^{1/5} n^{-1/5}, \tag{G.4}$$

where $B_i^2 = \lim_{n\to\infty} n^{-1}\sum_{t=1}^n (b_{i,t}^{\mathrm{sym}})^2$ and $V_i = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n v_{i,t}$ 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 on the other hand a one-sided kernel is used, the optimal bandwidth is

$$h_i^* = \left(\frac{V_i}{\tilde{B}_i^2}\right)^{1/3} n^{-1/3}, \tag{G.5}$$

where $\tilde{B}_i^2 = \lim_{n\to\infty} n^{-1}\sum_{t=1}^n (b_{i,t}^{\mathrm{one}})^2$, with the corresponding $\sqrt{IMSE}$ being of order $O\left(n^{-1/3}\right)$. Thus, the symmetric kernel estimator's root-MSE is generally smaller and substantially smaller if n is large.[17]

[17] The two exceptions 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.

H Bandwidth Choice for Long-Run Estimators

We sketch the outline of the derivation of an optimal bandwidth for estimating the integrated or long-run gammas.[18] We follow the same strategy as in Cattaneo, Crump, and Jansson (2010) and Ichimura and Linton (2005), among others. With $\hat{\gamma}_{\mathrm{LR},k}$ 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 $m_t = \Lambda_t \gamma_t$ and Λ_t,

$$\hat{\gamma}_{\mathrm{LR},k} - \gamma_{\mathrm{LR},k} = U_{1,n} + U_{2,n} + R_n, \tag{H.1}$$

where $R_n = O\left(\sup_{1\leq t\leq n} ||\hat{m}_{k,t} - m_{k,t}||^3\right) + O\left(\sup_{1\leq t\leq n} ||\hat{\Lambda}_t - \Lambda_t||^3\right)$ and

$$U_{1,n} = \frac{1}{n}\sum_{t=1}^{n}\left\{\Lambda_t^{-1}\left[\hat{m}_{k,t} - m_{k,t}\right] - \Lambda_t^{-1}\left[\hat{\Lambda}_t - \Lambda_t\right]\gamma_{k,t}\right\};$$

$$U_{2,n} = \frac{1}{n}\sum_{t=1}^{n}\left\{\Lambda_t^{-1}\left[\hat{\Lambda}_t - \Lambda_t\right]\Lambda_t^{-1}\left[\hat{\Lambda}_t - \Lambda_t\right]\gamma_{k,t} - \Lambda_t^{-1}\left[\hat{\Lambda}_t - \Lambda_t\right]\Lambda_t^{-1}\left[\hat{m}_{k,t} - m_{k,t}\right]\right\}.$$

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 a MSE expansion of the estimator as a function of the bandwidth h.

We obtain straightforwardly that with $\bar{\phi}(W_t)$ given in the proof of Theorem 2,

$$\mathrm{E}\left[U_{1,n}\right] = \mathrm{E}\left[\bar{\phi}(W_t)\right] \simeq \mu_r \left\{\frac{1}{n}\sum_{t=1}^{n}\gamma_{k,t}^{(r)}\right\} \times h^r \simeq B_{\mathrm{LR},k}^{(1)} \times h^r, \tag{H.2}$$

where $r \geq 1$ is the number of derivatives (or the degree of smoothness of the alphas and betas), and

$$B_{\mathrm{LR},k}^{(1)} = \mu_r \int_0^1 \gamma_k^{(r)}(\tau)\, d\tau, \quad \mu_r := \int K(z)\, z^r dz.$$

Next, assuming that

$$\Psi_{s,t} = E\left[\left(X_s X_s' - \Lambda_t\right)\Lambda_t^{-1}\left(X_s X_s' - \Lambda_t\right)\right] = \Psi\left((s-t)/n\right), \tag{H.3}$$

for some function $\Psi(\tau)$, it holds that

$$\mathrm{E}\left[U_{2,n}\right] = \frac{2}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n} K_{t,s}^2 \Lambda_t^{-1}\Psi_{s,t}\gamma_{k,t} \simeq \frac{2\kappa_2}{nh}\sum_{s=1}^{n}\Lambda_s^{-1}\Psi_{s,s}\gamma_{k,s} = \frac{1}{nh} \times B_{\mathrm{LR},k}^{(2)}, \tag{H.4}$$

where

$$B_{\mathrm{LR},k}^{(2)} := 2\kappa_2 \int_0^1 \Lambda^{-1}(\tau)\,\Psi(\tau)\,\gamma_k(\tau)\, d\tau.$$

To compute the variance of $U_{1,n}$, observe that

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

where ϕ and a are defined in the proof of Theorem 2. Here,

$$\begin{aligned}\mathrm{E}\left[a(w, W_t)\,a(w, W_t)'\right] &= \frac{1}{n}\sum_{t=1}^{n} K_{st}^2 \Lambda_s^{-1}\Omega_{kk,t}\Lambda_t\Lambda_s^{-1} + \frac{1}{n}\sum_{t=1}^{n} K_{st}^2 \Lambda_s^{-1}\Lambda_t\left[\gamma_{k,t} - \gamma_{k,s}\right]\left[\gamma_{k,t} - \gamma_{k,s}\right]'\Lambda_t\Lambda_s^{-1} \\ &\simeq \frac{1}{h} \times q_1(w),\end{aligned}$$

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

where $q_1(w) = \kappa_2 \Lambda_s^{-1} \Omega_{kk,s}$. Similarly,

$$\begin{aligned}
\mathrm{E}\left[a(W_t,w)\,a(W_t,w)'\right] &= \frac{1}{n}\sum_{t=1}^{n} K_{st}^2 \Lambda_t^{-1} xx' e_k^2 \Lambda_t^{-1} + \frac{1}{n}\sum_{t=1}^{n} K_{st}^2 \Lambda_t^{-1} xx' \left[\gamma_{k,t} - \gamma_{k,s}\right]\left[\gamma_{k,t} - \gamma_{k,s}\right]' xx' \Lambda_t^{-1} \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} K_{st}^2 \Lambda_t^{-1} xe \left[\gamma_{k,t} - \gamma_{k,s}\right]' xx' \Lambda_t^{-1} \\
&\simeq \frac{1}{h} \times q_2(w),
\end{aligned}$$

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

$$\Sigma_{\mathrm{LR},kk} = \mathrm{E}\left[q_1(W_t)\right] + \mathrm{E}\left[q_2(W_t)\right] = 2\kappa_2 \sum_{s=1}^{n} \Lambda_s^{-1} \Omega_{kk,s} = 2\kappa_2 \int_0^1 \Lambda^{-1}(\tau)\,\Omega_{kk}(\tau)\,d\tau,$$

while $\mathrm{var}[U_{2,n}]$ is of higher order and so can be ignored. In total,

$$MSE(\hat{\gamma}_{\mathrm{LR},k}) \simeq \left\| B_{\mathrm{LR},k}^{(1)} \times h^r + B_{\mathrm{LR},k}^{(2)} \times \frac{1}{nh} \right\|^2 + \frac{1}{n} \times V_{\mathrm{LR},kk} + \frac{1}{n^2 h} \times \Sigma_{\mathrm{LR},kk}. \tag{H.5}$$

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, and so the optimal bandwidth minimizing the squared component is

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

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

References

[1] Aït-Sahalia, Y., 1996. Nonparametric Pricing of Interest Rate Derivative Securities. Econometrica 64, 527-560.

[2] Aït-Sahalia, Y., Lo, A., 1998. Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices. Journal of Finance 53, 499-547.

[3] Aït-Sahalia, Y., Brandt, M., 2007. Nonparametric Consumption and Portfolio Choice with Option-Implied State Prices. Working paper, Duke University.

[4] Andersen, T., Bollerselv, T., Diebold, F., Labys, P., 2003. Modeling and Forecasting Realized Volatility. Econometrica 71, 529-626.

[5] 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.

[6] Andrews, D., 1993. Tests for Parameter Instability and Structural Change with Unknown Change Point. Econometrica 61, 821–856.

[7] Ang, A., Chen, J., 2007. CAPM Over the Long Run: 1926-2001. Journal of Empirical Finance 14, 1-40.

[8] Ang, A., Chen, J., Xing, Y., 2006. Downside Risk. Review of Financial Studies 19, 1191-1239.

[9] Bai, J., Perron, P., 1998. Testing for and Estimation of Multiple Structural Changes. Econometrica 66, 47-79.

[10] Bandi, F., 2002. Short-term Interest Rate Dynamics: A Spatial Approach. Journal of Financial Economics 65, 73-110.

[11] Bandi, F., Phillips, P., 2003. Fully Nonparametric Estimation of Scalar Diffusion Models. Econometrica 71, 241-283.

[12] Bansal, R., Viswanathan, S., 1993. No Arbitrage and Arbitrage Pricing: A New Approach. Journal of Finance 48, 1231-1262.

[13] 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. Working paper, Oxford University.

[14] Bekaert, G., Harvey, R., Lumsdaine, R., 2002. Dating the Integration of World Equity Markets. Journal of Financial Economics 65, 203–247.

[15] Bekaert, G., Wu, G., 2000. Asymmetric Volatility and Risk in Equity Markets. Review of Financial Studies 13, 1-42.

[16] Boguth, O., Carlson, M., Fisher, A., Simutin, M., 2010. Conditional Risk and Performance Evaluation: Volatility Timing, Overconditioning, and New Estimates of Momentum Alphas. Forthcoming Journal of Financial Economics.

[17] Brandt, M., 1999. Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach. Journal of Finance 54, 1609-1645.

[18] Cai, Z., 2007. Trending Time-varying Coefficient Time Series Models with Serially Correlated Errors. Journal of Econometrics 136, 163–188.

[19] Campbell, J., Cochrane, J., 1999. By Force of Habit. Journal of Political Economy 107, 105-251.

[20] Campbell, J., Vuolteenaho, T., 2004. Bad Beta, Good Beta. American Economic Review 94, 1249-1275.

[21] Cattaneo, M., Crump, R., Jansson, M., 2010. Generalized Jackknife Estimators of Weighted Average Derivatives. Working paper, University of Michigan.

[22] Chan, N., Wei, C., 1987. Asymptotic Inference for Nearly Nonstationary AR(1) Processes. Annals of Statistics 15, 1050-1063.

[23] 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.

[24] Cochrane, J., 2001. Asset Pricing. Princeton University Press, Princeton, NJ.

[25] Cumberland, W. G., Sykes, Z. M., 1982. Weak Convergence of an Autogressive Process Used in Modeling Population Growth. Journal of Applied Probability 19, 450-455.

[26] Denker, M., Keller, G., 1983. On U-Statistics and von Mises' Statistics for Weakly Dependent Processes. Zeitschrift für Wahrkeinlichkeitstheorie 64, 505-522.

[27] Diggle, P., Hutchinson, M., 1989. On Spline Smoothing with Correlated Errors. Australian Journal of Statistics 31, 166-82.

[28] Dimson, E., 1979. Risk Measurement When Shares Are Subject to Infrequent Trading. Journal of Financial Economics 7, 197-226.

[29] Fama, E., French, K., 1993. Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics 33, 3-56.

[30] Fama, E., French, K., 1997. Industry Costs of Equity. Journal of Financial Economics 43, 153-193.

[31] Fama, E., MacBeth, J., 1973. Risk, Return, and Equilibrium: Empirical Tests. Journal of Political Economy 81, 607-636.

[32] Fan, J., Huang, T., 2005. Profile Likelihood Inferences on Semiparametric Varying-Coefficient Partially Linear Models. Bernoulli 11, 1031-1057.

[33] Fan, J., Zhang, W., 2008. Statistical Methods with Varying Coefficient Models. Statistics and Its Interface 1, 179–195.

[34] Fan, J., Zhang, C., Zhang, W., 2001. Generalized Likelihood Ratio Statistics and Wilks Phenomenon. Annals of Statistics 29, 153-193.

[35] Ferson, W., Harvey, C., 1991. The Variation of Economic Risk Premiums. Journal of Political Economy 99, 385-415.

[36] Ferson, W., Schadt, R., 1996. Measuring Fund Strategy and Performance in Changing Economic Conditions. Journal of Finance 51, 425-462.

[37] Ferson, W., Qian, M., 2004. Conditional Evaluation Performance: Revisited. Research Foundation of the CFA Institute.

[38] Foster, D., Nelson, D., 1996. Continuous Record Asymptotics for Rolling Sample Variance Estimators. Econometrica 64, 139-174.

[39] French, K., Schwert, G., Stambaugh, R., 1987. Expected Stock Returns and Volatility. Journal of Financial Economics 19, 3-29.

[40] 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.

[41] Gibbons, M., Ross, S., Shanken, J., 1989. A Test of the Efficiency of a Given Portfolio. Econometrica 57, 1121-1152.

[42] Gomes, J., Kogan, L., Zhang, L., 2003. Equilibrium Cross Section of Returns. Journal of Political Economy 111, 693-732.

[43] 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.

[44] Hart, J., 1991. Kernel Regression Estimation with Time Series Errors. Journal of the Royal Statistical Society, Series B (Methodological) 53, 173-187.

[45] Harvey, C., 2001. The Specification of Conditional Expectations. Journal of Empirical Finance 8, 573-638.

[46] Hayashi, T., Yoshida, N., 2005. On Covariance Estimation of Non-Synchronously Observed Diffusion Processes. Bernoulli 11, 359-379.

[47] 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.

[48] Jagannathan, R., Wang, Z., 1996. The Conditional CAPM and the Cross Section of Expected Returns. Journal of Finance 51, 3-53.

[49] Jostova, G., Philipov, A., 2005. Bayesian Analysis of Stochastic Betas. Journal of Financial and Quantitative Analysis 40, 747-778.

[50] Kanaya, S., Kristensen, D., 2010. Estimation of Stochastic Volatility Models by Nonparametric Filtering. Working paper, Columbia University.

[51] Kristensen, D., 2009. Uniform Convergence Rates of Kernel Estimators with Heterogeneous, Dependent Data. Econometric Theory 25, 1433-1445.

[52] Kristensen, D., 2010a. Nonparametric Filtering of the Realised Spot Volatility: A Kernel-based Approach. Econometric Theory 26, 60-93.

[53] Kristensen, D., 20010b. Nonparametric Detection and Estimation of Structural Breaks. Working paper, Columbia University.

[54] Lettau, M., Ludvigson, S., 2001a. Consumption, Aggregate Wealth, and Expected Stock Returns. Journal of Finance 56, 815-849.

[55] 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.

[56] Lettau, M., Van Nieuwerburgh, S., 2007. Reconciling the Return Predictability Evidence. Review of Financial Studies 21, 1607–1652.

[57] Lewellen, J., Nagel, S., 2006. The Conditional CAPM does not Explain Asset-Pricing Anomalies. Journal of Financial Economics 82, 289-314.

[58] Li, Y., Yang, L., 2009. Underconditioning and Overconditioning: Testing the Conditional CAPM and the Conditional Fama-French Three-Factor Model. Working paper, Cornell University.

[59] Lin, C., Teräsvirta, T., 1994 Testing the Constancy of Regression Parameters Against Continuous Structural Change. Journal of Econometrics 62, 211–228.

[60] 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.

[61] Mamaysky, H., Spiegel M., Zhang, H., 2008. Estimating the Dynamics of Mutual Fund Alphas and Betas. The Review of Financial Studies 21, 233-264.

[62] Merton, R., 1980. On Estimating the Expected Return on the Market. Journal of Financial Economics 8, 323-361.

[63] Mykland, P., Zhang, L., 2006. ANOVA for Diffusions and Ito Processes. Annals of Statistics 34, 1931-1963.

[64] 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.

[65] Opsomer, J., Wang, Y., Yang, Y., 2001. Nonparametric Regression with Correlated Errors. Statistical Science 16, 134-153.

[66] Paye, B., Timmermann, A., 2006. Instability of Return Prediction Models. Journal of Empirical Finance 13, 274–315.

[67] Petkova, R., Zhang, L., 2005. Is Value Riskier than Growth? Journal of Financial Economics 78, 187-202.

[68] Phillips, P., 1987. Towards an Unified Asymptotic Theory for Autoregression. Biometrika 74, 535-547.

[69] Powell, J., Stock, J., Stoker, T., 1989. Semiparametric Estimation of Index Coefficients. Econometrica 57, 1403-1430.

[70] Powell, J., Stoker, T., 1996. Optimal Bandwidth Choice for Density-Weighted Averages. Journal of Econometrics 75, 291-316.

[71] Robinson, P., 1988. Root-N-Consistent Semiparametric Regression. Econometrica 56, 931-954.

[72] Robinson, P., 1989. Nonparametric Estimation of Time-Varying Parameters. In: Hackl, P. (Ed.), Statistical Analysis and Forecasting of Economic Structural Changes. Pp. 253-264.

[73] 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.

[74] Scholes, M., Williams, J., 1977. Estimating Betas from Non-synchronous Data. Journal of Financial Economics 5, 309-328.

[75] Shanken, J., 1990. Intertemporal Asset Pricing: An Empirical Investigation. Journal of Econometrics 45, 99-120.

[76] Stanton, R., 1997. A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. Journal of Finance 52, 1973-2002.

[77] Wang, K., 2003. Asset Pricing with Conditioning Information: A New Test. Journal of Finance 58, 161-196.

[78] Wooldridge, J., White, H., 1988. Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes. Econometric Theory 4, 210-230.

[79] Zhang, L., 2005. The Value Premium. Journal of Finance 60, 67-103.

Table 1: Summary Statistics of Factors and Portfolios

Panel A: Factors

			Correlations		
	Mean	Stdev	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	
	Mean	Stdev	$\hat{\alpha}_{OLS}$	$\hat{\beta}_{OLS}$
Book-to-Market Portfolios				
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 Portfolios				
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

Note to Table 1
We report summary statistics of Fama and French (1993) factors in Panel A and book-to-market and momentum portfolios in Panel B. Data is at a daily frequency and spans July 1964 to 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 the 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 ($\hat{\beta}_{OLS}$). These are obtained by regressing the daily portfolio excess returns onto daily market excess returns.

Table 2: Alphas and Betas of Book-to-Market Portfolios

	Bandwidth		Stdev of Conditional Estimates		Long-Run Estimates	
	Fraction	Months	Alpha	Beta	Alpha	Beta
1 Growth	0.0474	50.8	0.0121	0.0558	-0.0226	1.1705
					(0.0078)	(0.0039)
2	0.0989	105.9	0.0028	0.0410	-0.0037	1.0547
					(0.0069)	(0.0034)
3	0.0349	37.4	0.0070	0.0701	0.0006	0.9935
					(0.0072)	(0.0034)
4	0.0294	31.5	0.0136	0.0727	0.0043	0.9467
					(0.0077)	(0.0035)
5	0.0379	40.6	0.0113	0.0842	0.0077	0.8993
					(0.0083)	(0.0039)
6	0.0213	22.8	0.0131	0.0871	0.0187	0.8858
					(0.0080)	(0.0038)
7	0.0188	20.1	0.0148	0.1144	0.0275	0.8767
					(0.0084)	(0.0038)
8	0.0213	22.8	0.0163	0.1316	0.0313	0.8444
					(0.0082)	(0.0039)
9	0.0160	17.2	0.0184	0.1497	0.0373	0.8961
					(0.0094)	(0.0046)
10 Value	0.0182	19.5	0.0232	0.1911	0.0461	0.9556
					(0.0112)	(0.0055)
10-1 Book-to-Market Strategy	0.0217	23.3	0.0346	0.2059	0.0681	-0.2180
					(0.0155)	(0.0076)

Joint test for $\alpha_{LR,i} = 0, i = 1, ..., 10$
Wald statistic $W_0 = 31.6$, p-value = 0.0005

The table reports conditional bandwidths ($\hat{h}_{i,2}$ in equation (31)) and various statistics of conditional and long-run alphas and betas from a conditional CAPM of the book-to-market portfolios. The bandwidths are reported in fractions of the entire sample, which corresponds to 1, and in monthly equivalent units. We transform the fraction to a monthly equivalent unit by multiplying by $533 \times 1.96/0.975$, 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 estimates in the columns labeled "Stdev of Conditional 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 equation (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 W_0 in equation (20). 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.

Table 3: Tests of Constant Conditional Alphas and Betas of Book-to-Market Portfolios

	W_1		Critical Values	
	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

We test for constancy of the conditional alphas and betas in a conditional CAPM 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_1 given in equation (26) and 95% and 99% critical values of the asymptotic distribution. We mark rejections at the 99% level with **.

Table 4: Characterizing Conditional Betas of the Value-Growth Strategy

	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

We regress conditional betas of the value-growth strategy onto various macro variables. The betas are computed from a conditional CAPM and are plotted in Figure 3. The dividend yield is the sum of past 12-month dividends divided by current market capitalization of the CRSP value-weighted market portfolio. The default spread is the difference between BAA and 10-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 10-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 NBER recession variable is a zero/one indicator which takes on the variable one if the NBER defines a recession that month. All RHS 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.

Table 5: Long-Run Fama-French (1993) Alphas and Factor Loadings of Book-to-Market Portfolios

	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 $W_0 = 77.5$, p-value = 0.0000

The table reports long-run estimates of 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 equation (20). 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.

Table 6: Tests of Constant Conditional Fama-French (1993) Alphas and Factor Loadings of Book-to-Market Portfolios

	W_1				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

The table reports W_1 test statistics from equation (26) 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_1 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.

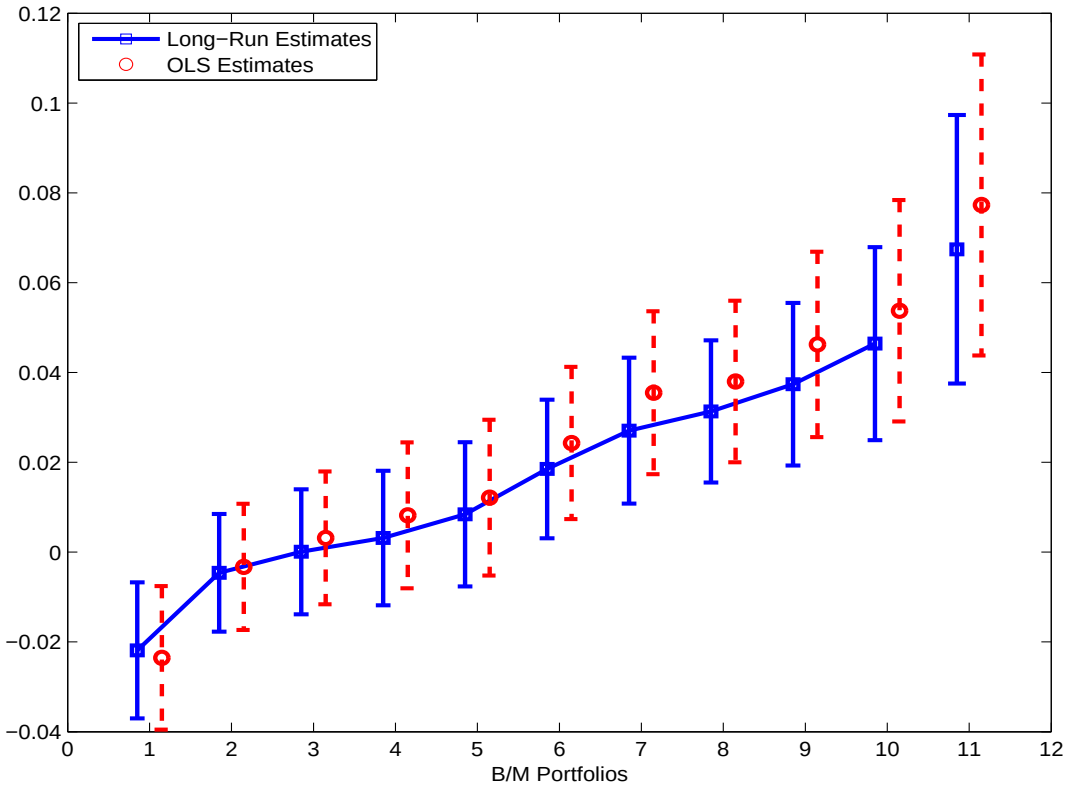
Table 7: Long-Run Fama-French (1993) Alphas and Factor Loadings of Momentum Portfolios

	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 $W_0 = 91.0$, p-value = 0.0000

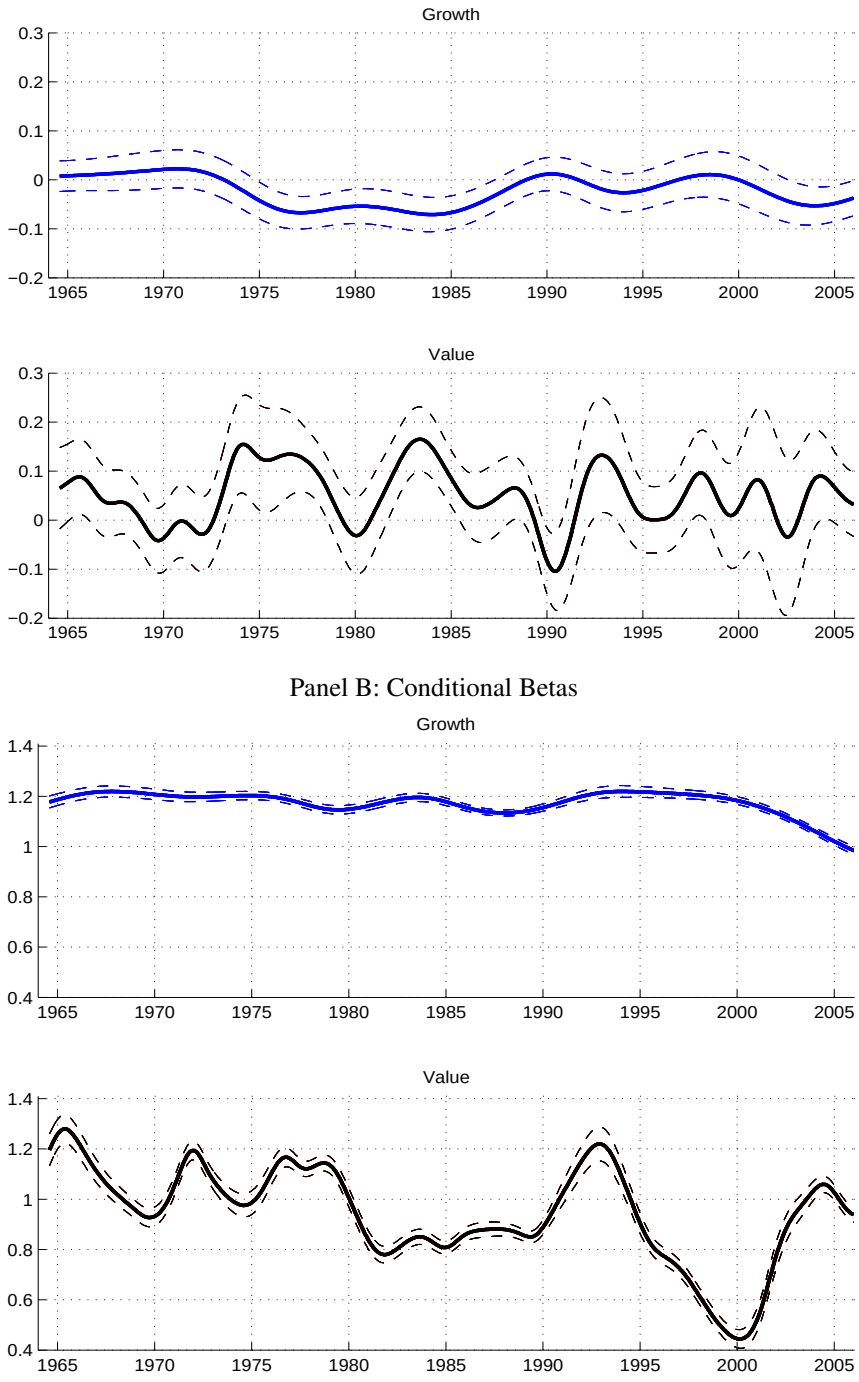
The table reports long-run estimates of alphas and factor loadings from a conditional Fama and French (1993) model applied to decile momentum portfolios and the 10-1 momentum strategy. The long-run estimates, with standard errors in parentheses, are computed following Theorem 2 and average daily estimates 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 equation (20). 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.

Figure 1: Long-Run Conditional CAPM Alphas versus OLS Alphas for the Book-to-Market Portfolios



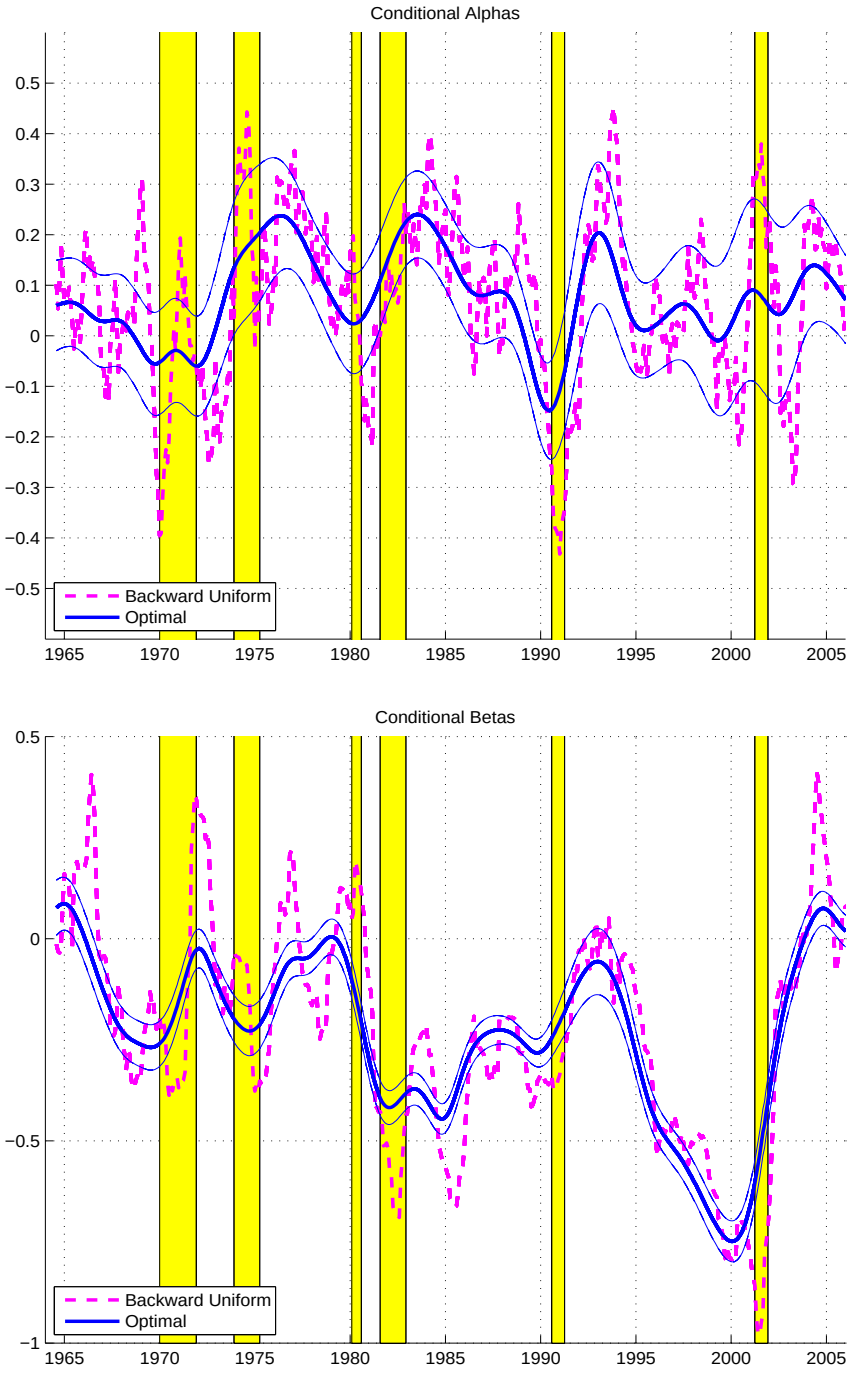
We plot long-run alphas implied by a conditional CAPM and OLS alphas for the book-to-market portfolios. We plot the long-run alphas using squares with 95% confidence intervals displayed by the solid error bars. The point estimates of the OLS alphas are plotted as circles with 95% confidence intervals 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.

Figure 2: Conditional Alphas and Betas of Growth and Value Portfolios



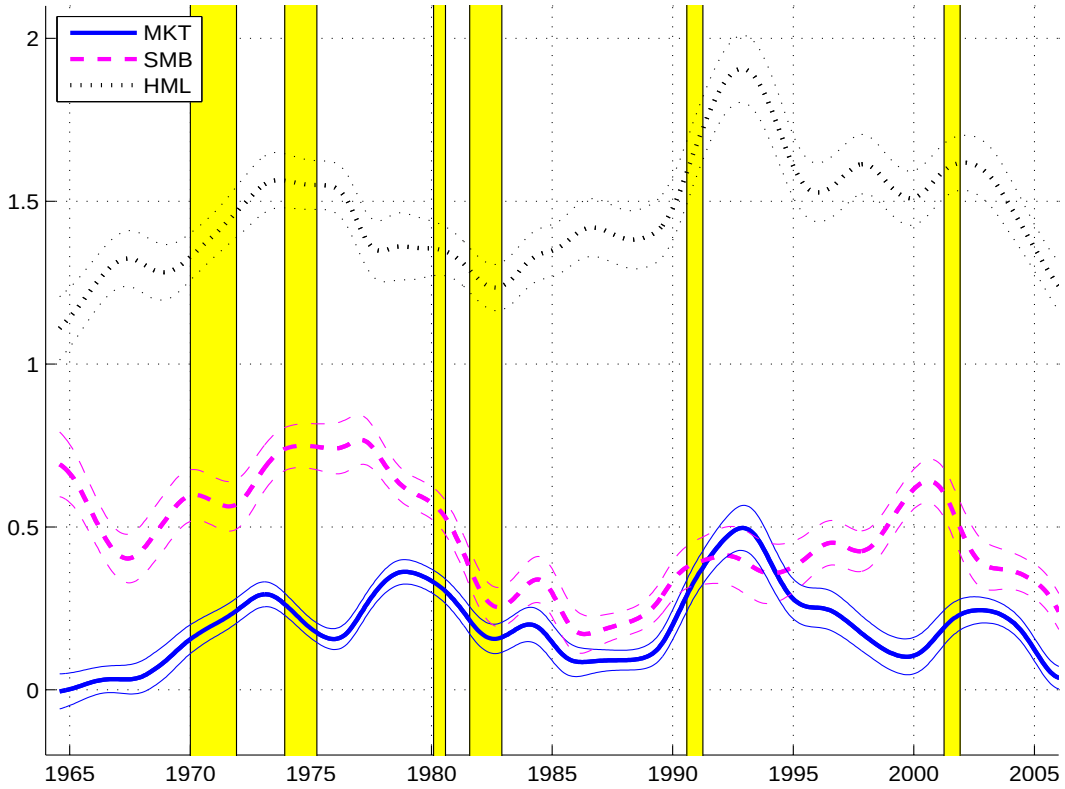
The figure shows monthly estimates of conditional alphas (Panel A) and conditional betas (Panel B) from a conditional CAPM of the first and tenth decile book-to-market portfolios (growth and value, respectively). We plot 95% confidence bands in dashed lines. The conditional alphas are annualized by multiplying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

Figure 3: Conditional Alphas and Betas of the Book-to-Market Strategy



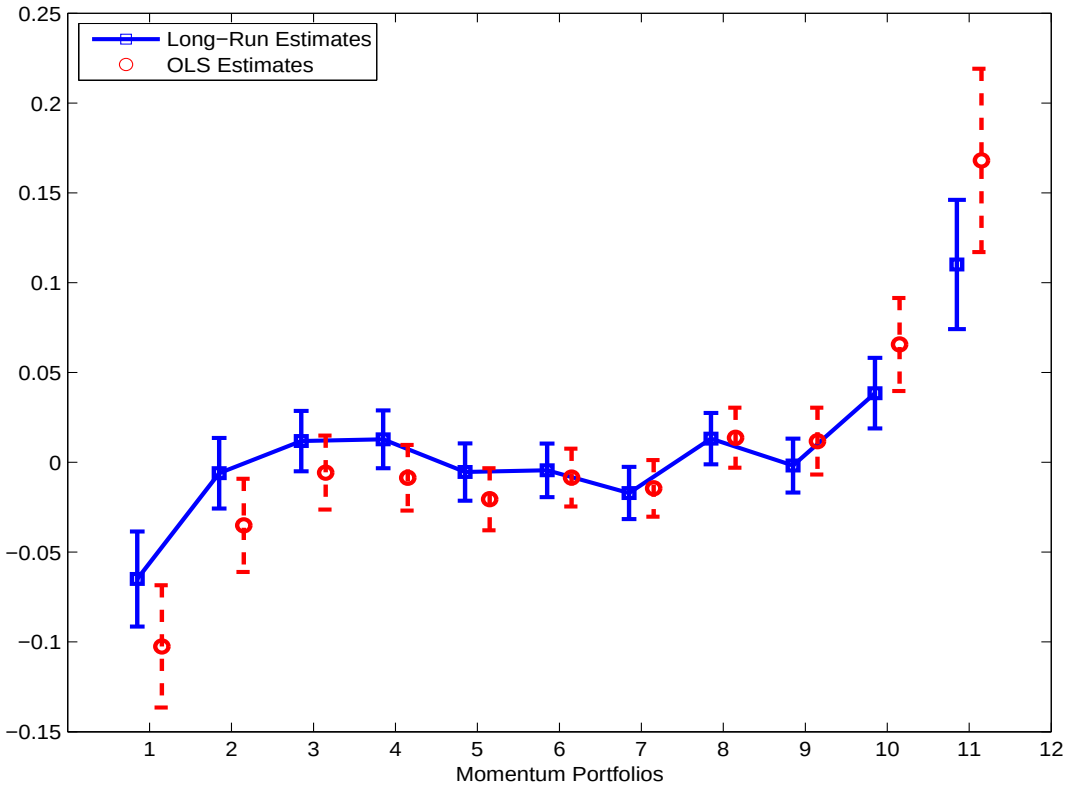
The figure shows monthly estimates of conditional alphas (top panel) and conditional betas (bottom panel) of the book-to-market strategy. We plot the optimal estimates in bold solid lines along with 95% confidence bands in regular solid lines. We also overlay the backward one-year uniform estimates in dashed lines. NBER recession periods are shaded in horizontal bars.

Figure 4: Conditional Fama-French (1993) Loadings of the Book-to-Market Strategy



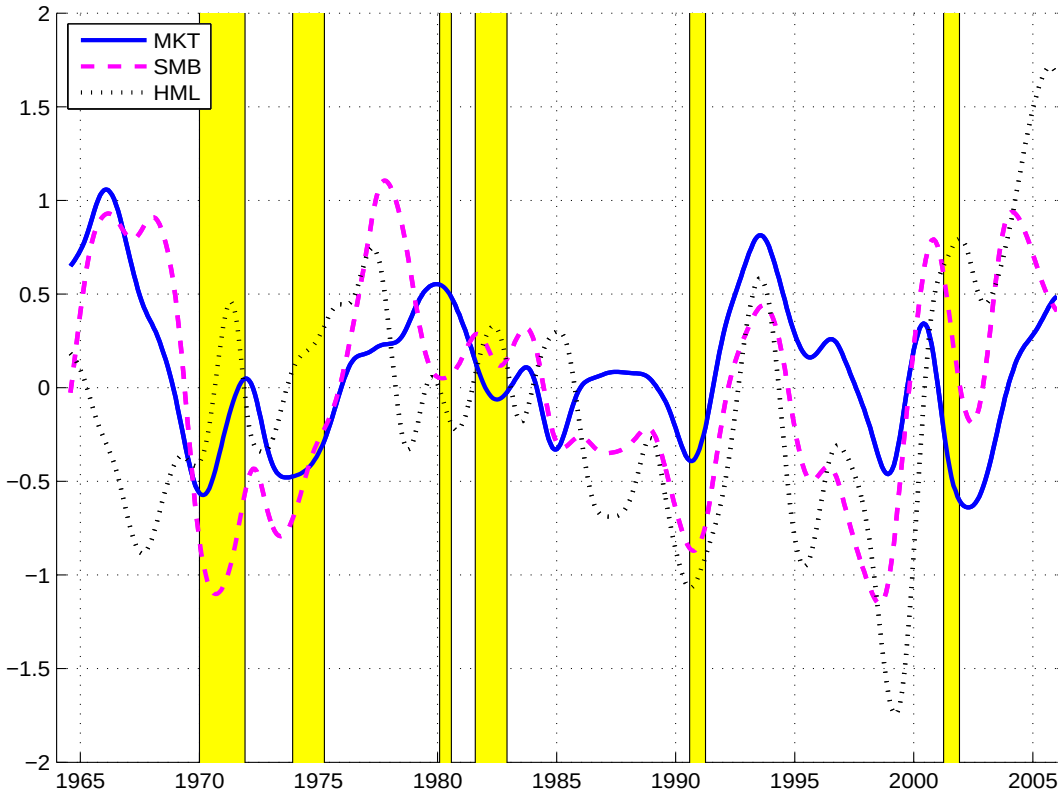
The figure shows monthly estimates of conditional Fama-French (1993) factor loadings of the book-to-market strategy, which goes long the 10th book-to-market decile portfolio and short the 1st book-to-market decile portfolio. We plot the optimal estimates in bold lines along with 95% confidence bands in regular lines. NBER recession periods are shaded in horizontal bars.

Figure 5: Long-Run Fama-French (1993) Alphas versus OLS Alphas for the Momentum Portfolios



We plot long-run alphas from a conditional Fama and French (1993) model and OLS Fama-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.

Figure 6: Conditional Fama-French (1993) Loadings of the Momentum Strategy



The figure shows monthly estimates of conditional Fama-French (1993) factor loadings of the momentum strategy, which goes long the 10th past return decile portfolio and short the 1st past return decile portfolio. NBER recession periods are shaded in horizontal bars.