Testing Conditional Factor Models*

Andrew Ang[†]
Columbia University and NBER

Dennis Kristensen[‡]
Columbia University and CREATES

This Version: 19 April, 2009

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

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

*We thank Pengcheng Wan for research assistance and Kenneth French for providing data. We thank Jonathan Lewellen, Serena Ng, and Jay Shanken for helpful suggestions, and seminar participants at Dartmouth University, Emory University, Oxford-Man Institute, Banff International Research Station Conference on Semiparametric and Nonparametric Methods in Econometrics, and Humboldt-Copenhagen Conference on Recent Developments in Financial Econometrics. Kristensen acknowledges funding from the Danish National Research Foundation.

[†]E-mail: aa610@columbia.edu

[‡]E-mail: dk2313@columbia.edu

Tests of Conditional Factor Models

Abstract

We develop a methodology for estimating time-varying factor loadings and conditional alphas based on nonparametric techniques. We test whether long-run alphas, or 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 when there is no time variation in the factor loadings. As applications of the methodology, we estimate conditional CAPM and Fama and French (1993) 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. Consequently, a popular time-series specification test of a factor model consists of testing whether the intercept term, or alpha, is equal to zero when the asset's excess return is regressed onto tradeable factors. Traditional tests of whether an alpha is equal to zero, like the widely used Gibbons, Ross and Shanken (1989) test, crucially 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 even at the portfolio level (see, among others, Fama and French, 1997; Lewellen and Nagel, 2006; Ang and Chen, 2007). The time variation in factor loadings distorts the standard factor model tests, which assume constant betas, for whether the alphas are equal to zero and, thus, renders traditional statistical inference for the validity of a factor model to be possibly misleading in the presence of time-varying factor loadings.

We introduce a methodology that tests for the significance of conditional alphas in the presence of time-varying betas. The tests can be run for an individual stock return, or jointly across assets. We build on the insights of Merton (1980), Foster and Nelson (1996), and Lewellen and Nagel (2006), among others, to use high frequency data to estimate factor loadings. We consider a class of models where as data are sampled at higher frequencies, estimates of variances and covariances, and hence betas, converge to their true values. Our insight is that, while high-frequency data characterize the distribution of covariances and hence betas, high-frequency data can also be used to characterize the distribution of conditional alphas. Unlike previous approaches which separate inference of conditional alphas and betas, our methodology derives their joint distribution, both at each moment in time and their long-run distributions across time. The tests can be applied to single assets or jointly specified across a system of assets.

In our methodology, tests of conditional alphas take into account the sampling variation of the conditional betas. These make our tests similar to the traditional maximum likelihood tests of the original CAPM developed by Gibbons (1982) and Gibbons, Ross and Shanken (1989). In the maximum likelihood set-up, the sampling variation of beta directly enters the standard error of the alpha when both the alpha and beta are estimated simultaneously.[1] The influential work by Gibbons, Ross and Shanken (1989) derives a statistic and distribution for testing whether the alphas of a set of base assets are jointly equal to zero and their tests of alphas take into account

[1] In a case with just one asset this is exactly the same as the standard error of a regular OLS intercept term depending on the mean and variance of the independent variable.

the sampling uncertainty of the factor loadings. Our tests can be viewed as the conditional analogue of the Gibbons, Ross and Shanken tests when the betas are time varying.

Our tests are straight forward to apply, powerful, and are based on no more than standard nonparametric estimates of OLS regression coefficients. We derive the asymptotic distribution of conditional alphas and betas to take into account the efficiency gains both from increasing the total length of the sample and from sampling at higher frequencies. With appropriate technical conditions, we derive a joint asymptotic distribution for the conditional alphas and betas at every point in time. We then construct a test statistic that 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 nonparametric estimators generally converge at slower rates than maximum likelihood 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 alpha equals zero are asymptotically equivalent to Gibbons, Ross and Shanken (1989).

We apply our tests to portfolios of stocks sorted on their book-to-market ratios and past returns. It is still a disputed question whether time variation in factor loadings in a conditional CAPM can explain the book-to-market effect. Petkova and Zhang (2005) and Ang and Chen (2007) argue that risk from time-varying market betas is enough to account for a substantial amount of the value premium. On the other hand, Lewellen and Nagel (2006) argue that the book-to-market effect cannot be explained by a conditional CAPM. The momentum effect of Jegadeesh and Titman (1993) is one of the most robust patterns of expected returns and so far no widely accepted factor model can explain the momentum effect, including the Fama-French (2003) model (see, e.g., Fama and French, 1996). Authors documenting significant time-varying factor exposure of momentum returns include Ball, Kothari and Shanken (1995) and Grundy and Martin (2001) and time-varying betas may account for a large part of momentum profitability. We find for both decile portfolios sorted by book-to-market ratios and past returns, long-run alphas are jointly significantly different from zero both for conditional versions of the one-factor market model and the three-factor Fama and French (1993) model. We also overwhelmingly reject the null that the conditional betas do not vary over time for both sets of portfolios.

Our approach builds on a literature advocating the use of short windows 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 (1986) estimate time-varying factor loadings and infer conditional alphas. Our work extends this literature in several important ways. First, 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 betas at any moment in time. Naturally, our optimal kernels can be adjusted to use one-sided, equal-weighted weighted filters which nest the approach of French, Schwert and Stambaugh (1987), Andersen et al. (2006), Lewellen and Nagel (2006), and others, as a special case.

Second, Lewellen and Nagel's (2006) procedure identifies the time variation of conditional betas and provides period-by-period estimates of conditional alphas. Lewellen and Nagel use only the time-series variation of these conditional alphas when conducting statistical tests of the alphas. But, since the alphas are a function of conditional betas which are also estimated, any inference on conditional alphas should take into account the sampling error of the time-varying factor loadings. Our procedure does precisely that. Our estimates of the sampling variation of conditional betas directly affect, and are simultaneously estimated with, the standard errors of the implied conditional alphas.

Third, we derive both univariate and joint tests of conditional alphas in the presence of time-varying betas. Similar to Gibbons, Ross and Shanken (1989) we are able to test for the significance of long-run alphas jointly over a set of portfolios. This is important because portfolios constructed by sorting over various characteristics are extensively used in finance to test factor models and it is common to test the efficiency of various small sets of investable assets. Following Fama and French (1993), and many others, a joint test over portfolios is useful for investigating whether a relation between conditional alphas and firm characteristics strongly exists across many portfolios.

Our work is most similar to tests of conditional factor models contemporaneously proposed 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. 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.

We apply our methodology to investigate if conditional CAPM and Fama-French (1993) models can price portfolios sorted on book-to-market ratios and past returns. 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, respectively. Section 6 concludes. We relegate all technical proofs to the appendix.

2 Statistical Methodology

In Section 2.1, we lay out the conditional factor model. Section 2.2 develops general conditional estimators and their distributions. We develop a test for long-run alphas and betas in Section 2.3 and tests for constancy of the conditional alphas and factor loadings in Section 2.4. Section 2.5 discusses the optimal bandwidth choice. We discuss other related finance literature in Section 2.6.

2.1 The Model

Let $R_{i,t}$ be the excess return of asset i at time t, for $i = 1, ..., N$ assets and $t = 1, ..., T$ periods. We wish to explain the returns through a set of J common tradeable factors, $f_t = (f_{1,t}, ..., f_{J,t})'$. We consider the following empirical specification:

$$\begin{aligned} R_{i,t} &= \alpha_{i,t} + \beta_{i1,t} f_{1,t} + ... + \beta_{iJ,t} f_{J,t} + \varepsilon_{i,t} \\ &= \alpha_{i,t} + \beta_{i,t}' f_t + \varepsilon_{i,t}, \end{aligned} \tag{1}$$

where $\alpha_{i,t} \in \mathbb{R}$ is the conditional alpha for stock i, $\beta_{i,t} = (\beta_{i1,t}, ..., \beta_{iJ,t})' \in \mathbb{R}^J$ is the vector of time-varying factor loadings for stock i, and the vector of error terms $\varepsilon_t = (\varepsilon_{1,t}, ..., \varepsilon_{N,t})'$ satisfies

$$E[\varepsilon_t | f_t, \beta_t] = 0 \quad \text{and} \quad E[\varepsilon_t \varepsilon_t' | f_t, \beta_t] = \Omega_t,$$

where $\Omega_t \in \mathbb{R}^{N \times N}$, $t = 1, ..., T$, is a sequence of covariance matrices. For convenience, we express equation (1) in vector notation:

$$R_t = \alpha_t + \beta_t' f_t + \varepsilon_t, \tag{2}$$

where $R_t = (R_{1,t}, ..., R_{N,t})'$ is the vector of returns, $\alpha_t = (\alpha_{1,t}, ..., \alpha_{N,t})' \in \mathbb{R}^N$ is the vector of conditional alphas across stocks $i = 1, ..., N$, and $\beta_t = (\beta_{1,t}, ..., \beta_{N,t})' \in \mathbb{R}^{J \times N}$ is the corresponding matrix of conditional betas. We collect the alphas and betas in $\gamma_t = (\alpha_t, \beta_t')' \in \mathbb{R}^{(J+1) \times N}$.

We are interested in the time series estimates of the conditional alphas, α_t, and the conditional factor loadings, β_t, along with their standard errors. Under the null of a factor model, the conditional alphas are equal to zero, or $\alpha_t = 0$. In equations (1) and (2), the time-varying conditional factor loadings can be random processes in their own right as long as they are weakly independent of the factors.

The model in equation (1) is strictly a conditional factor model in which betas vary over time but the factor premia do not exhibit time-varying expected returns. This setup has the advantage of avoiding the bias of Boguth et al. (2007) when both factor loadings and risk premia vary contemporaneously. However, the methodology we develop also applies to conditional factor models with time-varying factor premia. Jagannathan and Wang (1996) show that a conditional CAPM with time-varying factor premia is equivalent to an unconditional factor model by suitably expanding the set of unconditional factors. The expanded set of factors can be interpreted as managed returns (see Ferson and Harvey, 1991; Cochrane, 2001). Our framework applies to these settings if we can place additional factors on the RHS of equation (1) to capture the effect of time-varying risk premia.

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 define the long-run alphas and betas for asset i to be

$$\alpha_{\mathrm{LR},i} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \alpha_{i,t} \in \mathbb{R} \qquad \text{and} \qquad \beta_{\mathrm{LR},i} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \beta_{i,t} \in \mathbb{R}^J, \tag{3}$$

for $i = 1, ..., N$. We use the terminology "long run" to distinguish the conditional alpha at a point in time, $\alpha_{i,t}$, from the conditional alpha averaged over the sample, $\alpha_{\mathrm{LR},i}$. We are particularly interested in examining the hypothesis that the long-run alphas are jointly equal to zero across N assets:

$$H_0 : \alpha_{\mathrm{LR},i} = 0, \quad i = 1, ..., N. \tag{4}$$

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 in a setting where both conditional alphas and betas vary over time.

An important comment is that the model does not have a direct interpretation as a continuous-time model sampled at increasing frequency with an increasing sample length T. A continuous-time specification similar to equation (2) would be

$$ds_t = \mu_t dt + \beta_t' dX_t + \Omega_t^{1/2} dW_t,$$

where $s_t = \log(S_t)$, S_t is the vector of asset prices at time t, and X_t are the factors. This is the ANOVA model considered in Andersen et al. (2006) and Mykland and Zhang (2006). A discrete-time approximation of this continuous-time model takes the form of equation (1) where $R_t = s_{t\Delta} - s_{(t-1)\Delta}$, $\alpha_t = \mu_{t\Delta} - \mu_{(t-1)\Delta}$, $f_{n,t} = X_{t\Delta} - X_{(t-1)\Delta}$, and $\varepsilon_t = \Omega_{(t-1)\Delta}^{1/2}(W_{t\Delta} - W_{(t-1)\Delta})$, where Δ is the discretization interval. However, an important difference between the continuous-time model and the model in equation (2) is that in continuous time, the variance of the rescaled errors, $\Delta W_t \equiv W_{t\Delta} - W_{(t-1)\Delta} = \Omega_{(t-1)\Delta}^{-1/2}\varepsilon_t$, decreases as $\Delta \to 0$. Under this assumption in continuous time, conditional alphas cannot be identified. In contrast, we assume that the variance of the rescaled errors remains fixed as the sampling interval tends to zero, which allows us to identify both conditional and long-run alphas.[2] Our model should be viewed as a sequence of discrete-time models sampled at increasing frequency over a fixed horizon T. Our asymptotics are derived when the sampling interval tends to zero. This is the case in practice as an econometrician is given a sample of length T and can sample at monthly, weekly, daily, or higher intra-day frequencies. In our empirical work, we sample at the daily frequency.

2.2 Conditional Estimators

Suppose we have observed returns and factors observed over time $t = 1, ..., T$. We propose the following local least-squares estimators of $\alpha_{i,\tau}$ and $\beta_{i,\tau}$ in equation (1) at any point τ in the time interval $1 \leq \tau \leq T$:

$$(\hat{\alpha}_{i,\tau}, \hat{\beta}_{i,\tau}')' = \arg\min_{(\alpha,\beta)} \sum_{t=1}^{T} K_{h_i T}(t-\tau)\left(R_{i,t} - \alpha_i - \beta' f_t\right)^2, \tag{5}$$

[2] Merton (1980) shows that for a fixed time span in a simple diffusion model, we cannot recover the drift term, but can only identify the diffusion term, as we sample at increasing frequencies. Kristensen (2008a) extends this result to show that we cannot nonparametrically identify the drift term at a given point in time without further restrictions. Thus, in a diffusion setting, we cannot identify the short-run alphas. A continuous-time framework allows estimation only of $\lim_{T\to\infty} \frac{1}{T}\int_0^T \mu_t dt$, which is continuous-time counterpart to our long-run alphas, under the assumption of stationarity.

for each asset $i = 1, ..., N$, where $K_{h_iT}(z) = K(z/(h_iT))/(h_iT)$ with $K(\cdot)$ being a kernel and $h_i = h_{T,i} > 0$ a sequence of bandwidths. The estimators are simply kernel-weighted least squares, and it is easily seen that

$$(\hat{\alpha}_{i,\tau}, \hat{\beta}'_{i,\tau})' = \left[\sum_{t=1}^{T} K_{h_iT}(t-\tau) X_t X_t'\right]^{-1} \left[\sum_{t=1}^{T} K_{h_iT}(t-\tau) X_t R'_{i,t}\right], \tag{6}$$

where $X_t = (1, f_t')'$.

The proposed estimator gives 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 will 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.

The bandwidth $h_i > 0$ 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 and used in the estimation. Thus, the bandwidth controls the bias and variance of the estimator, and it should in general be chosen differently from one sample to another. In particular, as 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 further detail in Section 2.5.

We run the kernel regression (5) separately stock by stock for $i = 1, ..., N$. This is a generalization of the regular OLS estimators, which are also run stock by stock in the tests of unconditional factor models like Gibbons, Ross and Shanken (1989). If the same bandwidth h is used for all stocks, our estimator of alphas and betas across all stocks take the simple form of a weighted multivariate OLS estimator,

$$(\hat{\alpha}_\tau, \hat{\beta}'_\tau)' = \left[\sum_{t=1}^{T} X_t K_{Th}(t-\tau) X_t'\right]^{-1} \left[\sum_{t=1}^{T} X_t K_{Th}(t-\tau) R_t'\right].$$

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 very 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 will present the asymptotic results for the estimators $\hat{\alpha}_\tau$ and $\hat{\beta}_\tau$ 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 $T \to \infty$.

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. We build on the work of Robinson (1989), further extended by Robinson (1991), Orbe, Ferreira and Rodriguez-Poo (2005), and Cai (2007), who originally proposed to use kernel methods to estimate varying-coefficients models where the coefficients are functions of time. We utilize these results to obtain the asymptotic properties of the estimator.

We first state a result regarding the asymptotic properties of the local least-squares estimator. This is a direct consequence of Kristensen (2008b, Theorem 1).

Theorem 1 *Assume that (A.1)-(A.4) given in the Appendix hold and the bandwidth is chosen such that* $nh \to \infty$ *and* $nh^5 \to 0$*. Then, for any* $1 \leq \tau \leq T$, $\hat{\gamma}_\tau = (\hat{\alpha}_\tau, \hat{\beta}'_\tau)'$ *satisfies*

$$\sqrt{Th}(\hat{\gamma}_\tau - \gamma_\tau) \xrightarrow{d} N\left(0, \kappa_2 \Lambda_\tau^{-1} \otimes \Omega_\tau\right), \tag{7}$$

where $\Lambda_\tau = E\left[X_\tau X'_\tau\right]$, $X_\tau = (1, f'_\tau)'$*, and* $\kappa_2 = \int K^2(z)\, dz = 0.2821$ *for the normal kernel.*

Furthermore, for any two $\tau_1 \neq \tau_2$*,*

$$\mathrm{Cov}\left(\sqrt{Th}\left(\hat{\gamma}_{\tau_1} - \gamma_{\tau_1}\right), \sqrt{Th}\left(\hat{\gamma}_{\tau_2} - \gamma_{\tau_2}\right)\right) \xrightarrow{p} 0. \tag{8}$$

The result in Theorem 1 says that our estimator is able to pin down the full trajectory of the latent process γ_τ as we sample more and more frequently. Due to the nonparametric nature of the estimator, this result holds true for a wide range of data generating processes for γ_τ. That is, we do not have to assume a specific parametric model for the dynamics of γ_τ in order for our estimator to work.

For the estimator to be consistent, we have to let the sequence of bandwidths shrink towards zero as the sample size grows, $h \equiv h_T \to 0$ as $T \to \infty$. This is required in order to remove any biases of the estimator.[3] At the same time, the bandwidth sequence cannot go to zero too fast, otherwise the variance of the estimator will blow up. Thus, choosing the bandwidth should be done with care, since the estimates may be sensitive to the bandwidth choice. Unfortunately, the theorem is silent regarding how the bandwidth should be chosen for a given sample, which is a problem shared by most other nonparametric estimators. There are however many data-driven methods for choosing the bandwidths that work well in practice and we discuss our bandwidth selection procedure in Section 2.5.

The rate of convergence is $\sqrt{Th}$ as is standard for a nonparametric estimator. This is slower than the classical convergence rate of $\sqrt{T}$ since $h \to 0$. However, below, we show that a test for an average alpha across the sample equal to zero converges at the $\sqrt{T}$ 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, N, do not affect the performance of the estimator. A further advantage is that the point estimates $\hat{\alpha}_{i,\tau}$ and $\hat{\beta}_{i,\tau}$ are 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.

To make the result in Theorem 1 operational, we need estimators of the asymptotic variance. Simple estimators of the two terms appearing in the asymptotic variance are obtained as follows:

$$\hat{\Lambda}_\tau = \frac{1}{T}\sum_{t=1}^{T} K_{hT}\left(t-\tau\right) X_t X_t' \quad \text{and} \quad \hat{\Omega}_\tau = \frac{1}{T}\sum_{t=1}^{T} K_{hT}\left(t-\tau\right) \hat{\varepsilon}_t \hat{\varepsilon}_t', \tag{9}$$

where $\hat{\varepsilon}_t = R_t - \hat{\alpha}_t - \hat{\beta}_t' f_t$, $t = 1, ..., T$, are the fitted residuals. Due to the independence across different values of τ, pointwise confidence bands can easily be computed.

[3] 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. 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. In our empirical work, we believe that non-synchronous trading is not a major issue as we work with value-weighted portfolios at the daily frequency.

It is possible to use Theorem 1 to test the hypothesis that $\alpha_\tau = 0$ for a given value of $1 \leq \tau \leq T$:

$$W(\tau) = \hat{\alpha}_\tau' \hat{V}_{\tau,\alpha\alpha}^{-1} \hat{\alpha}_\tau \xrightarrow{d} \chi_1^2, \tag{10}$$

where $\hat{V}_{\tau,\alpha\alpha}$ consist of the first $N \times N$ components of $\hat{V}_\tau = \kappa_2 \hat{\Lambda}_t^{-1} \otimes \hat{\Omega}_t$. Due to the independence of the estimates at different values of τ, we can also test the hypothesis across any finite set of, say, $m \geq 1$ time points, $\tau_1 < \tau_2 < ... < \tau_m$:

$$\bar{W} = \sum_{k=1}^{m} W(\tau_k) \xrightarrow{d} \chi_m^2. \tag{11}$$

However, this test is not able to detect all departures from the null, since we only test for departures at a finite number of time points (which has to remain fixed as $T \to \infty$). To test the conditional alphas being equal to zero uniformly over time, i.e. $\alpha_\tau = 0$ for all $1 \leq \tau \leq T$, we advocate using the test for constancy of the conditional alphas which we present in Section 2.4. This test is similar to Shanken (1990) without external state variables and does not have a direct analogy with Gibbons, Ross and Shanken (1989), unlike the test for long-run alphas we present in Section 2.3.

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 is close to either 0 or T. In particular, the estimator is asymptotically biased when evaluated at the end points,

$$E[\hat{\gamma}_0] \to \frac{1}{2}\gamma_0 \quad \text{and} \quad E[\hat{\gamma}_T] \to \frac{1}{2}\gamma_T \quad \text{as } h \to 0.$$

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 that we simply refrain from reporting estimates close to the two boundaries: All our theoretical results are established under the assumption that our sample has been observed in the time interval $[-a, T+a]$ for some $a > 0$, and we then only estimate γ_τ for $\tau \in [0, T]$. 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, 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_τ and β_τ instead of a local constant one. Usage of these kernels does not affect the asymptotic distributions we derive for long-run alphas and betas in Section 2.3. We leave these technical extensions to future work.

2.3 Tests for Long-Run Alphas

To test the null of whether the long-run alphas are equal to zero (H_0 in equation (4)), we construct an estimator of the long-run alphas in equation (3) from the estimators of the conditional alphas, α_τ, and the conditional betas, β_τ, at any point in time $1 \leq \tau \leq T$. A natural way to estimate the long-run alphas and betas would be to simply plug the pointwise kernel estimators into the expressions found in equation (3):

$$\hat{\alpha}_{\mathrm{LR},i} = \frac{1}{T}\sum_{t=1}^{T}\hat{\alpha}_{i,t} \quad \text{and} \quad \hat{\beta}_{\mathrm{LR},i} = \frac{1}{T}\sum_{t=1}^{T}\hat{\beta}_{i,t}. \tag{12}$$

The proposed long-run alpha and beta estimators can be interpreted as two-step semiparametric estimators and as such share some features with other semiparametric estimators found in the literature. Two particular estimators that are closely related are nonparametric estimators of consumer surplus (Newey and McFadden, 1994, Section 8) and semiparametric estimation of index coefficients (Powell, Stock and Stoker, 1989). These estimators can be written as integrals over a ratio of two kernel estimators. Our long-run alpha and beta estimators fit into this class of estimators since the first-step estimators $\hat{\gamma}_\tau = (\hat{\alpha}_\tau, \hat{\beta}_\tau')'$ in equation (6) are ratios of kernel functions and the long-run alpha and betas are integrals (averages) of these kernel estimators.

The following theorem states the joint distribution of $\hat{\gamma}_{\mathrm{LR}} = (\hat{\alpha}_{\mathrm{LR}}, \hat{\beta}_{\mathrm{LR}}')' \in \mathbb{R}^{(J+1)\times N}$ where $\hat{\alpha}_{\mathrm{LR}} = (\hat{\alpha}_{\mathrm{LR},1}, ..., \hat{\alpha}_{\mathrm{LR},N})' \in \mathbb{R}^N$ and $\hat{\beta}_{\mathrm{LR}} = (\hat{\beta}_{\mathrm{LR},1}, ..., \hat{\beta}_{\mathrm{LR},N})' \in \mathbb{R}^{J\times N}$:

Theorem 2 *Assume that (A.1)-(A.6) given in the Appendix hold. Then,*

$$\sqrt{T}(\hat{\gamma}_{\mathrm{LR}} - \gamma_{\mathrm{LR}}) \xrightarrow{d} N\left(0, V_{\mathrm{LR}}\right), \tag{13}$$

where

$$V_{\mathrm{LR}} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Lambda_t^{-1}\otimes\Omega_t.$$

In particular,

$$\sqrt{T}(\hat{\alpha}_{\mathrm{LR}} - \alpha_{\mathrm{LR}}) \xrightarrow{d} N\left(0, V_{\mathrm{LR},\alpha\alpha}\right), \tag{14}$$

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

$$V_{\mathrm{LR},\alpha\alpha} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Lambda_{\alpha\alpha,t}^{-1}\otimes\Omega_t.$$

with

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

The asymptotic variance can be consistently estimated by:

$$\hat{V}_{\mathrm{LR}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\Lambda}_t^{-1} \otimes \hat{\Omega}_t, \tag{15}$$

where $\hat{\Lambda}_\tau$ and $\hat{\Omega}_\tau$ are given in equation (9).

An important observation is that the long-run estimators converge with standard parametric rate $\sqrt{T}$ despite the fact that they are based on preliminary estimators $\hat{\gamma}_\tau$ that converge at the slower, nonparametric rate $\sqrt{Th}$. 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 (6), as is well-known from other studies of semiparametric estimators; see, for example, Newey and McFadden (1994, Section 8) and Powell, Stock and Stoker (1989).

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

$$W_{\mathrm{LR}} = \hat{\alpha}_{\mathrm{LR}}' \hat{V}_{\mathrm{LR},\alpha\alpha}^{-1} \hat{\alpha}_{\mathrm{LR}} \in \mathbb{R}_+, \tag{16}$$

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

$$\hat{V}_{\mathrm{LR},\alpha\alpha} = \frac{1}{T}\sum_{t=1}^{T} \hat{\Lambda}_{\alpha\alpha,t}^{-1} \otimes \hat{\Omega}_t, \tag{17}$$

with

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

and

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

As a direct consequence of Theorem 2, we obtain

$$W_{\mathrm{LR}} \xrightarrow{d} \chi_N^2. \tag{18}$$

This is the conditional analogue Gibbons, Ross and Shanken (1989) and tests if long-run alphas are jointly equal to zero across all $i = 1, ..., N$ portfolios.

A special case of our model is when the factor loadings are constant with $\beta_t = \beta \in \mathbb{R}^{J\times N}$ 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 $\sqrt{T}(\hat{\alpha}_{\mathrm{LR}} - \alpha_{\mathrm{LR}})$ is identical to the standard Gibbons, Ross and Shanken (1989) test. This is shown in Appendix C. Thus, we pay no price asymptotically for the added robustness of our estimator. Furthermore, only in a setting with constant 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.

2.4 Tests for Constancy of the Alphas and Betas

In this section, we derive test statistics for the hypothesis that the conditional alphas or the betas, or both, are constant over time. The test can be applied to a subset of the full set of conditional alphas and betas. Since the proposed tests for constant alphas and betas are very similar, we treat them in a unified framework.

Suppose we wish to test for constancy of a subset of the time-varying parameters of stock i, $\gamma_{i,t} = \left(\alpha_{i,t}, \beta_{i,t}'\right)' \in \mathbb{R}^{J+1}$. We first 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:

$$R_{i,t} = \gamma_{i1,t}' X_{1,t} + \gamma_{i2,t}' X_{2,t} + \varepsilon_{i,t}. \tag{19}$$

We consider the following hypothesis:

$$H_1 : \gamma_{i1,t} = \gamma_{i1} \text{ for all } 0 \le t \le T. \tag{20}$$

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,[4]

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

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

$$\hat{\gamma}_{i2,\tau} = \arg\min_{\gamma_{i2}} \sum_{t=1}^{T} K_{h_i T}(t-\tau) \left[R_{i,t} - \gamma_{i1}' X_{1,t} + \gamma_{i2}' X_{2,t}\right]^2 = \hat{m}_{R_i,\tau} - \hat{m}_{1,\tau}\gamma_{i1}, \tag{21}$$

[4] To test $H_1 : \alpha_{i,t} = 0$ for all $0 \le t \le T$ simply set $\gamma_{i1} = \hat{\gamma}_{i1} = 0$.

where

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

are estimators of

$$\begin{aligned}m_{R_i,\tau} &= E\left[X_{2,\tau}X_{2,\tau}'\right]^{-1} E\left[X_{2,\tau}R_{i,\tau}'\right] \\ m_{1,\tau} &= E\left[X_{2,\tau}X_{2,\tau}'\right]^{-1} E\left[X_{2,\tau}X_{1,\tau}'\right].\end{aligned} \tag{22}$$

In the second stage, we obtain an estimator of the constant component γ_{i1}. We do this by substituting the conditional estimator $\hat{\gamma}_{2i,\tau}$ into the weighted least-squares criterion $Q_T(\gamma_{i1})$ given by:

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

where $\hat{X}_{1,t} = X_{1,t} - \hat{m}_{1,t}'X_{2,t} \in \mathbb{R}^m$ and $\hat{R}_{i,t} = R_{i,t} - \hat{m}_{R_i,t}'X_{2,t} \in \mathbb{R}$. Our estimator minimizes $Q_T(\gamma_{i1})$, which is again a simple least-squares problem with solution:

$$\hat{\gamma}_{i1} = \left[\sum_{t=1}^{T} \hat{\Omega}_{ii,t}^{-1}\hat{X}_{1,t}\hat{X}_{1,t}'\right]^{-1}\left[\sum_{t=1}^{T} \hat{\Omega}_{ii,t}^{-1}\hat{X}_{1,t}\hat{R}_{i,t}\right]. \tag{23}$$

The above estimator is akin to the residual-based estimator of Robinson (1988). It is also similar to the local linear profile estimator of Fan and Huang (2005) who demonstrate that in a cross-sectional framework with homoskedastic errors, the estimator of γ_{i1} is semiparametric efficient. Substituting equation (23) back into equation (21), the following estimator of the nonparametric component appears:

$$\hat{\gamma}_{i2,\tau} = \hat{m}_{R_i,\tau} - \hat{m}_{1,\tau}\hat{\gamma}_{i1}'. \tag{24}$$

Once the restricted estimators have been computed, we test H_1 by comparing the unrestricted and restricted model with a Wald test. Introducing the rescaled errors under the full model and under H_1 respectively as,

$$\hat{z}_t = \hat{\Omega}_{ii,t}^{-1/2}\hat{\varepsilon}_{it} \quad \text{and} \quad \hat{z}_{1,t} = \hat{\Omega}_{ii,t}^{-1/2}\hat{\varepsilon}_{i1,t},$$

where

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

are the residuals under the alternative and

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

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

$$SSR = \sum_{t=1}^{T} \hat{z}'_{i,t}\hat{z}_{i,t} \quad \text{and} \quad SSR_1 = \sum_{t=1}^{T} \hat{z}'_{i1,t}\hat{z}_{i1,t}.$$

The Wald test then takes the following form:

$$W_1 = \frac{T}{2}\frac{SSR_1 - SSR}{SSR}. \tag{25}$$

The proposed test statistic is related to the generalized likelihood-ratio test statistics advocated in Fan, Zhang and Zhang (2001):

Theorem 3 *Assume that (A.1)-(A.6) hold. Under H_1:*

$$\sqrt{T}(\hat{\gamma}_{i1} - \gamma_{i1}) \xrightarrow{d} N\left(0, \Sigma_{ii}^{-1}\right), \tag{26}$$

where, with $\hat{V}_t = X_{1,t} - m'_{1,t} X_{2,t}$,

$$\Sigma_{ii} = \lim_{T\to\infty} \sum_{t=1}^{T} \Omega_{ii,t}^{-1} V_t V'_t. \tag{27}$$

The test statistic satisfies

$$W_1 \xrightarrow{d} \chi^2_{q\mu}/q, \tag{28}$$

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 $c = 0.7737$.

The above result is in accordance with the results of Fan, Zhang and Zhang (2001). They demonstrate in a cross-sectional setting that test statistics of the form of W_1 are, in general, not dependent on nuisance parameters under the null and asymptotically converge to χ^2-distributions under the null. They further demonstrate that these statistics are 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.

2.5 Choice of Kernel and Bandwidth

As is common to all nonparametric estimators, the choice of the kernel and the bandwidth are important. Our theoretical results are based on using a kernel centered around zero and our main empirical results use the Gaussian kernel. In comparison, 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 (2006) corresponds to a uniform kernel on $[-1, 0]$ with $K(z) = \mathbb{I}\{-1 \leq z \leq 0\}$.

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 T 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 $O\left(T^{-2/5}\right)$ whereas the corresponding integrated RMSE is at most of order $O\left(T^{-1/3}\right)$. We provide further details in Appendix D.

There are two bandwidth selection issues unique to our estimators that we now discuss: Section 2.5.1 discusses bandwidth choices for the conditional estimates of alphas and betas while Section 2.5.2 treats the problem of specifying the bandwidth for the long-run alpha and beta estimators.

2.5.1 Bandwidth for Conditional Estimators

Our theoretical results establish rates at which the bandwidths should shrink to zero as the sample size grows. These are however not very informative about how one should choose the bandwidths for a given data set. We here propose data-driven rules for choosing the bandwidths in practice. We conducted simulation studies showing that the proposed methods work well in practice.

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.

To estimate the conditional bandwidths, we employ a plug-in method. For a symmetric kernel, the optimal bandwidth that minimizes the RMSE for stock i is

$$h_i^* = \left(\frac{||v_i||}{4||\zeta_i||^2} \right)^{1/5} T^{-1/5}, \tag{29}$$

where $v_i = T^{-1} \sum_{t=1}^{T} v_{i,t}$ and $\zeta_i = T^{-1} \sum_{t=1}^{T} \zeta_{i,t}$ are the integrated time-varying variance and bias components given by

$$v_{i,t} = \kappa_2 \Lambda_t^{-1} \otimes \Omega_t \quad \text{and} \quad \zeta_i = \Lambda_t^{-1} \left[\Lambda_t \gamma_{i,t}^{(2)} + 2\Lambda_t^{(1)} \gamma_{i,t}^{(1)} \right],$$

where $\Lambda_t^{(k)}$ and $\gamma_{i,t}^{(k)}$ for $k = 1, 2$ are the first and second order derivatives of Λ_t and $\gamma_{i,t}$, respectively. Ideally, we would compute v_i and ζ_i in order to obtain the optimal bandwidth given in equation (29). However, these depend on unknown components, Λ_t, $\gamma_{i,t}$, and Ω_t. In order to implement this bandwidth choice we therefore propose to obtain preliminary estimates of these through the following two-step method:[5]

1. Assume that $\Lambda_t = \Lambda$ and $\Omega_t = \Omega$ are constant, and $\gamma_t = a_0 + a_1 t + ... + a_p t^p$ is a polynomial. We then obtain parametric least-squares estimates $\hat{\Lambda}$, $\hat{\Omega}$ and $\hat{\gamma}_{i,t} = \hat{a}_{0,i} + \hat{a}_{1,i} t + ... + \hat{a}_{p,i} t^p$. Compute for each stock ($i = 1, ..., N$)

   $$\hat{v}_i = \kappa_2 \hat{\Lambda}^{-1} \otimes \hat{\Omega}_{ii} \quad \text{and} \quad \hat{\zeta}_i = \frac{1}{2T} \sum_{t=1}^{T} \hat{\gamma}_{i,t}^{(2)},$$

   where $\hat{\gamma}_{i,t}^{(2)} = 2\hat{a}_{2,i} + 6\hat{a}_{3,i} t + ... + p(p-1)\hat{a}_{p,i} t^{p-2}$. Then, using these estimates we compute the first-pass bandwidth

   $$\hat{h}_{i,1} = \left[\frac{||\hat{v}_i||}{\left(4||\hat{\zeta}_i||^2\right)} \right]^{1/5} \times T^{-1/5}. \tag{30}$$

2. Given $h_{i,1}$, compute the kernel estimators $\hat{\gamma}_{i,t} = \hat{\Lambda}_t^{-1} \sum_{t=1}^{T} K_{hT}(t - \tau) X_t R'_{i,t}$, $\hat{\Lambda}_t$ and $\hat{\Omega}_t$ as given in Section 2.2. We use these to obtain for each stock ($i = 1, ..., N$):

   $$\hat{v}_i = \kappa_2 \frac{1}{T} \sum_{t=1}^{T} \hat{\Lambda}_t^{-1} \otimes \hat{\Omega}_t \quad \text{and} \quad \hat{\zeta}_i = \frac{1}{2T} \sum_{t=1}^{T} \hat{\Lambda}_t^{-1} \left[\hat{\Lambda}_t \hat{\gamma}_{i,t}^{(2)} + 2\hat{\Lambda}_t^{(1)} \hat{\gamma}_{i,t}^{(1)} \right].$$

   These are in turn used to obtain a second-pass bandwidth:

   $$\hat{h}_{i,2} = \left[\frac{||\hat{v}_i||}{\left(4||\hat{\zeta}_i||^2\right)} \right]^{1/5} \times T^{-1/5}. \tag{31}$$

[5] Ruppert, Sheather and Wand (1995) discuss in detail how this can done in a standard kernel regression framework.

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.[6]

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.

2.5.2 Bandwidth for Long-Run Estimators

To estimate the long-run alphas and betas we re-estimate the conditional coefficients by undersmoothing relative to the bandwidth in 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 $E\left[||\hat{\gamma}_{\mathrm{LR},i} - \gamma_{\mathrm{LR},i}||^2\right]$, which we derive in Appendix E. As demonstrated there, the bandwidth used for the long-run estimators should be chosen to be of order $h_{\mathrm{LR},i} = O\left(T^{-2/(1+2r)}\right)$, where r is the number of derivatives required for the alpha and beta functions (or the degree of required smoothness). 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 $T^{-1/(1+2r)}$.

In our empirical work, we select the bandwidth for the long-run alphas and betas by first

[6] 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.

computing the optimal second-pass conditional bandwidth $\hat{h}_{i,2}$ in equation (31) and then scaling this down by setting

$$\hat{h}_{\text{LR},i} = \hat{h}_{i,2} \times T^{-1/(1+2r)}, \tag{32}$$

with a choice of $r = 1$.

2.6 Other Related Finance Literature

By taking advantage of nonparametric techniques to estimate latent quantities, we follow several papers in finance also using nonparametric estimators. Stanton (1997), Aït-Sahalia (1996), and Johannes (2004), among others, estimate drift and diffusion functions of the short rate using nonparametric estimators. Bansal and Viswanathan (1993), Aït-Sahalia and Lo (1998), and Wang (2003), among others, characterize the pricing kernel by nonparametric estimation. Brandt (1999) and Brandt and Aït-Sahalia (2007) present applications of nonparametric estimators to portfolio choice and consumption problems. Our work is the first, to our knowledge, to use nonparametric techniques to jointly estimate conditional alphas and betas in conditional factor models and, most importantly, to derive distributions of long-run alphas and factor loadings.

Our work is most motivated by Lewellen and Nagel (2006). Like Lewellen and Nagel our estimators of conditional alphas and betas use only information from high frequency data and ignore conditioning information from other instrumental variables. Alternative approaches taken by Shanken (1990) and Ferson and Harvey (1991, 1993), among many others, estimate time-varying factor loadings by instrumenting the factor loadings with macroeconomic and firm-specific variables. 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. Instead, only high frequency return data is used to obtain consistent estimates of alphas and betas. Thus, our estimator is in the same spirit of Lewellen and Nagel and uses local high frequency information, but we exploit a nonparametric structure.

A similar nonparametric approach is taken by Li and Yang (2009). Li and Yang also use nonparametric regressions to estimate conditional alphas and derive a similar test for average conditional alphas. However, they do not focus on conditional or long-run betas, or derive tests of constancy for 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. Finally, we show our test of long-run alphas across a set of base assets is a direct conditional analogue of Gibbons, Ross and Shanken (1989).

Our kernel specification used to estimate the time-varying betas nests several special cases in the literature. For example, French, Schwert and Stambaugh (1987) use daily data over the past month to estimate market variance. Lewellen and Nagel (2005) use daily returns over the past quarter or six months to estimate betas. Both of these studies use only truncated, backward-looking windows to estimate second moments. 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. 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 Philipov, 2005; Ang and Chen, 2007).

Because we use high frequency data to estimate second moments at lower frequencies, our estimator is also related to the realized volatility literature (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 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. Our estimators accommodate integrated betas as a special case by choosing a flat kernel and a fixed bandwidth.

3 Data

We apply our methodology to decile portfolios sorted by book-to-market ratios and decile portfolios sorted on past returns constructed by Kenneth French.[7] The book-to-market portfolios

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

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 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 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 reports 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 -2.19% with a standard error of 0.008 and the long-run alpha of the value portfolio is 4.64% with a standard error of 0.011. Both growth and value portfolios reject the conditional CAPM with p-values of their long-run alphas of 0.000. The long-run alpha of the book-to-market portfolio is 6.74% 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 -0.217 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 (16). The Wald test statistic is 32.95 with a p-value of 0.0003. 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 6.74%. Thus accounting for time-varying betas has reduced the OLS alpha by approximately only 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 $\sqrt{Th}$, which is less than the classical rate $\sqrt{T}$ 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 $\sqrt{T}$ 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 $\sqrt{T}$. 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.[8] However, the unconditional variation over the whole sample of the uniform filter estimates and the optimal estimates are similar. For example, the standard deviation of end-of-month conditional beta estimates from the uniform filter is 0.276, compared to 0.206 for the optimal two-sided conditional beta estimates. This implies that Lewellen and Nagel's (2007) 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

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

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 instruments in regression X measured at the beginning of the quarter. We define the market risk premium as the fitted value of this regression at the beginning of each quarter. We find that in regression XI, there is small positive covariation of conditional betas of value stocks with these fitted market risk premia with a coefficient of 0.37 and a standard error of 0.18. But, the adjusted R^2 of this regression is only 0.06. This is smaller than the covariation that Petkova and Zhang (2005) find because they specify betas as linear functions of the same state variables that drive the time variation of market risk premia. In summary, although conditional betas do covary with macro variables, 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

In this section, we examine alphas and factor loadings 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 2.03% for decile 1 to -1.67% 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 0.192. In contrast, the long-run SMB loading is relatively large at 0.450, 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 -0.670 for growth stocks to 0.804 for value stocks, with the book-to-market strategy having a long-run HML loading of 1.476. 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 for all book-to-market portfolios that the conditional alphas are constant. Whereas the conditional betas exhibited large time variation in the conditional CAPM, we now cannot reject that the conditional market factor loadings are constant. 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. 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 performed 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

In this section we test 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 -6.50% with a standard error of 0.014 for the first loser decile to 3.85% with a standard error of 0.010 to the tenth loser decile. The momentum strategy has a long-run alpha of 11.0% with a standard error of 0.018. 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 0.065. The long-run SMB loadings are small, except for the loser and winner deciles at 0.391 and 0.357, respectively. These effectively cancel in the momentum strategy, which is effectively SMB neutral. Finally, the long-run HML loadings are noticeably negative at -0.171 for the winner portfolio. The momentum strategy long-run HML loading is -0.113 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 11.0%. Thus, the conditional factor loadings have lowered the momentum strategy alpha by almost 7% 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.

In Table 8 we test for constancy of the Fama-French conditional alphas and factor loadings. Like the book-to-market portfolios, we cannot reject that conditional alphas are constant. However, for the momentum strategy all conditional factor loadings vary significantly through time. Table 8 shows that it is generally the loser and winner extreme quintiles that exhibit significant time-varying factor loadings and the middle quintiles generally fail to reject the null that the MKT, SMB, and HML loadings vary through time. We plot the time variation of these factor loadings for the momentum strategy in Figure 6.

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. 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 co-variation 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 single assets and also jointly for a system of assets. Like the classical time-series factor model tests which assume constant betas, the distributions of conditional alphas depend on, and are simultaneously estimated with, the distributions of conditional factor loadings. In the special case where there is no time variation in the factor loadings, 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. 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 refrained from reporting conditional estimates close to the beginning and end of the sample and so did 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:

$$R_{T,t} = \gamma_t' X_{T,t} + \Omega_t^{1/2} z_{T,t},$$

where $\gamma_t = (\alpha_t, \beta_t')'$ and $X_{T,t} = \left(1, f_{T,t}'\right)'$. We assume that we have observed data in the interval $[-a, T+a]$ for some fixed $a > 0$ to avoid any boundary issues and keep the notation simple. 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:

A.1 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 $\|u\| \geq L$. For some $r \geq 2$: $\int_{\mathbb{R}} K(z)\,dz = 1$, $\int_{\mathbb{R}} z^i K(z)\,dz = 0$, $i = 1, ..., r-1$, and $\int_{\mathbb{R}} |z|^r K(z)\,dz < \infty$.

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

A.3 $E[z_{T,t}|X_{T,t}] = 0$ and $E\left[z_{T,t} z_{T,t+k}'|X_{T,t}\right] = I_N$ if $k = 0$ and zero otherwise for all $1 \leq t \leq T$, $T \geq 1$.

A.3 The sequence γ_t is given by $\gamma_t = \gamma(t/T) + o(1)$ for some function $\gamma : [0,1] \mapsto \mathbb{R}^{(J+1)\times N}$ which lies in $C^r[0,1]$.

A.4 The matrix sequences $\Lambda_t \equiv E[X_t X_t']$ and Ω_t satisfy $\Lambda_t = \Lambda(t/T) + o(1)$ and $\Omega_t = \Omega(t/T) + o(1)$ for functions $\Lambda : [0,1] \mapsto \mathbb{R}^{(J+1)\times(J+1)}$ and $\Omega : [0,1] \mapsto \mathbb{R}^{N\times N}$ which are positive definite and with their individual elements lying in $C^r[0,1]$.

A.5 The covariance matrix Σ defined in equation (27) is non-singular.

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

The assumption (A.1) 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 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 negligible. The mixing and moment conditions in (A.2) are satisfied by most standard time-series models allowing, for example, f_t to solve 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. The smoothness conditions in (A.3)-(A.4) 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 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 is only needed for the semiparametric estimators and entails undersmoothing in the kernel estimation.

B Proofs

Proof of Theorem 1. The proof follows along the exact same lines as in Kristensen (2008b, Proof of Theorem 1), except that the response variable here is multivariate. This does not change any of the steps though. ■

Proof of Theorem 2. We write $K_{st} = K_h((s-t)/T)$ with similar notation for other variables. Define $\hat{\Lambda}_t = T^{-1}\sum_s K_{st} X_s X_s'$ and $\hat{m}_t = T^{-1}\sum_s K_{\tau,i} X_i R_i'$ such that:

$$\hat{\gamma}_t - \gamma_t = \hat{\Lambda}_t^{-1}\hat{m}_t - \Lambda_t^{-1} m_t, \quad \text{(B-1)}$$

where $m_t = E\left[X_t R_t'\right] = \Lambda_t \gamma_t$. By a second-order Taylor expansion of the right hand side,

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

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

$$\begin{aligned}\hat{m}_t &= m_t + O_P\left(h^r\right) + O_P\left(\sqrt{\log\left(T\right)/\left(Th\right)}\right) \\ \hat{\Lambda}_t &= \Lambda_t + O_P\left(h^r\right) + O_P\left(\sqrt{\log\left(T\right)/\left(Th\right)}\right), \end{aligned} \tag{B-3}$$

such that the two remainder terms are both $o_P(1/\sqrt{T})$ given the conditions imposed on h.

Using this result,

$$\begin{aligned}\hat{\gamma}_{\mathrm{LR}} - \gamma_{\mathrm{LR}} &= \frac{1}{T}\sum_{t=1}^{T}\left\{\Lambda_t^{-1}\left[\hat{m}_t - m_t\right] - \Lambda_t^{-1}\left[\hat{\Lambda}_t - \Lambda_t\right]\gamma_t\right\} + o_P(1/\sqrt{T}) \\ &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\{\Lambda_t^{-1}\left[K_{st}X_s R_s' - m_t\right] - \Lambda_t^{-1}\left[K_{st}X_s X_s' - \Lambda_t\right]\gamma_t\right\} + o_P(1/\sqrt{T}) \\ &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} a\left(Z_s, Z_t\right) + o_P(1/\sqrt{T}), \end{aligned} \tag{B-4}$$

where $Z_t = (\varepsilon_t, X_t, t)$ and

$$\begin{aligned} a\left(Z_s, Z_t\right) &= \Lambda_t^{-1}\left[K_{st}X_s R_s' - m_t\right] - \Lambda_t^{-1}\left[K_{st}X_s X_s' - \Lambda_t\right]\gamma_t \\ &= K_{st}\Lambda_t^{-1}X_s\left[R_s' - X_s'\gamma_t\right] \\ &= K_{st}\Lambda_t^{-1}X_s\left[X_s'\gamma_s + \varepsilon_s - X_s'\gamma_t\right] \\ &= K_{st}\Lambda_t^{-1}X_s\varepsilon_s' + K_{st}\Lambda_t^{-1}X_s X_s'\left[\gamma_s - \gamma_t\right]. \end{aligned}$$

Defining

$$\phi\left(Z_s, Z_t\right) = a\left(Z_s, Z_t\right) + a\left(Z_t, Z_s\right), \tag{B-5}$$

we may write

$$\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} a\left(Z_s, Z_t\right) = \frac{T-1}{T}U_T + \frac{1}{T^2}\sum_{t=1}^{T}\phi\left(Z_t, Z_t\right), \tag{B-6}$$

where $U_T = \sum_{s<T}\phi\left(Z_s, Z_t\right)/\left[T\left(T-1\right)\right]$. Here, $\sum_{t=1}^{T}\phi\left(Z_t, Z_t\right)/T^2 = O_P\left(1/T\right)$, while, by the Hoeffding decomposition, $U_T = 2\sum_{t=1}^{T}\bar{\phi}\left(Z_t\right)/T + \Delta_T$. The projection function $\bar{\phi}\left(z\right)$, $z = (e, x, \tau)$, is given by

$$\bar{\phi}\left(z\right) = E\left[\phi\left(z, Z_t\right)\right] = E\left[a\left(z, Z_t\right)\right] + E\left[a\left(Z_s, z\right)\right], \tag{B-7}$$

where

$$\begin{aligned} E\left[a\left(z, Z_t\right)\right] &= \int_0^T K_h\left(\tau - t\right)\Lambda^{-1}\left(t\right)xe'dt \\ &\quad + \int_0^T K_h\left(\tau - t\right)\Lambda^{-1}\left(t\right)xx'\left[\gamma\left(s\right) - \gamma\left(t\right)\right]dt \\ &= \Lambda_\tau^{-1}xe' + \Lambda_\tau^{-1}xx'\gamma_\tau^{(r)} \times h^r + o\left(h^r\right), \end{aligned}$$

$$\begin{aligned} E\left[a\left(Z_s, z\right)\right] &= \int_0^T K_h\left(s - \tau\right)\Lambda^{-1}\left(\tau\right)E\left[X_s\varepsilon_s'\right]ds \\ &\quad + \int_0^T K_h\left(s - \tau\right)\Lambda^{-1}\left(\tau\right)E\left[X_s X_s'\right]\left[\gamma\left(s\right) - \gamma\left(\tau\right)\right]ds \\ &= \gamma_\tau^{(r)} \times h^r + o\left(h^r\right), \end{aligned}$$

and $\gamma_t^{(r)} = \gamma^{(r)}(t/T)$ with $\gamma^{(r)}(\cdot)$ denoting the rth order derivative of $\gamma(\cdot)$. In total,

$$\bar{\phi}(Z_t) = \Lambda_t^{-1} \times X_t \varepsilon_t' + \Lambda_t^{-1} X_t X_t' \gamma_t^{(r)} \times h^r + \gamma_t^{(r)} \times h^r + o(h^r). \tag{B-8}$$

By Denker and Keller (1983, Proposition 2), the remainder term of the decomposition, Δ_T, satisfies $\Delta_T = O_P\left(T^{-1+\epsilon/2} s_{T,\delta}\right)$ for any $\epsilon > 0$, where $s_{T,\delta} \equiv \sup_{s,t} E\left[|\phi(Z_s, Z_t)|^{2+\delta}\right]^{1/(2+\delta)}$. Thus,

$$\begin{aligned}
s_{T,\delta}^{2+\delta} &\equiv \sup_{s,t} E\left[|\phi(Z_s, Z_t)|^{2+\delta}\right] && \text{(B-9)} \\
&\leq 2\sup_{s,t} E\left[|a(Z_s, Z_t)|^{2+\delta}\right] \\
&\leq 2\int_0^T \int_0^T |K_h(s-t)|^{2+\delta} \left\|\Lambda^{-1}(s)\right\|^{2+\delta} E\left[\|X_t\|^{2+\delta}\right] E\left[\|\varepsilon(t)\|^{2+\delta}\right] dsdt \\
&\quad +2\int_0^T \int_0^T |K_h(s-t)|^{2+\delta} \left\|\Lambda^{-1}(s)\right\|^{2+\delta} E\left[\|X_t\|^{2+\delta}\right] |\gamma(t) - \gamma(s)|^{2+\delta} dsdt \\
&\leq \frac{C}{h^{1+\delta}} \int_0^T \left\|\Lambda^{-1}(t)\right\|^{2+\delta} E\left[\|X_t\|^{2+\delta}\right] E\left[\|\varepsilon_t\|^{2+\delta}\right] dt + O(1) \\
&= O\left(h^{-(1+\delta)}\right). && \text{(B-10)}
\end{aligned}$$

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

$$\begin{aligned}
\sqrt{T}(\hat{\gamma}_{\text{LR}} - \gamma_{\text{LR}}) &= \sqrt{T} \sum_{t=1}^{T} \Lambda_t^{-1} X_t \varepsilon_t + O_P\left(\sqrt{T} h^r\right) \\
&\quad + O_P\left(\log(T) / \left(\sqrt{T} h\right)\right) + O_P\left(1/\left(T^{(1-\epsilon)/2} h^{7/8}\right)\right), && \text{(B-11)}
\end{aligned}$$

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

$$\sqrt{T} \sum_{t=1}^{T} \Lambda_t^{-1} X_t \varepsilon_t \to^d N(0, V_{\text{LR}}). \tag{B-12}$$

■

Proof of Theorem 3. This follows from Kristensen (2008b, Theorem 2). ■

C 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 $\tilde{\gamma}_{\text{LR}} = (\tilde{\alpha}_{\text{LR}}, \tilde{\beta}_{\text{LR}})$ is a standard least squares estimator of the form

$$\begin{aligned}
\tilde{\gamma}_{\text{LR}} &= \left[\sum_{t=0}^{T} X_t X_t'\right]^{-1} \left[\sum_{t=0}^{T} X_t R_t'\right] \\
&= \left[\sum_{t=0}^{T} X_t X_t'\right]^{-1} \sum_{t=0}^{T} X_t X_t' \gamma_t + \left[\sum_{t=0}^{T} X_t X_t'\right]^{-1} \sum_{t=0}^{T} X_t \varepsilon_t' \\
&= \bar{\gamma}_{\text{LR}} + U, && \text{(C-1)}
\end{aligned}$$

where

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

and

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

To separately investigate the performance of $\tilde{\alpha}_{\text{LR}}$, we note that $\bar{\gamma}_{\text{LR}} = \left(\bar{\alpha}_{\text{LR}}, \bar{\beta}_{\text{LR}}\right)'$ can be written as

$$\begin{aligned} \bar{\beta}_{\text{LR}} &= \left[\int_0^T \text{Var}\left(f_s\right)ds\right]^{-1}\int_0^T \text{Cov}\left(f_s, R_s\right)ds \\ \bar{\alpha}_{\text{LR}} &= \int_0^T \left(E\left[R_s\right]ds - \bar{\beta}_{\text{LR}}E\left[f_s\right]\right)ds, \end{aligned} \tag{C-2}$$

while $\gamma_{\text{LR}} = \left(\alpha_{\text{LR}}, \beta_{\text{LR}}\right)'$ can be written as

$$\begin{aligned} \beta_{\text{LR}} &= \int_0^T \left[\text{Var}\left(f_s\right)ds\right]^{-1}\text{Cov}\left(f_s, R_s\right)ds \\ \alpha_{\text{LR}} &= \int_0^T \left(E\left[R_s\right] - \beta_s E\left[f_s\right]\right)ds. \end{aligned} \tag{C-3}$$

From these two sets of expressions, we see that in general the GRS estimator will be inconsistent since it is centered around $\bar{\gamma}_{\text{LR}} \neq \gamma_{\text{LR}}$. However, in the case where $\beta_s = \beta$ is constant, $\bar{\beta}_{\text{LR}} = \beta_{\text{LR}}$ which in turn implies that $\bar{\alpha}_{\text{LR}} = \alpha_{\text{LR}}$. Thus, when the betas exhibit no time variation, $\tilde{\gamma}_{\text{LR}}$ is a consistent estimator of γ_{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.

D Two-Sided versus One-Sided Filters

One can show 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

$$\text{var}\left(\hat{\gamma}_{i,t}\right) = \frac{1}{Th}v_{i,t} + o\left(1/\left(Th\right)\right) \quad \text{with} \quad v_{i,t} = \kappa_2\Lambda_t^{-1}\Omega_{ii,t}, \tag{D-1}$$

while the bias is given by

$$\text{Bias}\left(\hat{\gamma}_{i,t}\right) = h^2\zeta_{i,t}^{\text{sym}} + o\left(h^2\right) \quad \text{with} \quad \zeta_{i,t}^{\text{sym}} = \mu_2\Lambda_t^{-1}\left[\frac{1}{2}\Lambda_t\gamma_{i,t}^{(2)} + \Lambda_t^{(1)}\gamma_{i,t}^{(1)}\right], \tag{D-2}$$

where we have assumed that Λ_t and $\gamma_{i,t}$ are twice differentiable with first and second order derivatives $\Lambda_t^{(k)}$ and $\gamma_{i,t}^{(k)}$, $k = 1, 2$. 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

$$\text{Bias}\left(\hat{\gamma}_{i,t}\right) = h\zeta_{i,t}^{\text{one}} + o\left(h\right) \quad \text{with} \quad \zeta_{i,t}^{\text{one}} = \mu_1\Lambda_t^{(1)}\gamma_{i,t}^{(1)}. \tag{D-3}$$

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

As a consequence, for the symmetric kernel the optimal bandwidth is

$$h_i^* = \left(\frac{||v_i||}{4||\zeta_i^{\text{sym}}||^2}\right)^{1/5}T^{-1/5}, \tag{D-4}$$

where $\zeta_i^{\text{sym}} = T^{-1}\sum_{t=1}^T \zeta_{i,t}^{\text{sym}}$ and $v_i = T^{-1}\sum_{t=1}^T v_{i,t}$ are the integrated versions of the time-varying bias and variance components. With this bandwidth choice, the integrated RMSE is of order $O\left(T^{-2/5}\right)$, where the integrated RMSE is defined as

$$\left(\int E[\|\hat{\gamma}_{i,\tau} - \gamma_{i,\tau}\|^2]d\tau\right)^{\frac{1}{2}}.$$

If on the other hand a one-sided kernel is used, the optimal bandwidth is

$$h_i^* = \left(\frac{||v_i||}{2||\zeta_i^{\text{one}}||^2}\right)^{1/3} T^{-1/3}, \tag{D-5}$$

with the corresponding integrated RMSE being of order $O\left(T^{-1/3}\right)$. Thus, the symmetric kernel integrated RMSE is generally smaller and substantially smaller if T is large.[9]

E Bandwidth Choice for Long-Run Estimators

We here follow the arguments of Härdle, Hall and Marron (1992), Stoker (1993) and Powell and Stoker (1996) to derive an optimal bandwidth for estimating the integrated or long-run gammas. We first note that, c.f. Proof of Theorem 2,

$$\hat{\gamma}_{\text{LR}} - \gamma_{\text{LR}} = \frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s<t} \phi(Z_s, Z_t) + o_P\left(1/\sqrt{T}\right), \tag{E-1}$$

where $Z_t = (\varepsilon_t, X_t, t)$, $\phi(Z_s, Z_t) = a(Z_s, Z_t) + a(Z_s, Z_t)$, and

$$a(Z_s, Z_t) = K_h(s-t)\Lambda_t^{-1} X_t \varepsilon_t' + K_h(s-t)\Lambda_t^{-1} X_t X_t' [\gamma_s - \gamma_t].$$

Thus, our estimator is approximately on the form of a U-statistic and the general result of Powell and Stoker (1996, Proposition 3.1) [PS] can be applied if we can verify their Assumptions 1 and 2. Their Assumptions 1-2 state that the function $a(Z_s, Z_t)$ has to satisfy (PS.i) $E[\phi(z, Z_i)] = s(z) h^{a_1} + o(h^{a_1})$ and (PS.ii) $E\left[\|\phi(z, Z_i)\|^2\right] = q(z) h^{-a_2} + o(h^{-a_2})$ for some $a_1, a_2 > 0$ and some functions $s(z)$ and $q(z)$.[10]

First, we verify (PS.i): Define

$$\bar{\phi}(z) = E[\phi(z, Z_i)] = E[a(z, Z_j)] + E[a(Z_i, z)]. \tag{E-2}$$

With $z = (e, x, \tau)$, it follows from the proof of Theorem 2 that

$$\begin{aligned} E[a(z, Z_t)] &= \Lambda^{-1}(\tau) xe' + \Lambda^{-1}(\tau) xx'\gamma^{(r)}(\tau) \times h^r + o(h^r) \\ E[a(Z_s, z)] &= \gamma^{(r)}(\tau) \times h^r + o(h^r), \end{aligned} \tag{E-3}$$

where $\gamma^{(r)}(t)$ is the rth order derivative of $\gamma(t)$. In total,

$$\bar{\phi}(Z_t) = \Lambda_t^{-1} X_t \varepsilon_t' + \Lambda_t^{-1} X_t X_t' \gamma_t^{(r)} \times h^r + \gamma_t^{(r)} \times h^r + o(h^r). \tag{E-4}$$

Thus, (PS.i) holds with $a_1 = r$ and

$$s(Z_t) = \Lambda_t^{-1} X_t X_t' \gamma_t^{(r)} + \gamma_t^{(r)}.$$

To verify (PS.ii), note that

$$\begin{aligned} E\left[\phi(Z_s, z)\phi(Z_s, z)'\right] &= E\left[a(Z_s, z) a(Z_s, z)'\right] + E\left[a(z, Z_t) a(z, Z_t)'\right] \\ &= \int K_h^2(s-\tau) ds \Lambda_\tau^{-2} xx'ee' + \int K_h^2(s-\tau)\Lambda_\tau^{-2} \|xx'\|^2 \|\gamma_s - \gamma_\tau\|^2 ds \\ &\quad + \int K_h^2(\tau - t)\Lambda_t^{-2} E[X_t X_t] E[\varepsilon_t \varepsilon_t'] dt \\ &\quad + \int K_h(\tau - t)\Lambda_t^{-1} E\left[\|X_t X_t'\|^2\right] \|\gamma_\tau - \gamma_t\|^2 dt \\ &= h^{-1}\kappa_2 \times \left[\Lambda_\tau^{-2} xx'ee' + \Lambda_\tau^{-1}\Omega_\tau\right] + o\left(h^{-1}\right). \end{aligned} \tag{E-5}$$

[9] 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.

[10] Their results are derived under the assumption of I.I.D. observations, but can be extended to hold under our assumptions (A.1)-(A.5).

Thus, (PS.ii) holds with $a_2 = 1$ and

$$q(Z_t) = \kappa_2 \Lambda_t^{-2} \left[X_t X_t' \varepsilon_t \varepsilon_t' + \Lambda_t \Omega_t \right].$$

It now follows from Powell and Stoker (1996, Proposition 3.1) that an approximation of the optimal bandwidth is given by

$$h_{\mathrm{LR},i} = \left[\frac{||E[q_i(Z)]||}{r||E[s_i(Z)]||^2} \right]^{1/(1+2r)} \times \left[\frac{1}{T} \right]^{2/(1+2r)}, \tag{E-6}$$

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

$$E[q_i(Z)] = 2\kappa_2 \times \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \Lambda_t^{-1} \Omega_{ii,t} \quad \text{and} \quad E[s_i(Z)] = 2 \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \gamma_{i,t}^{(r)}.$$

References

Aït-Sahalia, Y., 1996, "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527-560.

Aït-Sahalia, Y., and A. W. Lo, 1998, "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53, 499-547.

Aït-Sahalia, Y., and M. W. Brandt, 2007, "Nonparametric Consumption and Portfolio Choice with Option Implied State Prices," working paper, Duke University.

Andersen, T. G., T. Bollerselv, F. X. Diebold, and P. Labys, 2003, "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 529-626.

Andersen, T. G., T. Bollerselv, F. X. Diebold, and J. Wu, 2006, "Realized Beta: Persistence and Predictability," in T. Fomby and D. Terrell, 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, 1-40.

Andrews, D.W.K., 1991, "Asymptotic Optimality of Generalized C_L Cross-Validation, and Generalized Cross-Validation in Regression with Heteroskedastic Errors," *Journal of Econometrics*, 47, 359-377.

Ang, A., and J. Chen, 2007, "CAPM Over the Long Run: 1926-2001," *Journal of Empirical Finance*, 14, 1-40.

Ang, A., J. Chen, and Y. Xing, 2006, "Downside Risk," *Review of Financial Studies*, 19, 1191-1239.

Ball, R., S. P. Kothari, and J. Shanken, 1995, "Problems in Measuring Portfolio performance: An Application to Contrarian Investment Strategies," *Journal of Financial Economics*, 38, 79-107.

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

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2009, "Multivariate Realised Kernels: Consistent Positive Semi-definite Estimators of the Covariation of Equity Prices with Noise and Non-synchronous Trading," working paper, Oxford University.

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

Boguth, O., M. Carlson, A. Fisher, and M. Simutin, 2007, "Conditional Risk, Overconditioning, and the Performance of Momentum Strategies," working paper UBC.

Brandt, M. W., 1999, "Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach," *Journal of Finance*, 54, 1609-1645.

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

Campbell, J. Y., and J. Cochrane, 1999, "By Force of Habit," *Journal of Political Economy*, 107, 105-251.

Campbell, J. Y., and T. Vuolteenaho, 2004, "Bad Beta, Good Beta," *American Economic Review*, 94, 1249-1275.

Cochrane, J., 2001, *Asset Pricing*, Princeton University Press, Princeton, NJ.

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

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

Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3-56.

Fama, E. F., and K. R. French, 1996, "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 51, 55-84.

Fama, E. F., and K. R. French, 1997, "Industry Costs of Equity," *Journal of Financial Economics*, 43, 153-193.

Fan, J. and T. Huang, 2005, "Profile Likelihood Inferences on Semiparametric Varying-Coefficient Partially Linear Models," *Bernoulli*, 11, 1031-1057.

Fan, J., and W. Zhang, 2008, "Statistical Methods with Varying Coefficient Models," manuscript, Princeton.

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

Ferson, W. E., and C. R. Harvey, 1991, "The Variation of Economic Risk Premiums," *Journal of Political Economy*, 99, 2, 385-415.

Ferson, W. E., and C. R. Harvey, 1993, "The Risk and Predictability of International Equity Returns," *Review of Financial Studies*, 6, 3, 527-566.

Foster, D. P., and D. B. Nelson, 1996, "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica*, 64, 1, 139-174.

French, K. R., G. W. Schwert, and R. F. Stambaugh, 1987, "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3-29.

Ghysels, E., 1998, "On Stable Factor Structures in the Pricing of Risk: Do Time Varying Betas Help or Hurt? " *Journal of Finance*, 53, 549-574.

Gibbons, M. R., 1982, "Multivariate Tests of Financial Models," *Journal of Financial Economics*, 10, 3-27.

Gibbons, M. R., S. A. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 5, 1121-1152.

Gomes, J. F., L. Kogan, and L. Zhang, 2003, "Equilibrium Cross Section of Returns," *Journal of Political Economy*, 111, 4, 693-732.

Grundy, B. D., and J. S. Martin, 2001, "Understanding the Nature and Risks and the Source of the Rewards to Momentum Profits," *Journal of Financial Studies*, 14, 1, 29-78.

Härdle, W., P. Hall, and J. S. Marron, 1988, "How Far Are Automatically Chosen Regression Smoothing Parameters from Their Optimum?" *Journal of the American Statistical Association*, 83, 86-101.

Härdle, W., J. D. Hart, J. S. Marron, and A. B. Tsybakov, 1992, "Bandwidth Choice for Average Derivative Estimation," *Journal of the American Statistical Association*, 87, 218-233.

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

Harvey, C. R., 2001, "The Specification of Conditional Expectations," *Journal of Empirical Finance*, 8, 573-638.

Hayashi, T. and N. Yoshida, 2005, "On Covariance Estimation of Non-Synchronously Observed Diffusion Processes," *Bernoulli*, 11, 359-379.

Hurvich, C. M., J. S. Simonoff, and C. L. Tsai, 1998, "Smoothing Parameter Selection in Nonparametric Regression Using An Improved Akaike Information Criterion," *Journal of the Royal Statistical Society, Series B*, 60, 271-293.

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

Jegadeesh, N., and S. Titman, 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance*, 48, 65-91.

Johannes, M., 2004, "The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models," *Journal of Finance*, 59, 227-260.

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

Kanya, S., and D. Kristensen, 2008, "Estimation of Stochastic Volatility Models by Nonparametric Filtering," working paper, Columbia University.

Koch, I., and A. Pope, 1996, "Asymptotic Bias and Variance of A Kernel-based Estimator for the location of A Discontinuity," *Journal of Nonparametric Statistics*, 8, 45-64.

Kristensen, D., 2008a, "Nonparametric Filtering of the Realised Spot Volatility: A Kernel-based Approach," forthcoming in *Econometric Theory*.

Kristensen, D., 2008b, "Nonparametric Testing of Time-Varying Regression Models," working paper, Columbia University.

Kristensen, D., 2009, “Uniform Convergence Rates of Kernel Estimators with Heterogeneous, Dependent Data,” forthcoming in *Econometric Theory*.

Lewellen, J., and S. Nagel, 2006, “The Conditional CAPM does not Explain Asset-Pricing Anomalies,” *Journal of Financial Economics*, 82, 289-314.

Lettau, M., and S. Ludvigson, 2001a, “Consumption, Aggregate Wealth, and Expected Stock Returns,” *Journal of Finance*, 56, 815-849.

Lettau, M., and S. Ludvigson, 2001b, “Resurrecting the (C)CAPM: A Cross-Sectional Test when Risk Premia are Time Varying,” *Journal of Political Economy*, 109, 1238-1287.

Li, Y., and L. Yang, 2009, “Underconditioning and Overconditioning: Testing the Conditional CAPM and the Conditional Fama-French Three-Factor Model,” working paper, Cornell University.

Loader, C. R. 1996, “Change Point Estimation Using Nonparametric Regression,” *Annals of Statistics*, 24, 1667-1678.

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.

Merton, R. C., 1980, “On Estimating the Expected Return on the Market,” *Journal of Financial Economics*, 8, 323-361.

Mykland, P. A., and L. Zhang, 2006, “ANOVA for diffusions and Ito processes,”forthcoming *Annals of Statistics*

Newey, W. K., and D. L. McFadden, 1994, “Large Sample Estimation and Hypothesis Testing,” in Engle, R. F., and D. L. McFadden, eds., *Handbook of Econometrics*, vol. IV, 2111-2245.

Orbe, S., E. Ferreira, and J. Rodriguez-Poo, 2005, “Nonparametric Estimation of Time Varying Parameters Under Shape Restrictions,” *Journal of Econometrics*, 126, 53-77.

Petkova, R., and L. Zhang, 2005, “Is Value Riskier than Growth?” *Journal of Financial Economics*, 78, 187-202.

Powell, J. L., J. H. Stock, and T. M. Stoker, 1989, “Semiparametric Estimation of Index Coefficients,” *Econometrica*, 57, 1403-1430.

Powell, J. L., and T. M. Stoker, 1996, “Optimal Bandwidth Choice for Density-Weighted Averages,” *Journal of Econometrics*, 75, 291-316.

Robinson, P. M., 1988, “Root-N-Consistent Semiparametric Regression,” *Econometrica*, 56, 931-954.

Robinson, P. M., 1989, “Nonparametric Estimation of Time-Varying Parameters,” in P. Hackl, ed. *Statistical Analysis and Forecasting of Economic Structural Changes*, 253-264.

Rice, J., 1984, “Bandwidth Choice for Nonparametric Regression,” *Annals of Statistics*, 12, 1215-1230.

Ruppert, D., S. J. Sheather, and M. P. Wand, 1995, “An Effective Bandwidth Selector for Local Least Squares Regression,” *Journal of the American Statistical Association*, 90, 1257–1270.

Santos, T., and P. Veronesi, 2006, “Labor Income and Predictable Stock Returns,” *Review of Financial Studies*, 19, 1-44.

Scholes, M. and J. Williams, 1977, “Estimating Betas from Non-synchronous Data,” *Journal of Financial Economics*, 5, 309-328.

Shanken, J., 1990, “Intertemporal Asset Pricing: An Empirical Investigation,” *Journal of Econometrics*, 45, 99-120.

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

Stoker, T. M., 1993, “Smoothing Bias in Density Derivative Estimation,” *Journal of the American Statistical Association*, 88, 855-863.

Wang, K. Q., 2003, “Asset Pricing with Conditioning Information: A New Test,” *Journal of Finance*, 58, 161-196.

Wooldridge, J. M., and H. White, 1988, “Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes,” *Econometric Theory*, 4, 210-230.

Zhang, L., 2005, “The Value Premium,” *Journal of Finance*, 60, 67-103.

Zhang, W., S. Y. Lee, and X. Song, 2002, ”Local polynomial Fitting in semivarying coefficient models,” *Journal of Multivariate Analysis*, 82, 166-188.

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	-0.0219	1.1721
					(0.0077)	(0.0040)
2	0.0989	105.9	0.0028	0.0410	-0.0046	1.0563
					(0.0067)	(0.0034)
3	0.0349	37.4	0.0070	0.0701	0.0001	0.9938
					(0.0071)	(0.0035)
4	0.0294	31.5	0.0136	0.0727	0.0031	0.9456
					(0.0076)	(0.0035)
5	0.0379	40.6	0.0113	0.0842	0.0084	0.8990
					(0.0082)	(0.0039)
6	0.0213	22.8	0.0131	0.0871	0.0185	0.8865
					(0.0079)	(0.0038)
7	0.0188	20.1	0.0148	0.1144	0.0270	0.8777
					(0.0083)	(0.0039)
8	0.0213	22.8	0.0163	0.1316	0.0313	0.8444
					(0.0081)	(0.0039)
9	0.0160	17.2	0.0184	0.1497	0.0374	0.8966
					(0.0092)	(0.0046)
10 Value	0.0182	19.5	0.0232	0.1911	0.0464	0.9568
					(0.0110)	(0.0055)
10-1 Book-to-Market Strategy	0.0217	23.3	0.0346	0.2059	0.0674	-0.2170
					(0.0153)	(0.0077)

Joint test for $\alpha_{LR,i} = 0, i = 1, ..., 10$
Wald statistic $W_0 = 32.95$, p-value = 0.0003

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 $T = 11202$ 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 (16). 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

	Alpha			Beta		
		Critical Values			Critical Values	
	W_1	95%	99%	W_1	95%	99%
1 Growth	82	232	242	515**	232	242
2	17	116	123	512**	116	123
3	46	311	322	476**	311	322
4	83	367	378	496**	367	378
5	55	287	298	630**	287	298
6	96	501	515	695**	501	515
7	131	566	581	769**	566	581
8	136	502	516	850**	502	516
9	156	659	675	1047**	659	675
10 Value	209	583	598	1162**	583	598
10-1 Book-to-Market Strategy	212	491	505	951**	491	505

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 (25) 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 3-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	0.0203	0.9781	-0.1781	-0.6701
	(0.0055)	(0.0042)	(0.0061)	(0.0075)
2	0.0118	0.9693	-0.0650	-0.2776
	(0.0060)	(0.0044)	(0.0065)	(0.0080)
3	0.0056	0.9698	-0.0202	-0.1136
	(0.0067)	(0.0050)	(0.0074)	(0.0089)
4	-0.0054	0.9976	0.0178	0.1535
	(0.0072)	(0.0052)	(0.0076)	(0.0095)
5	-0.0014	0.9671	0.0021	0.2599
	(0.0075)	(0.0055)	(0.0082)	(0.0101)
6	-0.0014	0.9827	0.0637	0.2996
	(0.0072)	(0.0053)	(0.0079)	(0.0096)
7	-0.0119	1.0053	0.0871	0.4252
	(0.0071)	(0.0051)	(0.0077)	(0.0093)
8	-0.0132	1.0361	0.1057	0.7108
	(0.0057)	(0.0041)	(0.0062)	(0.0076)
9	-0.0163	1.1021	0.1369	0.7738
	(0.0068)	(0.0050)	(0.0077)	(0.0092)
10 Value	-0.0167	1.1699	0.2717	0.8040
	(0.0090)	(0.0066)	(0.0098)	(0.0121)
10-1 Book-to-Market Strategy	-0.0375	0.1924	0.4501	1.4756
	(0.0102)	(0.0075)	(0.0111)	(0.0136)

Joint test for $\alpha_{LR,i} = 0, i = 1, ..., 10$
Wald statistic $W_0 = 78.92$, 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 (16). 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

	Alpha		MKT		SMB		HML	
	W_1	95%	W_1	95%	W_1	95%	W_1	95%
1 Growth	102	518	499	518	782**	518	3285**	518
2	117	635	431	635	478	635	1090**	635
3	77	516	230	516	256	516	491	516
4	68	429	254	429	263	429	451**	429
5	55	391	245	391	230	391	713**	391
6	67	436	243	436	357	436	831**	436
7	92	644	256	644	483	644	1220**	644
8	119	673	278	673	547	673	3485**	673
9	74	492	234	492	548**	492	3173**	492
10 Value	79	440	420	440	649**	440	2237**	440
10-1 Book-to-Market Strategy	85	467	338	467	1089**	467	4313**	467

The table reports W_1 test statistics from equation (25) 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	-0.0650	1.1836	0.3913	-0.0560
	(0.0135)	(0.0093)	(0.0134)	(0.0164)
2	-0.0061	1.0442	0.0944	0.0292
	(0.0100)	(0.0073)	(0.0103)	(0.0128)
3	0.0118	0.9754	-0.0266	0.0495
	(0.0086)	(0.0062)	(0.0090)	(0.0111)
4	0.1284	0.9622	-0.0525	0.0778
	(0.0082)	(0.0060)	(0.0087)	(0.0106)
5	-0.0054	0.9360	-0.0564	0.0575
	(0.0081)	(0.0058)	(0.0084)	(0.0103)
6	-0.0045	0.9579	-0.0344	0.1081
	(0.0076)	(0.0055)	(0.0081)	(0.0099)
7	-0.0171	0.9831	-0.0255	0.0960
	(0.0074)	(0.0054)	(0.0079)	(0.0096)
8	0.0131	1.0228	-0.0270	0.0944
	(0.0073)	(0.0053)	(0.0079)	(0.0096)
9	-0.0018	1.0868	0.0758	0.0331
	(0.0076)	(0.0056)	(0.0083)	(0.0102)
10 Winners	0.0385	1.2501	0.3572	-0.1705
	(0.0100)	(0.0075)	(0.0107)	(0.0134)
10-1 Momentum Strategy	0.1101	0.0653	-0.0341	-0.1127
	(0.0184)	(0.0128)	(0.0184)	(0.0228)

Joint test for $\alpha_{LR,i} = 0, i = 1, ..., 10$
Wald statistic $W_0 = 159.7$, p-value = 0.0000

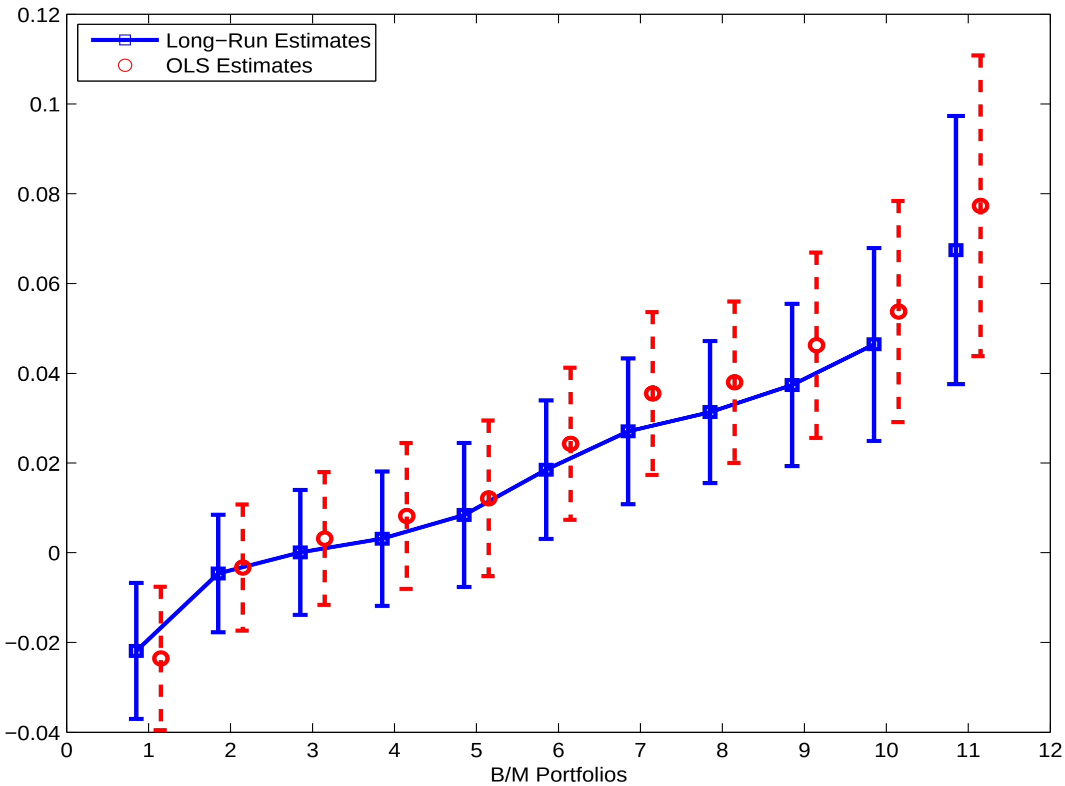
The table reports long-run estimates of alphas and factor loadings from a conditional Fama and French (1993) model applied to decile momentum portfolios and the 10-1 momentum strategy. The long-run estimates, with standard errors in parentheses, are computed following Theorem 2 and average daily estimates 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 (16). 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 8: Tests of Constant Conditional Fama-French (1993) Alphas and Factor Loadings of Momentum Portfolios

	Alpha		MKT		SMB		HML	
	W_1	95%	W_1	95%	W_1	95%	W_1	95%
1 Losers	195	677	743^{**}	677	1178^{**}	677	611^{**}	677
2	165	734	776^{**}	734	603	734	730^{*}	734
3	134	620	586	620	460	620	473	620
4	85	476	340	476	356	476	429	476
5	52	321	224	321	263	321	340^{**}	321
6	60	324	141	324	224	324	277	324
7	77	399	187	399	287	399	270	399
8	129	707	389	707	512	707	466	707
9	142	786	655	786	689	786	657	786
10 Winners	157	631	848^{**}	631	1090^{**}	631	897^{**}	631
10-1 Momentum Strategy	245	748	1000^{**}	748	945^{**}	748	906^{**}	748

The table reports W_1 test statistics in equation (25) 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 95% and 99% level with * and **, respectively. 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.

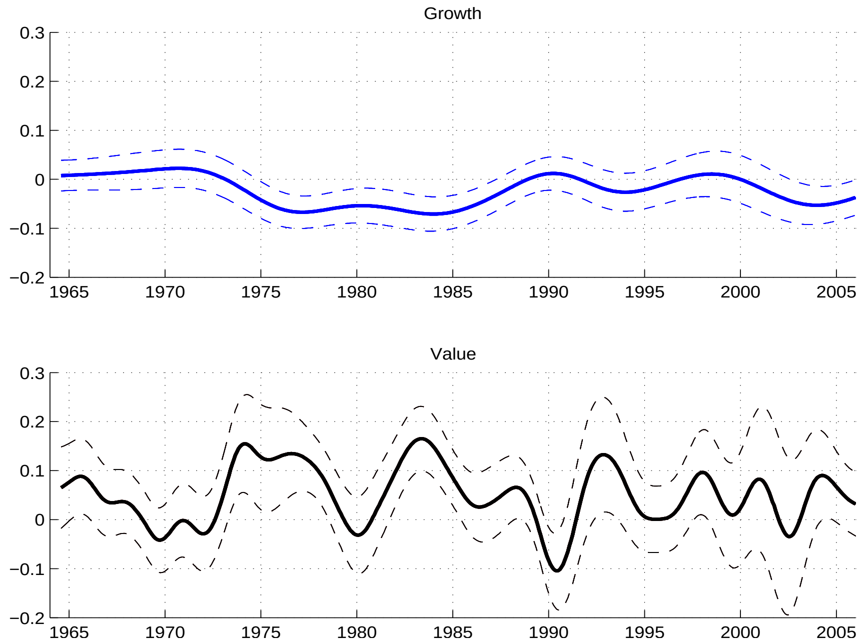
Figure 1: Long-Run Conditional CAPM Alphas versus OLS Alphas for the Book-to-Market Portfolios



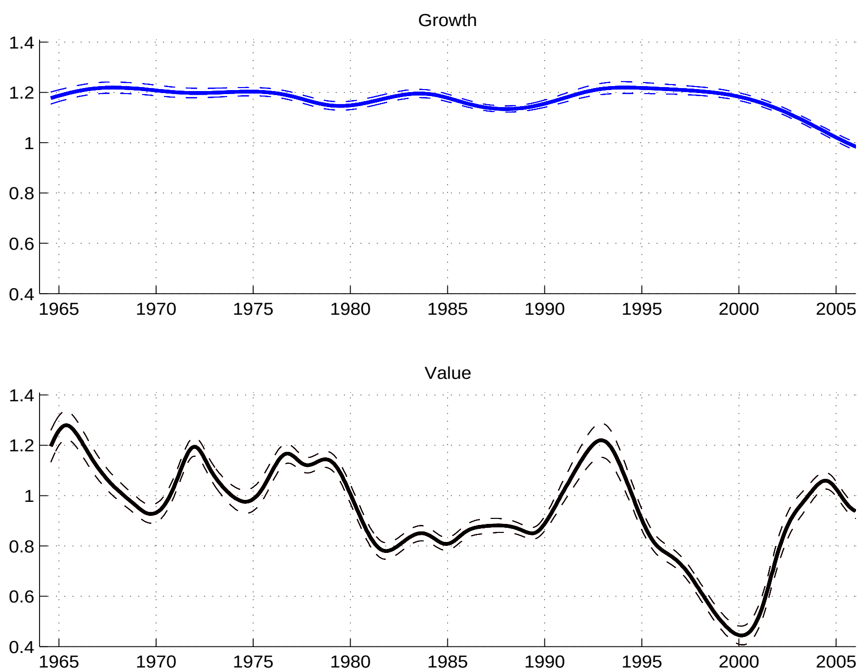
We plot long-run alphas implied by a conditional CAPM and OLS alphas for the book-to-market portfolios. We plot the long-run alphas using squares with 95% confidence intervals displayed by the solid error bars. The point estimates of the OLS alphas are plotted as circles with 95% confidence intervals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

Panel A: Conditional Alphas

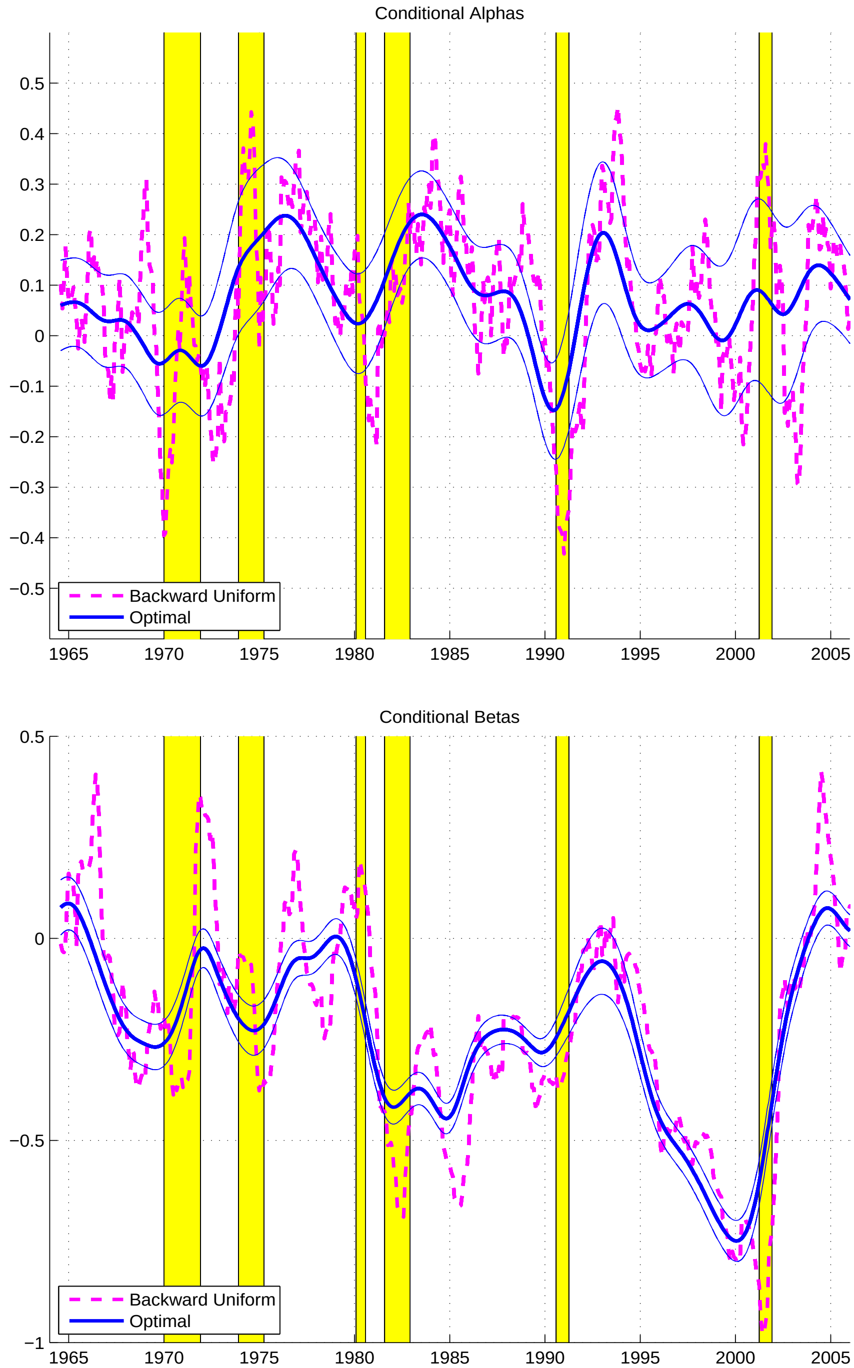


Panel B: Conditional Betas



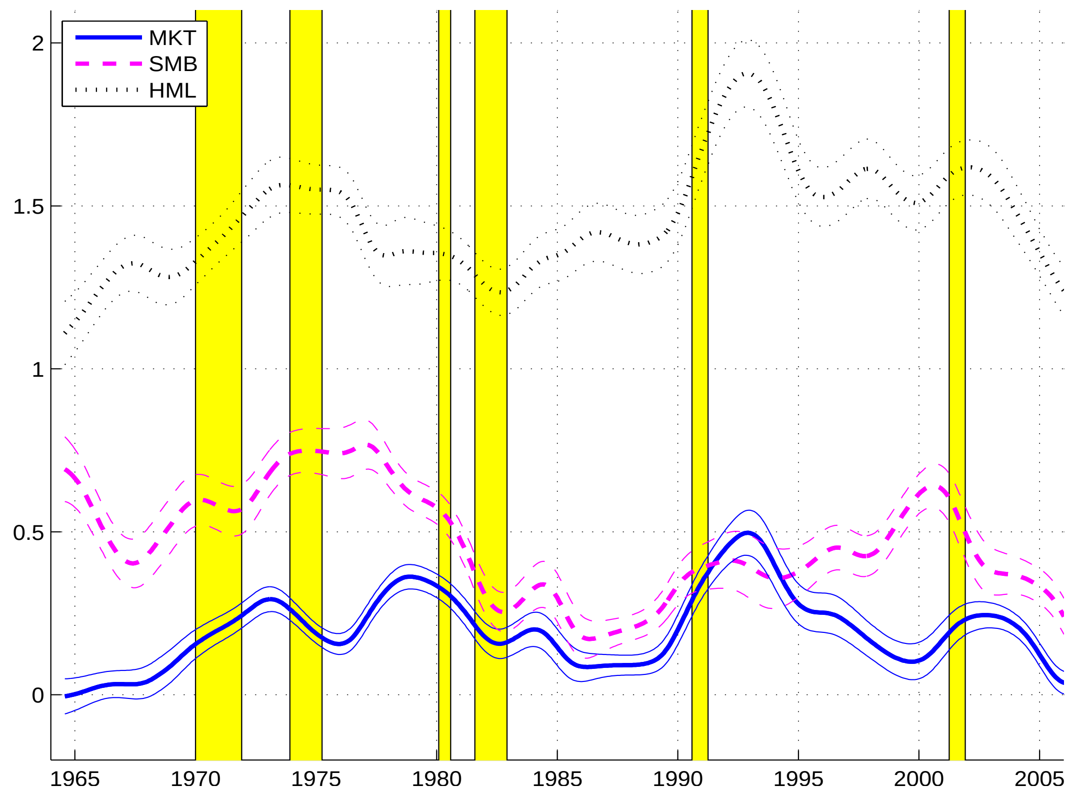
The figure shows monthly estimates of conditional alphas (Panel A) and conditional betas (Panel B) from a conditional CAPM of the first and tenth decile book-to-market portfolios (growth and value, respectively). We plot 95% confidence bands in dashed lines. The conditional alphas are annualized by multiplying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

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



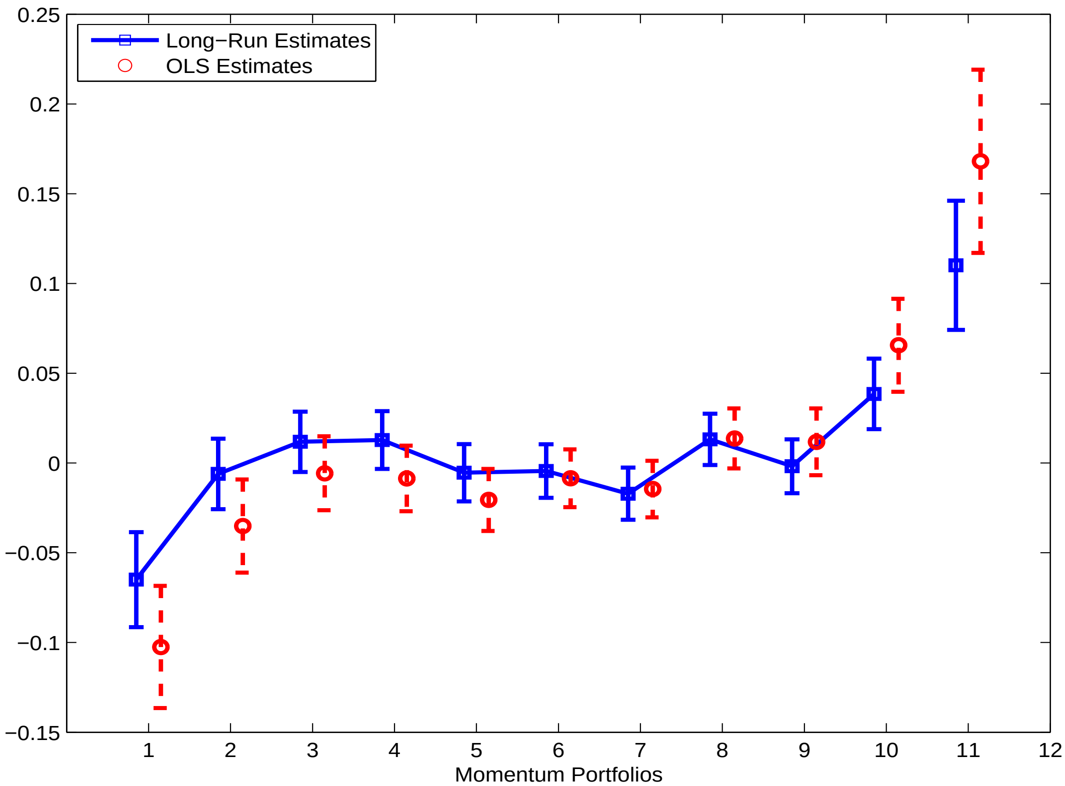
The figure shows monthly estimates of conditional alphas (top panel) and conditional betas (bottom panel) of the book-to-market strategy. We plot the optimal estimates in bold solid lines along with 95% confidence bands in regular solid lines. We also overlay the backward one-year uniform estimates in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

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



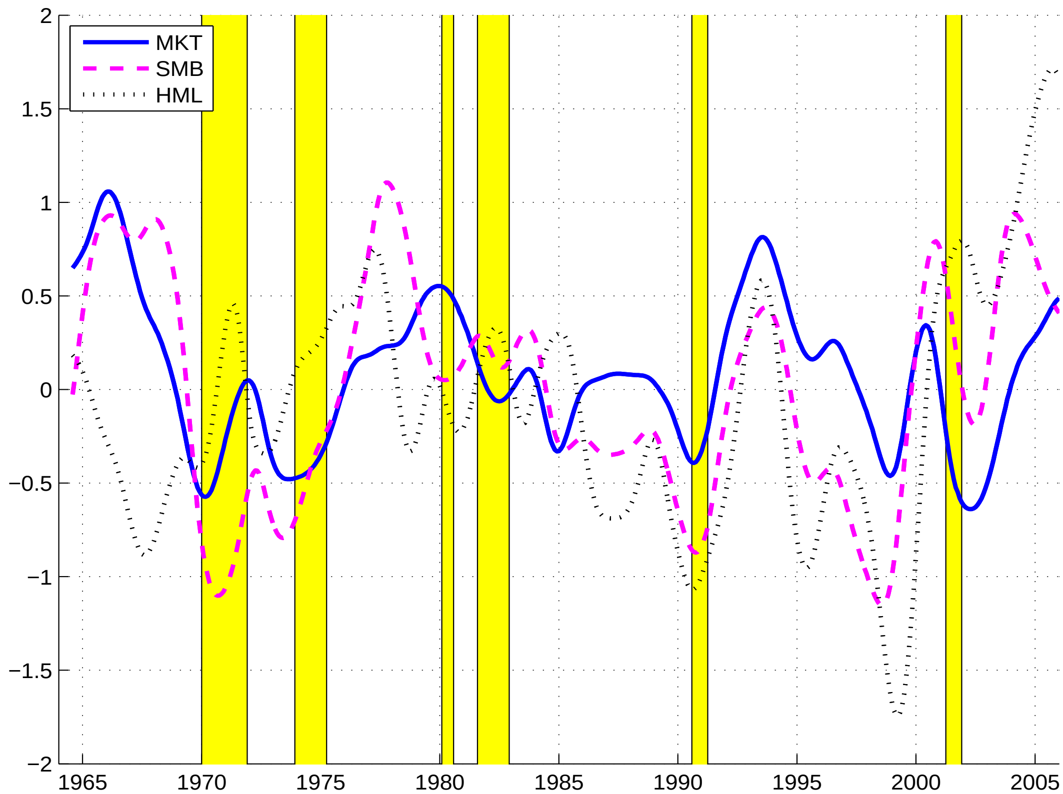
The figure shows monthly estimates of conditional Fama-French (1993) factor loadings of the book-to-market strategy, which goes long the 10th book-to-market decile portfolio and short the 1st book-to-market decile portfolio. We plot the optimal estimates in bold lines along with 95% confidence bands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