Journal of Banking & Finance 61 (2015) 117–126

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

# A semiparametric conditional capital asset pricing model ☆

Zongwu Cai [a,b], Yu Ren [b,*], Bingduo Yang [c]

[a] Department of Economics, University of Kansas, Lawrence, KS 66045, USA
[b] Wang Yanan Institute for Studies in Economics, MOE Key Lab of Econometrics, and Fujian Key Lab of Statistical Sciences, Xiamen University, Xiamen, Fujian 361005, China
[c] School of Finance, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China

ARTICLE INFO

*Article history:*
Received 29 October 2014
Accepted 7 September 2015
Available online 16 September 2015

*JEL classification:*
C13
C52
G12

*Keywords:*
Conditional capital asset pricing model
Functional coefficient regression
Smoothly clipped absolute deviation penalty
Variable selection

ABSTRACT

This paper proposes using a functional coefficient regression technique to estimate time-varying betas and alpha in the conditional capital asset pricing model (CAPM). Functional coefficient representation relaxes the strict assumptions regarding the structure of betas and alpha by combining the predictors into an index. Appropriate index variables are selected by applying the smoothly clipped absolute deviation penalty. In such a way, estimation and variable selection can be done simultaneously. Based on the empirical studies, the proposed model performs better than the alternatives in explaining asset returns and we find no strong evidence to reject the conditional CAPM.



## 1. Introduction

The capital asset pricing model (CAPM) plays a cornerstone role in theoretical and empirical finance. It states that a linear relationship exists between the excess return of a risky asset and the beta of that asset with respect to the market return. The betas in the CAPM are commonly assumed to be constant over time. However, recent empirical studies provide ample evidence against this assumption because the relative risk of firm's cash flow varies over the business cycle and the state of the economy; see, for example, Fama and French (1997), Ferson and Harvey (1997), Lettau and Ludvigson (2001), Zhang (2005), Lewellen and Nagel (2006) and the references therein. In other words, it is more reasonable to believe that the CAPM holds under the condition of current information sets, which leads to the conditional CAPM. In the conditional CAPM, the corresponding betas should be adjusted accordingly because information sets are updated over time. This implies that betas are time-varying. How to estimate the time-varying betas is of great importance because only when the betas are estimated appropriately, the pricing errors of the conditional CAPM are able to be measured correctly and the validity of the model can be evaluated. For this purpose, in this paper, we propose a new methodology to estimate the time-varying betas.

Estimation of the time-varying betas has already been discussed extensively through two different approaches in the finance literature. First, betas can be regarded as a function of time; for example, see the papers by Johnstone and Silverman (1997) and Robinson (1997). For parametric models in this approach, betas are assumed to be either a discrete function of time such as the threshold CAPM proposed by Akdeniz et al. (2003), or a continuous function like a smooth transition model developed by Lin and Teräsvirta (1994). For nonparametric models, betas are simply a nonparametric function of time, as described by Ang and Kristensen (2012) and Connor et al. (2012). This approach is criticized for hiding the economics driving force behind betas. It does not show how and why betas

☆ We thank the editor and the referees for their constructive and helpful comments. We also thank the seminar participants in UC Riverside, University of Oklahoma, National University of Singapore, Canadian Economics Association 2010 Meeting, Tsinghua Econometrics Workshop 2010 and Taipei FERM Meeting 2010. Cai's research was supported, in part, by the National Nature Science Foundation of China grants (#71131008 (Key Project), #70871003 and #70971113). Ren's research was supported by the Fundamental Research Funds for the Central Universities (#2013221022) and the Natural Science Foundation of China grants (#71301135, #71203189, #71131008). Yang's research was supported by the National Nature Science Foundation of China grant (#71401066) and Specialized Research Fund for the Doctoral Program of Higher Education (#20130161120023).
* Corresponding author at: Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China. Tel.: +86 592 2186025.
*E-mail address:* renxmu@gmail.com (Y. Ren).

http://dx.doi.org/10.1016/j.jbankfin.2015.09.002
0378-4266/

vary over time. The second approach is that betas are assumed to be affected by some variables. These variables can be the proxies of latent variables as pointed out by Ang and Chen (2007) or of some observable macro variables as in Ferson and Harvey (1999). It is clear that the latter approach can provide more economic intuition on the movement of betas than the former one. Because there is no overwhelming argument for either approach, this paper circumvents this debate and assumes that betas are functions of some observable variables. These variables are often called financial instruments.

There are two advantages of using financial instruments to track the movement of betas. The first advantage is that this method can reveal the close relationship between the relative risk of firm's cash flow as it varies over the business cycle and the state of the economy. As suggested by Aït-Sahalia and Brandt (2001), all of potential instrument variables can be combined into an index that best captures time variations in betas, and the index can be explained as an economic state variable. This index has a good interpretation as follows. From a statistical standpoint, the index avoids the curse of dimensionality because it allows us to reduce the multivariate problem to one. Therefore, we can implement the nonparametric approach (see Section 2 later) in a univariate setting; see Aït-Sahalia and Brandt (2001) for details. From an economic perspective, this index offers a convenient univariate summary statistic that describes the current state of the various time-varying economic indicators related to investment opportunities for portfolio investment. From a normative perspective, the index can help investors with any set of preferences to determine which economic variables they should track and, more importantly, in what single combination. The other advantage is that this approach considers not only the variation across averages of betas in each short time window but also the variation of the actual betas within each window. Indeed, Campbell and Vuolteenaho (2004), Fama and French (2005) and Lewellen and Nagel (2006), among others, assumed discrete changes in betas across sub-samples but constant betas within sub-samples.

However, there are still three pitfalls in this time-varying betas. First, there is a strict assumption about the relationship between the betas and the instrument variables. Ferson and Harvey (1999) imposed the assumption that the betas are linear functions of the index, whereas Wang (2002, 2003) found strong evidences against this assumption and argued that this strong assumption might lead to a model misspecification. As shown in Ghysels (1998), inference and estimation based on misspecification can be very misleading. In addition, Ghysels (1998) showed that among several well-known time-varying beta models, a serious misspecification might produce time variation in the beta that is highly volatile, and it might lead to large pricing errors. Thus, it is important to analyze the time-varying betas by relaxing the aforementioned assumptions. Second, as pointed out by Aït-Sahalia and Brandt (2001), it is often in the literature to choose instruments and to estimate model in two different model frameworks (under two different objective functions). In such a way, it might produce an inconsistent estimation or an inappropriate instrument selection. Therefore, to select instruments and best fit the model, the selection procedure should be conducted simultaneously with the estimation approach. Last, Harvey (2001) showed that the estimates of betas obtained using instrumental variables are very sensitive to the choice of instruments used as proxies for time-variation in the conditional betas.

To address the aforementioned issues, we propose using a functional coefficient regression (FCR) technique introduced by Cai et al. (2000) to estimate the time-varying betas and at the same time, adopting a penalty function to select the instrument variables. A FCR model estimates betas nonparametrically, by assuming that the coefficients of financial covariates are deterministic functions of some instrument variables. The estimates are obtained by using any nonparametric methods such as local linear fitting; see Fan and Gijbels (1996). Thus, a FCR model can relax the strong assumption of linearity and avoid model misspecification. The reader is referred to the survey paper by Cai and Hong (2009) on how to apply a FCR model in economics and finance. In addition, we can estimate betas and select instrument variables simultaneously by adding a penalty term. For example, we can choose the smoothly clipped absolute deviation penalty (SCAD) function introduced by Fan and Li (2001) although other penalty functions might be applicable. By doing so, the model estimation and variable selection can be implemented simultaneously so that important instrument variables are chosen automatically for the regression model. Thus, all of the potential candidates can be included into the model without examining whether a relationship exists between any individual instrument variable and the asset return. The main contribution of this paper is that a new FCR model with instrument variable selection is proposed from the conditional CAPM point of view. Also, the mathematical proofs for our model under time series settings are provided and indeed, they are different from the linear model set up in the paper by Fan and Li (2001) for the independent identically distributed (iid) sample. Moreover, the attractive point of this instruments selection for the FCR model is that it is not only critical for the conditional CAPM, but also flexible to be applied to other related economic and financial areas where the important variables should be selected from very large scale candidates.

In this paper, a discussion concerning about time-varying betas is under the framework of the conditional CAPM. However, the conditional CAPM, as an alternative to the static CAPM, is quite controversial in the finance literature; see Lewellen and Nagel (2006) for the detailed arguments. This means that it is difficult to choose the conditional CAPM or the static CAPM in a real application. Theoretically, the conditional CAPM can hold perfectly from period to period, and it can be regarded as a base for many other models; for example, the premium-labor model in Jagannathan and Wang (1996). Conversely, Ferson and Harvey (1999), Wang (2003) and Lewellen and Nagel (2006) found that the conditional CAPM is rejected by their empirical analysis, although their results might not be convincing. As for the aforementioned papers, the model in Ferson and Harvey (1999) may be misspecified; see Section 4 later for the detailed arguments. Wang (2003) used four instrument variables to perform nonparametric estimations and tests. However, the number of observations in his data might not be large enough to produce reliable inferences. Finally, Lewellen and Nagel (2006) did a simple comparison to evaluate whether the variation in the betas and the equity premium is large enough to explain important asset-pricing anomalies. However, the size of their test method is challenged by Li et al. (2015). Thus, the validity of choosing the conditional CAPM over the static CAPM is still an open research topic.

Recently, Ferreira et al. (2011) proposed a nonparametric two-stage estimator for conditional beta pricing models by allowing for flexibility not only in the betas but also in the risk premium. Our method can be considered as a generalization of the first stage estimation in Ferreira et al. (2011) with the advantage that there is no need to select the instrumental variables in advance. In addition, Ferreira et al. (2011) used multivariate kernel to estimate the beta, and their method suffers from the curse of dimensionality. For this reason, in their empirical analysis, they only used two instrument variables. While, our model can alleviate this problem in the estimation. Connor et al. (2012) developed a characteristic-based weighted additive regression for the factor model. Due to the curse of dimensionality, univariate nonparametric functions are considered as the characteristic-betas in the paper. In their empirical analysis, four characteristics including size, value, momentum and own-volatility are used to estimate the additive nonparametric characteristic-based functions respectively. Ang and Kristensen (2012) proposed a nonparametric method for estimating and testing conditional alphas and betas and long-run alphas and betas. In their paper, the alphas and betas are assumed to be deterministic functions of time $t$.

To evaluate the conditional CAPM fairly, there are many aspects to consider. Due to the infeasibility to analyze all of them, our focus here is only on the following two points in this paper. The first is to see whether the data support the time-varying betas and the second is to exam whether the pricing error, the alpha, is statistically insignificant. We do test on these two phenomena based on our model. By applying the well known Fama–French data sets, we find that we can not reject the hypotheses that the betas are time varying and that the alpha is statistically insignificant. Therefore, at this stage, we do not find any strong evidence against the conditional CAPM.

The rest of this paper is organized as follows. Section 2 is devoted to the descriptions of our estimation model and instrument selection as well as the related econometric issues. Section 3 presents some simulation results to demonstrate the finite sample performance of the proposed model, and Section 4 reports the empirical analysis of 25 portfolios based on Fama–French data sets. Finally, Section 5 concludes the paper. All the mathematical proofs are relegated to the Appendices.

## 2. Econometric model

### 2.1. Model and estimation procedure

We consider the model

$$y_i = g^T(c^T Z_i)X_i + \varepsilon_i, \quad 1 \leqslant i \leqslant n, \tag{1}$$

where $y_i$ is a dependent variable, $X_i = (1, x_i)^T$, $x_i$ is a factor, $Z_i$ is a $d \times 1$ vector of local variables, $\{\varepsilon_i\}$ are iid with mean 0 and standard deviation $\sigma, c \in R^d$ is a $d \times 1$ vector of unknown parameters and $g(\cdot) = \left(g_1(\cdot), g_2(\cdot)^T\right)^T$ is a vector of 2-dimensional unknown functional coefficients. In the context of the conditional CAPM, $g_1(\cdot)$ is the pricing error, or $\alpha(\cdot)$; $g_2(\cdot)$ is the factor loading, or $\beta(\cdot)$. Namely, $\alpha(\cdot) = g_1(\cdot)$ and $\beta(\cdot) = g_2(\cdot)$. We assume that $||c|| = 1$ and the first element of $c$ is positive for identification. Note that for simplicity, $x_i$ is assumed to be a single factor. Extension to multiple factors is straightforward and all modeling procedures and theories continue to hold. Further, the residuals $\{\varepsilon_i\}$ can be extended to heteroscedasticity case in which the estimators are still consistent if the sample size is large.

One may consider more general set up with $Y_i = g(X_i, Z_i)$ and estimate the function by a non-parametric (or semi-parametric) method. However, by following the CAPM model and its related literature, we consider this FCCAPM form in this paper since this setting may cover several other existing conditional CAPM models (Johnstone and Silverman, 1997; Robinson, 1997; Ferson and Harvey, 1999; Ang and Kristensen, 2012; Connor et al., 2012) as a special case. Meanwhile, it can overcome the difficulty of the so-called curse of dimensionality if there are many conditional variable $Z$'s."

The functional coefficient capital asset pricing model (FCCAPM) is given by

$$y_i = g_1(z_i) + g_2(z_i)x_i + \varepsilon_i, \tag{2}$$

where $z_i = c^T Z_i$. By premultiplying $X_i$ on (2) and taking $E(\cdot|c^T Z_i)$ on both sides, it leads to $g(c^T Z_i) = \left[E(X_i X_i'|c^T Z_i)\right]^{-1} E(X_i y_i'|c^T Z_i)$, which can be considered as an extension of Eqs. (7)–(10) in the paper by Wang (2003).

To estimate $g_1(\cdot)$ and $g_2(\cdot)$ in (2), one can use a local linear fitting scheme (see Fan and Gijbels, 1996). For a given grid point $z_0$, we can approximate $g_1(z)$ and $g_2(z)$ locally by a linear[1] function $g_1(z) \approx a_0 + a_1(z - z_0)$ and $g_2(z) \approx b_0 + b_1(z - z_0)$, respectively, when $z$ is in a neighborhood of $z_0$. The local linear estimators of $g_1(z_0)$ and $g_2(z_0)$ are defined as $\widehat{g}_1(z_0) = \widehat{a}_0$ and $\widehat{g}_2(z_0) = \widehat{b}_0$, where $\{\widehat{a}_j, \widehat{b}_j\}$ $(j = 0, 1)$ minimize the following locally weighted least squares

$$\sum_{i=1}^{n}\{y_i - [a_0 + a_1(z_i - z_0)] - [b_0 + b_1(z_i - z_0)]x_i\}^2 K_h(z_i - z_0), \tag{3}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function on $\mathbb{R}^1$ and $h$ is a bandwidth, which controls the amount of smoothing used in the estimation. Next, we re-write the minimization problem (3) in a matrix form. Let $B = (a_0, b_0, a_1, b_1)'$, $Y = (y_1, y_2, \ldots, y_n)'$,

$$U = \begin{pmatrix} 1 & x_1 & (z_1 - z_0) & (z_1 - z_0)x_1 \\ 1 & x_2 & (z_2 - z_0) & (z_2 - z_0)x_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & (z_n - z_0) & (z_n - z_0)x_n \end{pmatrix}, \quad \text{and} \quad W = \text{diag}\{K_h(z_i - z_0)\}.$$

Then, the minimization problem can be transformed to $\min_B (Y - UB)'W(Y - UB)$ and the solution is

$$\widehat{B} = (U'WU)^{-1}U'WY, \tag{4}$$

which is the well known weighted least squares estimate. Clearly, (4) provides a formula for computational implementation, which can be carried out by any standard statistical package. When $z_0$ moves over the domain of $z_i$, the estimated curves of $\alpha(\cdot)$ and $\beta(\cdot)$ are obtained. In the practical implementation, $z_0$ can be taken any value in the domain of $z_i$.

To do the estimation and variable selection simultaneously, we implement the procedures as follows. Given $\widehat{g}(\cdot)$, we minimize the penalized global least squares $Q(c, \widehat{g})$ (or maximize the penalized global likelihood), where

$$Q(c, \widehat{g}) = \frac{1}{2}\sum_{i=1}^{n}\left(y_i - \widehat{g}^T(c^T Z_i)X_i\right)^2 + n\sum_{k=1}^{d} P(|c_k|) \tag{5}$$

with $P(\cdot)$ being a penalty function. Here, an initial estimator $\widehat{g}(\cdot)$ can be obtained by various algorithms such as the method in Fan et al. (2003).

As for the penalty function, we choose the so-called smoothly clipped absolute deviation (SCAD) penalty, where the first order derivative of $P_{\lambda,\upsilon}(\cdot)$ is defined as

$$P'_{\lambda,\upsilon}(|c_k|) = \lambda I(|c_k| \leqslant \lambda) + \frac{(\upsilon\lambda - |c_k|)_+}{\upsilon - 1} I(|c_k| > \lambda), \tag{6}$$

and $\lambda$ and $\upsilon$ are two tuning parameters. By choosing two optimal tuning parameters $\lambda$ and $\upsilon$, the resulting estimates from (5) can be obtained in terms of out-of-sample MSE performance. Note that other penalty functions can be used in (5); see Fan and Lv (2009) for more discussions on the choice of various penalty functions. As shown in Fan and Li (2001), the SCAD penalty function leads to the estimators with three desired properties that can not be achieved by either the $L_p$ penalty function or the hard penalty function. The three properties are as follows: unbiasedness for the non-zero coefficient to avoid unnecessary estimation bias, sparsity for estimating a coefficient as small as 0 to reduce model complexity, and continuity of the resulting estimator to avoid unnecessary variation in model prediction. More discussions regarding these properties can be found in the paper by Fan and Li (2001).

[1] A polynomial approach is applicable; see Fan and Gijbels, 1996 for details.

To make our method more adaptive to big data paradigm, we may use the sure independence screening (Fan and Lv, 2008). The first is to squeeze down the size of instrumental variables $Z$ by doing regression on functional coefficient model with each variable $Z$, respectively. Then we apply SCAD to select the instrumental variables.

In view of (6), the choices of $\lambda$ and $\upsilon$ are important in the practical implementation. Here, we follow the idea from Fan and Li (2001) and use the following multi-fold cross-validation. Denote the full data set by $\mathcal{T}$, and denote cross-validation training and test set by $\mathcal{T}-\mathcal{T}^q$ and $\mathcal{T}^q$, respectively, for $q=1,\ldots,Q$. For every $\lambda,\upsilon$ and $q$, we get the estimators $\widehat{g}_1(\cdot)$ and $\widehat{g}_2(\cdot)$ by using $\mathcal{T}-\mathcal{T}^q$ training set. Then, we choose $\lambda$ and $\upsilon$ to minimize

$$\mathrm{CV}(\lambda,\upsilon)=\sum_{q=1}^{Q}\sum_{(y_q,x_q,z_q)\in\mathcal{T}^q}\left[y_q-\widehat{g}_1(z_q)-\widehat{g}_2(z_q)x_q\right]^2.$$

As suggested by Fan and Li (2001), we choose $Q=5$.

Intuitively, if instrument variables are not helpful in explaining the time-varying betas, the estimated coefficients for these instruments are shrunk exactly to zero. Hence, the instruments with the estimated zero value coefficients are not included in formatting the state variable. This variable selection method is entirely data-driven and reveals the potential effects of instrument variables, while it maintains simultaneously the analysis within the framework of FCCAPM. Furthermore, because the number of the instrument variables are no longer restricted, the SCAD method removes automatically the "useless" instruments and selects the significant instruments.

### 2.2. Asymptotic properties

Let $\{(X_i,Z_i,y_i)\}$ be a strictly stationary and strong mixing sequence, $f(z,c)$ be the density function of $z=c^TZ$, where $c$ is an interior point of the compact set $\mathcal{C}$. Define $\mathcal{A}_z=\{Z: f(Z,c)\geqslant\Lambda,\forall c\in\mathcal{C}\}$, where $\Lambda$ is a small positive constant. For $c$ an interior point of the compact set $\mathcal{C}$, define penalized least squares

$$Q(c,\widehat{g})=\frac{1}{2}\sum_{i=1}^{n}\left(y_i-\widehat{g}^T(c^TZ_i)X_i\right)^2+n\sum_{k=1}^{d}P_{\lambda_n}(|c_k|). \tag{7}$$

We assume the first $d_1$ coefficients of $c$ are nonzero, and all rest of parameters are zero, i.e., $c_0=(c_{10}^T,c_{20}^T)^T$, all elements of $c_{10}$ with dimension $d_1$ are nonzero, and $d-d_1$ dimensional coefficients $c_{20}=0$. Finally, define $V_n=\sum_{i=1}^n(Z_i-E(Z_i|c_0^TZ_i))\dot{g}^T(c_0^TZ_i)X_i\varepsilon_i$, where $\dot{g}(\cdot)$ is the first derivative of function $g(\cdot)$ vector, and $\varepsilon_i$ is iid with mean 0 and standard deviation $\sigma$. Let $\widetilde{V}_0=\frac{1}{n}Var(V_n)/\sigma^2$, and define $\mathbf{e}$ be an asymptotically standard normal random $d$-dimensional vector such that $V_n=n^{1/2}\sigma\widetilde{V}_0^{1/2}e$. $V_{1n}=\sum_{i=1}^n(Z_{1i}-E(Z_{1i}|c_{10}^TZ_{1i}))\dot{g}^T(c_{10}^TZ_{1i})X_i\varepsilon_{1i}$, where $\varepsilon_{1i}$ is the same as $\varepsilon_i$ since $c_{20}=0$. Similarly, we define $\widetilde{V}_{10}=\frac{1}{n}Var(V_{1n})/\sigma^2$ and $\mathbf{e}_1$ be an asymptotically standard normal random $d_1$-dimensional vector such that $V_{1n}=n^{1/2}\sigma\widetilde{V}_{10}^{1/2}e_1$.

To study the asymptotic distribution of the penalized least squares estimator $\widehat{c}$, we impose some technical conditions as below.

(A1) The vector functions $g(\cdot)$ have continuous second order derivatives with respect to the support of $\mathcal{A}_z$.
(A2) For any $c\in\mathcal{C}$ and $Z\in\mathcal{A}_z$, the density function $f(\cdot,c)$ is continuous and there exists a small positive constant $\Lambda$ such that $f(\cdot,c)>\Lambda$.
(A3) The kernel function $K(\cdot)$ is a bounded density with a bounded support region. Let $\mu_2=\int\upsilon^2K(\upsilon)d\upsilon$ and $\nu_0=\int K^2(\upsilon)d\upsilon$.
(A4) $\lim_{n\to\infty}\inf_{\theta\to0^+}P'_{\lambda_n}(\theta)/\lambda_n>0,\lambda_n\to0,\quad\sqrt{n}\lambda_n\to\infty$ and $h=O(n^{-1/5})$.
(A5) $\{(X_i,Z_i,y_i)\}$ is a strictly stationary and strong mixing sequence with mixing coefficient satisfying $\alpha(m)=O(\rho^m)$ for some $0<\rho<1$.
(A6) $E(\varepsilon_i|X_i,Z_i)=0, E(\varepsilon_i^2|X_i,Z_i)=\sigma^2, E|X_i|^m<\infty$ and $E|y_i|^m<\infty$ for all $m>0$.

**Remark 1.** The second order differentiability of vector functions $g(\cdot)$ in A1 and kernel function $K(\cdot)$ in A3 lead to that the order of bias term for nonparametric estimator is $O_p(h^2)$. These assumptions are standard for a nonparametric method. The assumptions in A4 indicate the oracle property in Theorem 2. An alternative condition for bandwidth in Ichimura (1993) is $nh^8\to0$. However, the condition $nh^8\to0$ is still satisfied with our condition $h=O(n^{-1/5})$ in A4. Assumptions in A5 are the common conditions with weak dependent data. Most financial models satisfy these conditions, such as ARCH and GARCH models; see Cai, 2002. For Assumption A6, it is not hard to extend to the heteroscedasticity case, $E(\varepsilon_i^2|X_i,Z_i)=\sigma^2(X_i,Z_i)$, and it requires some higher moment conditions of $X_i$ and $y_i$ so that Chebyshev inequality can be applied.

**Theorem 1.** *Let $\{(X_i,Z_i,y_i)\}$ be a strictly stationary and strong mixing sequence, $a_n=\max\{P'_{\lambda_n}(|c_k|):c_k\neq0\}$, and $\widehat{c}=\mathrm{argmin}_{c\in\mathcal{C}}Q(c,\widehat{g})$. Under Assumptions A1–A6 and if $\max\{P''_{\lambda_n}(|c_k|):c_k\neq0\}\to0$, then the order of $\|\widehat{c}-c_0\|$ is $O_p(n^{-1/2}+a_n)$. If the penalty function is SCAD, $a_n=0$ as sample size $n\to\infty$, and $\|\widehat{c}-c_0\|=O_p(n^{-1/2})$.*

**Theorem 2** (*Oracle property*). *Let $\{(X_i,Z_i,y_i)\}$ be a strictly stationary and strong mixing sequence. Under Assumptions A1–A6, by assuming $\lambda_n\to0$ and $\sqrt{n}\lambda_n\to\infty$ as $n\to\infty$, then,*
(a) *Sparsity:*

$\widehat{c}_2=0.$

(b) *Asymptotic normality:*

$$\sqrt{n}(\widehat{c}_1-c_{10})\to N\left(0,\widetilde{V}_{10}^{-1}V_{10}\widetilde{V}_{10}^{-1}\right),$$

*where $\widetilde{V}_{10}$ is defined earlier and $V_{10}=\Gamma(0)+2\sum_{\ell=1}^{\infty}\Gamma(\ell)$ with $\Gamma(\ell)=\mathrm{Cov}(\Gamma_i,\Gamma_{i-\ell})$ and $\Gamma_i=(Z_{1i}-E(Z_{1i}|c_{10}^TZ_{1i}))\dot{g}^T(c_{10}^TZ_{1i})X_i\varepsilon_{1i}$.*

When the random variables $\{\Gamma_i\}_{i=1}^{\infty}$ are either iid or martingale difference sequence, $V_{10}$ becomes $V_{10}=\Gamma(0)=\mathrm{Var}(\Gamma_i)$. Otherwise, the autocovariance function $\Gamma(\ell)$ may not be zero at least for some lag orders $\ell>0$ due to the serial correlation. Theorem 2 shows that our variable selection procedures of minimizing penalized least squares enjoy the oracle property.

### 2.3. Bandwidth selection

It is well known that the bandwidth is important in nonparametric estimation since it can balance the trade-off between the bias and the variance of the nonparametric estimates. There are several bandwidth selection methods available in the literature, here we prefer the optimal bandwidth $h_{opt}$ by minimizing the following nonparametric version of bias-corrected Akaike information criterion (AIC) due to its simplicity (Cai and Tiwari, 2000; Cai, 2002),

$$\mathrm{AIC}(h)=\log(\widehat{\sigma}^2)+2(n_h+1)/(n-n_h-2), \tag{8}$$

where $\widehat{\sigma}^2=\sum_{i=1}^n(\widehat{y}_i-y_i)^2/n$ and $n_h$ is the trace of the hat matrix $H_h$ which makes $\widehat{Y}=H_hY$.

This selection criterion counteracts the over/under-fitting tendency of the generalized cross-validation and the classical AIC; see Cai and Tiwari (2000) and Cai (2002) for more details. However, the rigorous theoretical properties for the optimality of this bias-corrected AIC selector need further research and they can be regarded as the future research topics. Alternatively, one might use some other existing methods in the time series literature although they may require more computing (Fan and Gijbels, 1996; Cai et al., 2000).

### 2.4. Testing the conditional CAPM

In order to check the validity of the conditional CAPM, we have to test the significance of the pricing error. In the conditional CAPM given in (2), the pricing error is denoted as $g_1(\cdot)$ in the model. Hence, the testing problem can be formulated as

$$H_0: g_1(z) = 0, \quad \text{versus} \quad H_1: g_1(z) \neq 0 \quad \text{forall } z. \tag{9}$$

The generalized $F$-type test statistic proposed by Cai and Tiwari (2000) can be applied here and is defined as

$$J_n = \mathrm{RSS}_0/\mathrm{RSS}_1 - 1. \tag{10}$$

Here, the sum square residuals (RSS) under the null hypothesis is

$$\mathrm{RSS}_0 = n^{-1}\sum_{i=1}^{n}[y_i - \tilde{g}_2(z_i)x_i]^2,$$

where $\tilde{g}_2(\cdot)$ is estimated under the null, and the RSS under the alternative is

$$\mathrm{RSS}_1 = n^{-1}\sum_{i=1}^{n}\left[y_i - \widehat{g}_1(z_i) - \widehat{g}_2(z_i)x_i\right]^2,$$

where $\widehat{g}_1(\cdot)$ and $\widehat{g}_2(\cdot)$ are estimated under the alternative.

For computational simplicity, the following nonparametric bootstrap approach is used to obtain the $p$-value of the statistic $J_n$ given in (10):

1. Collect the residuals $\{\tilde{e}_i\}$ by $\tilde{e}_i = y_i - \tilde{g}_2(z_i)x_i$.
2. Generate the bootstrap residuals $\{e_i^*\}$ from the empirical distribution of the centered residuals $\{\tilde{e}_i - \bar{\tilde{e}}_i\}$.
3. Define the bootstrap sample as $y_i^* = \tilde{g}_2(z_i)x_i + e_i^*$.
4. Calculate the bootstrap test statistic $J_n^*$ based on the sample $\{y_i^*, x_i, z_i\}$.
5. Compute the $p$-value of the test based on the relative frequency of the event $\{J_n^* \geqslant J_n\}$ in the replications of the bootstrap sampling.

Note that the validity of this bootstrap can be found in Cai et al. (2000) and Kreiss et al. (2008).

It is clear that the hypothesis testing formulation given in (9) can easily generalized to test if the model proposed in Ferson and Harvey (1999) is appropriate. That is to test if both $g_1(z)$ and $g_2(z)$ in (2) are linear. Then, (9) becomes to

$$\begin{aligned} &H_0: \text{both } g_1(z) \text{ and } g_2(z) \text{ are linear versus} \\ &H_1: \text{at least one is not linear.} \end{aligned} \tag{11}$$

Similarly, we can define the test statistic for (11). Other types of testing problems can be formulated in the same fashion.

## 3. Monte Carlo studies

In this section, we illustrate the proposed modeling methods using simulated data sets. This data set mimics the actual portfolio returns and market returns in the conditional CAPM model. In addition, we generate three instrument variables but assume that only two of them formulate the state of the economy. We then determine whether our model can deliver consistent estimates in terms of the mean absolute deviation (MAD) and determine the actual instrument variables in the model.

In our simulations, the optimal tuning parameters are chosen by the fivefold cross-validation as in Section 2.1 and the bandwidth is selected by AIC as in Section 2.3. The Epanechnikov kernel $K(x) = 0.75(1-x^2)$ if $|x| \leqslant 1$ is used. Because the key point to demonstrate the validity of the conditional CAPM is to test whether the pricing error is significant, the bootstrap testing procedure outlined in Section 2.4 should have the appropriate size and indeed is powerful to make correct inferences. Thus, based on the simulated data, we also check the size and the power of our testing procedure given in Section 2.4.

### 3.1. Linear model

We generate $n = 300$, $n = 500$ and $n = 1000$ data points, respectively. $Z_{1i}, Z_{2i}$ and $Z_{3i}$ are drawn from a normal distribution with mean zero and standard deviation of 0.38 as our instrument variables.[2] $x_i$ is generated from a uniform distribution on $[0, 1]$. The data generating process (DGP) for $y_i$ is

$$y_i = (c_1 Z_{1i} + c_2 Z_{2i} + c_3 Z_{3i})x_i + e_i, \tag{12}$$

where $e_i$ is an error term distributed normally with mean zero and standard deviation of 0.08, and $c_1 = c_2 = \sqrt{2}/2$ but $c_3 = 0$. Although only two instruments are involved in DGP in (12), we consider all three instrument variables in our estimation procedure. We expect our model to account for this fact and decrease the effect of the irrelevant instruments somehow. We repeat the simulation 1000 times, and report the estimates of $c_1$, $c_2$ and $c_3$ in terms of their mean absolute deviation (MAD) and the shrinkage rate for $\widehat{c}_3$, defined as the number of $\{\widehat{c}_3 \neq 0\}/1000$, respectively. The median and the standard deviation of the absolute deviation of the estimators are showed in the column of linear model listed in Table 1.

It is surprising that $\widehat{c}_3$ shrinks to 0 in all 1000 simulations. This indicates that our model can delete the third instrument, which is not included in the true DGP, from the instrument variables automatically. In addition, the median and standard deviation of 1000 MAD values for the estimates of $c_1$ and $c_2$ decrease for all settings when the sample size $n$ increases. It implies that the estimates of $c_1$ and $c_2$ are indeed consistent. This finding is consistent with what our theoretical model indicates.

### 3.2. Nonlinear Model

We use the same settings to simulate $Z_{1i}, Z_{2i}, Z_{3i}$, $x_i$ and $e_i$, whereas $y_i$ is generated nonlinearly as follows

$$y_i = 0.1\exp(c_1 Z_{1i} + c_2 Z_{2i} + c_3 Z_{3i})x_i + e_i. \tag{13}$$

Similarly, $c_1 = c_2 = \sqrt{2}/2$ but $c_3 = 0$. The estimation results are reported in the column of nonlinear model listed in Table 1.

As expected, the estimates of $c_1$ and $c_2$ are consistent and the estimated $c_3$ can be shrunk to zero when the sample size is large. Also, one can see that for this setting, the shrinkage rate of the estimated $c_3$ is not very high when $n = 300$ and it is slightly lower than that for the linear model. This phenomenon is not surprising because the performance of our model depends on the complexity of the true DGP. Therefore, in the overall, the performance of our model works fairly well.

[2] We also simulate $Z_{1i}, Z_{2i}$ and $Z_{3i}$ by time series models in this and the following experiments. The results are similar. In order to save space, we do not report them in this paper. They are available upon request.

**Table 1**
Simulation results: mean absolute deviation and shrinkage rate. We generate three instruments, and only use the first two to form the instrument variables. The excess returns are simulated linearly in (12) and nonlinearly in (13), respectively. The median and standard deviation for the absolute deviation of the estimates are reported based on 1000 simulations. $c_1, c_2$ and $c_3$ are coefficients of three instruments. "Shrinkage rate" indicates the relative frequency that the estimator of $c_3$ equals exactly to 0.

| | Linear model | | | Nonlinear model | | |
|---|---|---|---|---|---|---|
| | $c_1$ | $c_2$ | $c_3$ | $c_1$ | $c_2$ | $c_3$ |
| $n = 300$ | | | | | | |
| Median | 0.0022 | 0.0022 | | 0.0180 | 0.0160 | |
| Std | 0.0017 | 0.0017 | | 0.0186 | 0.0196 | |
| Shrinkage rate | | | 100% | | | 80% |
| $n = 500$ | | | | | | |
| Median | 0.0019 | 0.0019 | | 0.0112 | 0.0114 | |
| Std | 0.0010 | 0.0010 | | 0.0133 | 0.0143 | |
| Shrinkage rate | | | 100% | | | 95% |
| $n = 1000$ | | | | | | |
| Median | 0.0014 | 0.0014 | | 0.0106 | 0.0107 | |
| Std | 0.0010 | 0.0010 | | 0.0059 | 0.0073 | |
| Shrinkage rate | | | 100% | | | 98% |

### 3.3. Size and power of test

Because we use the bootstrap method to test the significance of the pricing error, we need to check if the proposed test has a right test size and is powerful. If the size or power distortion is large, then the inferences based on the test results are not reliable.

We still use the DGPs described previously as the representatives of the linear and the nonlinear models for the conditional CAPM. We simulate $n = 100, n = 300$ and $n = 500$ data points, respectively and test the null hypothesis that the intercept function is zero. Then we repeat the entire procedure 1000 times and report the frequency of the rejections of the null hypothesis. Because the true DGP is based on the conditional CAPM, the rejection frequency should be close to the nominal level. These results are summarized in the column of size of test listed in Table 2, and we can conclude that the proposed test can give a right test size.

In order to check the power of the test, we consider the so-called local alternative so that we add a local time-varying intercept function to the linear and the nonlinear models described before. The reason of considering the local alternative is to remove the effect of increasing sample size. The local alternative is

$$y_i = \frac{c_0}{n^{2/5}}(c_1 Z_{1i} + c_2 Z_{2i}) + (c_1 Z_{1i} + c_2 Z_{2i}) x_i + e_i$$

for the linear model and

$$y_i = \frac{c_0}{n^{2/5}}(c_1 Z_{1i} + c_2 Z_{2i}) + 0.1 \exp(c_1 Z_{1i} + c_2 Z_{2i}) x_i + e_i$$

for the nonlinear model.[3] We obtain the rejection frequencies based on 1000 replications for various values of $c_0$ which measures the degree of the alternative hypothesis departing from the null hypothesis. We consider different values of $c_0$ and $n$ by setting the significant level at 5%. The rejection frequencies are reported in the column of power of test listed in Table 2.

One can find that the power of our test increases sharply with $c_0$ increasing for all sample sizes and it reaches almost 100% when $c_0 = 0.5$. As expected, powers for all sizes are almost same due to the local alternative. Therefore, the power of our test is warranted to be useful in empirical applications.

[3] The convergence rate in the local alternative is based on the convergence rate for nonparametric estimates of $\alpha(\cdot)$ and $\beta(\cdot)$.

**Table 2**
Size and power of test. We use the linear and the nonlinear models with a local pricing error to generate the data, and test the null hypothesis that the pricing error is insignificant. The rejection rates with nominal level 1%, 5% and 10% are reported in the column of size of test. The rejection rates with six different values of $c_0$ under $\alpha = 5\%$ are showed in the column of power of test.

| | Size of test | | | Power of test | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1% | 5% | 10% | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.5 |
| $n = 100$ | | | | | | | | | |
| Linear model (%) | 0.3 | 3.3 | 7.6 | 10.5 | 29 | 43.5 | 63.5 | 79.5 | 98.0 |
| Nonlinear model (%) | 1.3 | 3.7 | 9.2 | 11.0 | 25 | 42 | 62 | 83.5 | 97.5 |
| $n = 300$ | | | | | | | | | |
| Linear model (%) | 0.7 | 3.9 | 8.2 | 22.0 | 45.5 | 70.5 | 86.5 | 95.5 | 99.5 |
| Nonlinear model (%) | 0.9 | 5.7 | 10.9 | 17.0 | 46.5 | 70.5 | 86.5 | 96.0 | 100 |
| $n = 500$ | | | | | | | | | |
| Linear model (%) | 1.2 | 4.6 | 9.8 | 26.5 | 49 | 76 | 91.5 | 97.5 | 100 |
| Nonlinear model (%) | 1.1 | 4.8 | 9.5 | 25.5 | 50 | 76 | 95.5 | 99 | 100 |

## 4. Empirical analysis

### 4.1. Data

We collect monthly returns of the Fama–French 25 portfolios from July 1963 to December 2009. The financial instruments, following Ferson and Harvey (1999), are the spread between the returns of the three-month and the one-month Treasury bill, the spread between Moody's Baa and Aaa corporate bond yield, the spread between a ten-year and one-year Treasury bond yields and the one-month Treasury bill yield. To match the model, we obtain the one-month lagged data for the instrument variable.

We do not claim that these instrument variables are the set of all potential instruments. These are the most popular ones used in practice because interest rates and spreads are usually benchmark indexes for the business cycle. It is believed in the literature that they are major forces to drive betas to be time-varying. By putting them into our model, we can estimate which variable plays a more important role than the others. Certainly, it is feasible to incorporate more macro variables into the model because the SCAD penalty criterion can shrink the coefficients of those "useless" variables to zero. By doing so. it will not hurt our estimation, but only needs more computing time.

### 4.2. Data smoothing

We use our model to model the relationship between the instrument variables and the asset returns. To compare the relative performances, we also employ the model in Ferson and Harvey (1999) (denoted by FH model) to the data. We report the mean square error (MSE) in Table 3.

The first column demonstrates that the portfolios for studies are divided into 25 groups which are sorted by the orders of the size ("S") and the book-to-market ratio ("B"). "S1" (and "B1") denotes the lowest order and "S5" (and "B5") represents the highest order. The mean square errors estimated in both our model and FH model are reported in the second and the third columns, respectively. We can see that the numbers in the second column are always smaller than those in the third column for all portfolios. The improvement of the MSE can be as large as 12% shown in portfolio "S2/B3". For the others, our model can decrease the MSE for approximately 7% on average. The relatively poor performance of the FH model may be explained by the model's misspecification. The linearity assumption imposed on the relationship between the instrument variables and the time-varying betas seems too strong. The fitting performance increases when this assumption is relaxed, which is consistent with the findings in Wang (2002, 2003).

**Table 3**
Mean square error. The instruments are the spread between the returns of a three-month and a one-month Treasury bill (r3m-r1m), the spread between Moody's Baa and Aaa corporate bond yield (BmA), the spread between a ten-year and one-year Treasury bond yield (r10y-r1y) and the one-month Treasury bill yield (r1m). In order to compare the relative performances, we also employ the model in v (FH model) to fit the data. The first column demonstrates that the portfolios for studies are divided into 25 groups which are sorted by the orders of the size ("S") and the book-to-market ratio ("B"). "S1" (and "B1") denotes the lowest order and "S5" (and "B5") represents the highest order. The second column reports the MSE when we use FCCAPM with SCAD variable selection to find the state variables and the third column reports the MSE by FH model. The MSE delivered by FCCAPM using another two state variables are reported in the fourth and the fifth columns, respectively. The coefficients of four instrument variables estimated by our SCAD method are reported the second panel. We also use the rolling 360-month to estimate the coefficients of the instruments, and then use the nonparametric method to estimate the portfolio return in the next period. The corresponding out of sample MSE are summarized in the last two columns.

| | In sample | | | | | | | | Out of sample MSE | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MSE | | | | Parametric coefficients | | | | | |
| | FCCAPM +SCAD | FH | FCCAPM +OLS | FCCAPM +FH | BmA | r3m | r10y | r1m | FCCAPM +SCAD | FH |
| S1/B1 | 22.49 | 23.33 | 23.43 | 23.15 | 0.30 | 0.52 | 0.80 | 0 | 34.91 | 35.96 |
| S1/B2 | 17.52 | 17.72 | 17.45 | 17.62 | 0.55 | 0.59 | 0.58 | 0 | 25.58 | 26.80 |
| S1/B3 | 10.77 | 12.04 | 11.73 | 11.63 | 0.66 | 0.71 | 0.22 | 0 | 14.01 | 15.86 |
| S1/B4 | 10.61 | 11.42 | 11.14 | 11.12 | 0.60 | 0.68 | 0.33 | 0.21 | 12.65 | 14.63 |
| S1/B5 | 12.33 | 13.40 | 13.45 | 13.36 | 0.61 | 0.75 | 0.18 | 0.11 | 13.45 | 15.44 |
| S2/B1 | 12.50 | 13.36 | 12.97 | 13.16 | 0.29 | 0.52 | 0.80 | 0 | 18.73 | 19.99 |
| S2/B2 | 8.19 | 8.62 | 8.31 | 8.46 | 0.68 | 0.71 | 0.16 | 0 | 10.56 | 11.69 |
| S2/B3 | 6.44 | 7.38 | 7.26 | 7.20 | 0.48 | 0.69 | 0.44 | 0.29 | 9.03 | 9.77 |
| S2/B4 | 7.35 | 7.42 | 7.26 | 7.29 | 0.22 | 0.97 | 0 | 0 | 9.83 | 11.19 |
| S2/B5 | 9.83 | 10.83 | 10.67 | 10.66 | 0.52 | 0.56 | 0.39 | 0.50 | 14.23 | 16.04 |
| S3/B1 | 8.94 | 9.21 | 8.89 | 9.01 | 0.29 | 0.51 | 0.80 | 0 | 13.86 | 14.95 |
| S3/B2 | 4.98 | 5.11 | 4.91 | 4.97 | 0.73 | 0.45 | 0.37 | 0.33 | 6.78 | 7.17 |
| S3/B3 | 4.93 | 5.24 | 5.08 | 5.07 | 0.47 | 0.51 | 0.44 | 0.56 | 6.82 | 7.08 |
| S3/B4 | 5.37 | 5.81 | 5.66 | 5.59 | 0.30 | 0.66 | 0.54 | 0.41 | 8.41 | 9.16 |
| S3/B5 | 8.13 | 9.23 | 9.01 | 8.79 | 0.57 | 0.54 | 0.34 | 0.50 | 10.28 | 12.46 |
| S4/B1 | 4.82 | 5.18 | 4.98 | 5.02 | 0.60 | 0.21 | 0.46 | 0.61 | 7.75 | 8.30 |
| S4/B2 | 3.39 | 3.51 | 3.52 | 3.48 | 0.18 | 0.51 | 0.66 | 0.50 | 5.33 | 5.49 |
| S4/B3 | 4.14 | 4.63 | 4.33 | 4.46 | 0.62 | 0.34 | 0.25 | 0.65 | 7.29 | 7.99 |
| S4/B4 | 4.52 | 5.09 | 4.89 | 4.99 | 0.43 | 0.49 | 0.45 | 0.60 | 6.73 | 7.20 |
| S4/B5 | 7.53 | 8.38 | 8.41 | 8.14 | 0.56 | 0.52 | 0.39 | 0.50 | 10.73 | 11.86 |
| S5/B1 | 2.38 | 2.71 | 2.67 | 2.58 | 0.31 | 0.49 | 0.67 | 0.45 | 2.44 | 2.51 |
| S5/B2 | 2.48 | 2.61 | 2.64 | 2.61 | 0.61 | 0.36 | 0.50 | 0.48 | 3.93 | 4.09 |
| S5/B3 | 3.60 | 3.94 | 3.96 | 3.85 | 0.31 | 0.61 | 0.61 | 0.37 | 5.05 | 5.49 |
| S5/B4 | 5.16 | 5.34 | 5.37 | 5.15 | 0.50 | 0.56 | 0.39 | 0.51 | 8.28 | 8.44 |
| S5/B5 | 8.43 | 9.02 | 8.70 | 8.67 | 0.50 | 0.47 | 0.45 | 0.56 | 12.03 | 12.71 |

To explore further, we check how the choice of the instrument variables affects our results. We run FCCAPM by using different state variables. Another two different state variables are studied here. For the first state variable, we run a linear regression of the asset returns on the instrument variables and treat the fitted value as the state variable. The second is the state variable implied by the time-varying betas in the FH model.[4] All of the state variables used for the smoothing variables are implemented in the framework of FCCAPM, and we report their MSEs in the fourth and the fifth columns in Table 3.

On average, SCAD gives us the best state variable among these three models. The majority of the MSE values in the second column are smaller than those in the other columns. The reason is that SCAD selects the variables to fit directly the asset returns and chooses the index that is the best for the model estimation. This method avoids the problems caused by the separation of instrument selection and model estimation.

When our model is employed, it produces the estimated coefficients of the instruments summarized in Table 3. To avoid the identification problem, we search for the estimates over a unit circle. Interestingly, the coefficients of the one-month interest rate are zero for many portfolios. This is consistent with the empirical evidence in Bernanke, 1990 that the interest rate spreads play a more important role than the rates in financial markets. This feature can not be observed in the paper by Ferson and Harvey (1999).

Moreover, we also analyze the out-of-sample forecasting for the Fama–French 25 portfolios. We use the rolling 360-month to estimate the coefficients of the instruments, and then use the nonparametric method to estimate the portfolio return in the next period. The results are summarized in the last two columns in Table 3. We can see that the MSE for the out-of-sample forecasting are quite comparable with the MSE for the in-sample fitting. In addition, our model delivers a smaller MSE than the FH model for all the 25 portfolios.

Knowing whether or not the betas are really time-varying is crucial to the conditional CAPM. We plot the estimates of both betas and alpha with their 95% pointwise confidence intervals in Fig. 1. The confidence intervals are obtained by bootstrapping the data 1000 times. To save space, we report only the S4/B2 portfolio. Fig. 1 shows clearly that the alphas and the betas are not constant over the time period studied. Also, we can see that the pricing errors fluctuate around zero. These conclude that the conditional CAPM is appropriate for this data set.

### 4.3. Model testing

After estimating the time-varying betas and alpha, we are curious about whether our model is over parametrized and whether our model is sufficient to explain the portfolio returns. To answer these two questions, we first test the model specified in Ferson and Harvey (1999). In addition, we test the significance of the pricing errors using both time series and cross-sectional methods. If the conditional CAPM is valid and our model estimates the betas correctly, then the pricing error should be insignificant.

#### 4.3.1. Testing linearity

Ferson and Harvey (1999) used a linear model to describe the time-varying betas, while our paper uses the nonparametric model

[4] The FH model allows the betas and alpha to have different state variables.

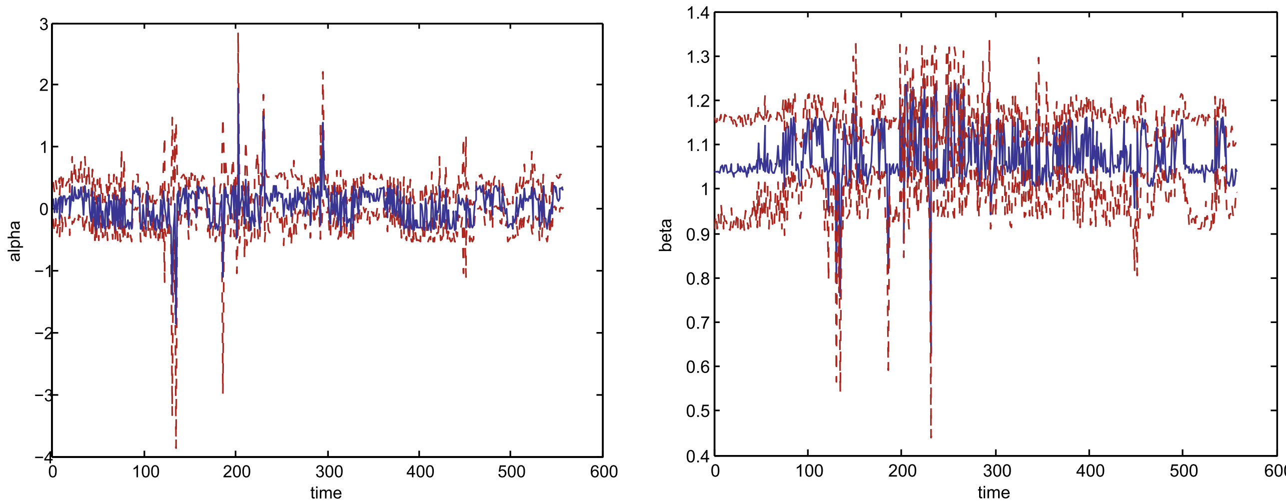


**Fig. 1.** Estimated alphas and betas with 95% pointwise confidence intervals for the S4/B2 portfolio.

to fulfill this job. It is necessary for us to test the model specification. Namely, the null hypothesis is

$$H_0 : g_1(Z_i) = a_0 + a_1' Z_i \quad \text{and} \quad g_2(Z_i) = b_0 + b_1' Z_i$$

and the alternative hypothesis is

$$H_1 : \text{either } g_1(Z_i) = f_1(Z_i) \text{ or } g_2(Z_i) = f_2(Z_i),$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are two unknown functions and at least one of them is nonlinear so that they can be estimated using the nonparametric method described in (3). The *p*-values for 25 portfolios are summarized in Table 4, Panel A.

The portfolios in each row are sorted by the book-to-market ratios and in each column by the sizes. We find that we can reject the null hypothesis for 17 out of 25 portfolios at 5% significance level, which means that the majority of the Fama–French 25 portfolios do not satisfy the linearity assumption imposed in Ferson and Harvey (1999). This finding supports partially the conclusion observed in Wang (2002, 2003). This nonlinearity is even more prominent for the portfolios that have high book-to-market ratios or small sizes. This nonlinearity implies the instruments capture the time lag information mainly through their high-order term. These findings seem to be new in the literature.

*4.3.2. Pricing error of time series analysis*

In the conditional CAPM, the alpha represents the abnormal returns of the assets. The abnormal returns occur when there are other systematic risks in the market that the model does not capture. Thus, we need to test the significance of the alpha. When the estimated alpha is significant, then the model is rejected. Specifically, if the model is misspecified, the test results may be misleading. Because we have already rejected the linear time-varying betas and alpha, we are warranted to test the significance of the alpha based on our model.

We estimate the time-varying alpha for the entire time period of the data and then perform a significance test on the alpha. The null hypothesis is $H_0 : g_{1,i} = g_1(Z_i) = 0$ and $H_1 : g_{1,i} \neq 0$ for $i = 1, 2, \ldots, n$. Here, we only impose restrictions on the alpha, and allow the betas to be time-varying. The *p*-values are obtained using the bootstrapping procedure described in Section 2.4 and they are summarized in Table 4, Panel B which shows that 20 out of 25 (80%) portfolios have insignificant alphas at the level of 5%. This result is quite different from the results found by Ferson and Harvey (1999), due to the model specification. As argued in Ghysels (1998), the inferences made from the misspecified model can be very misleading. Thus, the correct conditional CAPM can deliver insignificant pricing errors. Conversely, the insignificant pricing errors also suggest that the conditional CAPM can capture the systematic risk from period to period. This result also gives a strong support for the models that are based on the conditional CAPM, such as the premium-labor model proposed by Jagannathan and Wang (1996). This finding is innovative in the literature.

**Table 4**
Testing results. Panel A reports the *p*-values of testing linearity of the betas and alpha. The null hypothesis is the linear model in Ferson and Harvey (1999) and the alternative is our nonparametric model. The *p*-values are obtained by bootstrapping the data 1000 times. The portfolios in each row are sorted by the book-to-market ratio and in each column by the size. Panel B reports the *p*-values for the test on the significance of the alphas when the model has time-varying betas. The null hypothesis is $H_0 : \alpha_i = 0$ and the alternative is $H_1 : \alpha_i \neq 0$. Panel C demonstrates parameter estimates from the cross-sectional regression of excess returns on constant, time-varying beta and fitted conditional expected return. The parameters are estimated by two methods of rolling sample and expanding sample, and the corresponding Fama–MacBeth *t*-ratios are reported in parentheses.

| | B1 | B2 | B3 | B4 | B5 |
|---|---|---|---|---|---|
| *Panel A: p-values of testing linearity* | | | | | |
| S1 | 0.2500 | 0.1954 | 0.0002 | 0.0000 | 0.0000 |
| S2 | 0.2082 | 0.0125 | 0.0011 | 0.0001 | 0.0088 |
| S3 | 0.2344 | 0.0412 | 0.0168 | 0.0248 | 0.0038 |
| S4 | 0.0050 | 0.0002 | 0.0136 | 0.0184 | 0.0006 |
| S5 | 0.2508 | 0.0360 | 0.0702 | 0.2628 | 0.0806 |
| *Panel B: p-values of testing pricing errors: time series approach* | | | | | |
| S1 | 0.6840 | 0.5480 | 0.0020 | 0.0000 | 0.0002 |
| S2 | 0.6520 | 0.3980 | 0.0100 | 0.1500 | 0.1060 |
| S3 | 0.0900 | 0.2620 | 0.5300 | 0.0300 | 0.1480 |
| S4 | 0.0860 | 0.0800 | 0.7020 | 0.0500 | 0.4880 |
| S5 | 0.5400 | 0.0940 | 0.6160 | 0.6660 | 0.5480 |
| *Panel C: cross-sectional regression* | | | | | |
| | $\gamma_0$ | $\gamma_1$ | $\gamma_2$ | | |
| Rolling sample | 0.6815 | −0.0057 | 0.8408 | | |
| (*t*-statistic) | (0.1171) | (−0.0010) | (0.1622) | | |
| Expanding sample | 0.8599 | −0.3111 | 0.9330 | | |
| (*t*-statistic) | (0.1166) | (−0.0437) | (0.2138) | | |

*4.3.3. Cross sectional analysis of pricing errors*

In addition, we also consider testing the pricing errors using a cross-sectional regression technique. Since the conditional CAPM model is expressed as

$$E[R_{j,i} | I_{i-1}] = \gamma_{1,i-1} \beta_{j,i-1}, \tag{14}$$

where $\gamma_{1,i-1}$ is the conditional market risk premium, following Ferson and Harvey (1999), the cross-sectional regression can be setup as

$$R_{j,i} = \gamma_{0,i-1} + \gamma_{1,i-1}\beta_{j,i-1} + \gamma_{2,i-1}\delta'_{j,i-1}Z_{i-1} + e_{j,i}, \quad j = 1, \cdots, 25, \quad (15)$$

where $\gamma_{0,i-1}$ is the intercept, $\gamma_{1,i-1}$, $\gamma_{2,i-1}$ are the slope coefficients and $\beta_{j,i-1}$ are estimated from our previous time series regression. $\delta'_{j,i-1}Z_{i-1}$ is the fitted conditional expected return, where $\delta_{j,i-1}$ is estimated by regressing the return on the lagged variable $Z_{i-1}$, using the data up to time $i-1$. $\gamma_{0,i-1}$ denotes the pricing error in the cross-sectional regression, and $\gamma_{2,i-1}$ indicates how much information is not captured by the conditional betas. Therefore, if the conditional CAPM is valid, $\gamma_{0,i-1}$ and $\gamma_{2,i-1}$ should be insignificant.

We run the cross-sectional regression via two methods. One method uses the rolling 120-month prior estimation period, and the other uses an expanding sample.[5] The time-series averages of the cross-sectional regression coefficients are shown along with their Fama–MacBeth $t$-ratios (two-pass regression) in Table 4, Panel C.

The results in Panel C suggest that $\gamma_0$ is insignificant in both cases. Although the pricing errors of some portfolios are significant in the time-series analysis, we can still obtain insignificant pricing errors for the cross-sectional regression. This is because, in the time-series analysis, pricing error is a function of the instruments, $g_{1,i} = g_1(Z_i)$. The realized value of this function is not zero, but its expected value can be zero. Moreover, we find that $\gamma_2$ is insignificant as well. This implies that the public information has already been revealed in the time-varying betas. These two findings can justify the conditional CAPM from the cross-sectional point of view.

## 5. Conclusion

This paper uses a functional coefficient regression with an index to estimate the time-varying betas and alpha in the conditional capital asset pricing model. Functional coefficient representation relaxes the strict assumptions regarding the structure of betas and alpha by combining the predictors into an index. The index captures time variations in betas and alpha and can be estimated in nonparametric way. In such a way, it helps us to determine which economic variables we should track and, more importantly, in what combination. We select appropriate index variables by using a smoothly clipped absolute deviation penalty on functional coefficients. In this manner, estimation and variable selection can be performed simultaneously.

Our findings are quite interesting. First, empirical estimation results show that important instruments are selected automatically in our model, and it has better fit compared with other methods in terms of mean square error. The 95% pointwise confidence interval plots for alphas and betas suggest time-varying CAPM exists for the data set. Second, our testing results show that majority of the Fama–French 25 portfolios do not satisfy the linearity assumption. Insignificant pricing errors are found from time series analysis and cross sectional analysis under the framework of FCCAPM. These support the conventional wisdom about the conditional CAPM, which holds from period to period. It should be pointed out that our FCCAPM model is established in the CAPM framework with single market factor. The performances of the FCCAPM model and the FH model in the framework of Fama and French (1993) 3-factor model are available from the authors upon request. For the future research, it would be interesting to consider its practical performance with other multiple factors, such as the Fama and French (1993) 3-factor model, the Carhart, 1997 4-factor model, the Fama and French (2015) 5-factor model and the Hou et al. (2015) empirical $q$-factor model.

[5] For the expanding sample, we use 120 months for the regression at first. Then, we add the 121th month into the sample and run the regression again. This procedure is iterated until all the observations are included in the regression sample.

## Appendix A. Proofs

**Proof of Theorem 1.**

$$\widehat{Q}_1(c,h) = \widetilde{S}(c) + T(h) + R_1(c,h) + R_2(h),$$

where $\widehat{Q}_1(c,h) = \sum_{i=1}^n (y_i - \widehat{g}^T(c^T Z_i)X_i)^2, T(h)$ and $R_2(h)$ do not depend on $c$, and $R_1(c,h)$ is an ignorable term. Furthermore,

$$\widetilde{S}(c) = n[\widetilde{V}_0^{1/2}(c-c_0) - n^{-1/2}\sigma\varepsilon]^T[\widetilde{V}_0^{1/2}(c-c_0) - n^{-1/2}\sigma\varepsilon] + R_3 + R_4(c),$$

where $R_3$ does not depend on $c$ and $h$, and $R_4(c)$ is an ignorable term.

Let $\delta_n = n^{-1/2} + a_n, t = (t_1, \ldots t_d)^T$. For any small $\Lambda > 0$, if we can show there exists a large constant $C$, such that

$$P\left\{\inf_{\|t\|=C} Q(c_0 + \delta_n t, \widehat{g}) > Q(c_0, \widehat{g})\right\} > 1 - \Lambda,$$

then

$$\|\widehat{c} - c_0\| = O_p(\delta_n).$$

Define $D_n = Q(c_0 + \delta_n t, \widehat{g}) - Q(c_0, \widehat{g})$. Then,

$$D_n \geqslant \frac{1}{2}\sum_{i=1}^n \left(y_i - \widehat{g}^T(c_0^T Z_i + \delta_n t^T Z_i)X_i\right)^2 - \frac{1}{2}\sum_{i=1}^n \left(y_i - \widehat{g}^T(c_0^T Z_i)X_i\right)^2 + n\sum_{k=1}^{d_1} P_{\lambda_n}(|c_{10k} + \delta_n t_k|) - n\sum_{k=1}^{d_1} P_{\lambda_n}(|c_{10k}|) \quad (\text{by } c_{20} = 0)$$

and

$$\begin{aligned} & n\sum_{k=1}^{d_1} P_{\lambda_n}(|c_{10k} + \delta_n t_k|) - n\sum_{k=1}^{d_1} P_{\lambda_n}(|c_{10k}|) \\ &= n\sum_{k=1}^{d_0}\left[\delta_n P'_{\lambda_n}(|c_{10k}|)sgn(c_{10k})t_k + \frac{1}{2}\delta_n^2 P''_{\lambda_n}(|c_{10k}|)t_k^2\right] + o_p(n\delta_n^2) \\ &\leqslant \sqrt{d_1}n\delta_n a_n\|t\| + \frac{1}{2}n\delta_n^2 max_{1\leqslant k\leqslant d_1}\{P''_{\lambda_n}(|c_{10k}|)\}\|t\|^2 + o_p(n\delta_n^2) \\ &\quad (\text{by Cauchy-Schwarz inequality}) \\ &\leqslant n\delta_n^2\sqrt{d_1}C + O_p(n\delta_n^2) \quad \text{as} n \to \infty \text{ and } max_{1\leqslant k\leqslant d_0}\{P''_{\lambda_n}(|c_{10k}|)\} \to 0 \end{aligned}$$

and

$$\begin{aligned} & \tfrac{1}{2}\sum_{i=1}^n \left(y_i - \widehat{g}^T(c_0^T Z_i + \delta_n t^T Z_i)X_i\right)^2 - \tfrac{1}{2}\sum_{i=1}^n \left(y_i - \widehat{g}^T(c_0^T Z_i)X_i\right)^2 \\ &= \tfrac{1}{2}n\left[\widetilde{V}_0^{1/2}\delta_n t - n^{-1/2}\sigma\varepsilon\right]^T\left[\widetilde{V}_0^{1/2}\delta_n t - n^{-1/2}\sigma\varepsilon\right] - \tfrac{1}{2}n\left[n^{-1/2}\sigma\varepsilon\right]^T\left[n^{-1/2}\sigma\varepsilon\right] \\ &\quad + R_1(c_0 + \delta_n t, h) - R_1(c_0, h) + o_p(1) \quad (\text{by the theorem in Xia and Li, 1999}) \\ &= \tfrac{1}{2}n\delta_n^2 t^T\widetilde{V}_0 t - n^{1/2}\delta_n t^T\widetilde{V}_0^{1/2}\sigma\varepsilon + R_1(c_0 + \delta_n t, h) - R_1(c_0, h) + o_p(1) \\ &= \tfrac{1}{2}n\delta_n^2 t^T\widetilde{V}_0 t - \delta_n t^T V_n + R_1(c_0 + \delta_n t, h) - R_1(c_0, h) + o_p(1). \end{aligned}$$

Since $R_1$ are negligible terms as $n \to \infty$ and $\frac{1}{\sqrt{n}}V_n = O_p(1)$. then $-\delta_n t^T V_n = C \cdot O_p(\delta_n\sqrt{n}) = C \cdot O_p(\delta_n^2 n)$. By choosing a sufficient large $C$, the term $\frac{1}{2}n\delta_n^2 t^T\widetilde{V}_0 t = C^2 \cdot O_p(\delta_n^2 n)$ will dominate others. Hence, $D_n \geqslant 0$ holds. □

**Proof of Theorem 2.** Let $\widehat{c}_1 - c_{10} = O_p(n^{-1/2})$. We want to show that $(\widehat{c}_1, 0)^T = \text{argmin}_{(c_1^T, c_2^T)^T \in \mathbb{B}} Q\left((c_1^T, c_2^T)^T, \widehat{g}\right)$. It suffices to show that for some constant $C$ and $k = q_0 + 1, \cdots q$,

$$\frac{\partial Q\left((c_1^T, c_2^T)^T, \widehat{g}\right)}{\partial c_k} \begin{array}{ll} > 0 & \text{for } 0 < c_k < Cn^{-1/2} \\ < 0 & \text{for } -Cn^{-1/2} < c_k < 0. \end{array}$$

Note that

$$\frac{\partial \widehat{Q}_1(c,h)}{\partial c_k} = \frac{\partial \widetilde{S}(c)}{\partial c_k} + R_m = e_k^T \frac{\partial \widetilde{S}(c)}{\partial c} + R_m$$
$$= 2ne_k^T \widetilde{V}_0(c - c_0) - 2n^{1/2}\sigma e_k^T \widetilde{V}_0^{1/2}\varepsilon + R_m$$
$$= 2ne_k^T \widetilde{V}_0(c - c_0) - 2e_k^T V_n + R_m$$

where $R_m$ represents small order term and $e_k$ is a $d$-dimensional vector with $k$th element being one and all others being zero.

Since $c - c_0 = O_p(1/\sqrt{n})$ and $V_n = O_p(\sqrt{n})$, then,

$$\frac{\partial \widehat{S}(c,h)}{\partial c_k} = O_p(\sqrt{n})$$

and

$$\frac{\partial Q\left((c_1^T, c_2^T)^T, \widehat{g}\right)}{\partial c_k} = \frac{1}{2}\frac{\partial \widehat{Q}_1(c,h)}{\partial c_k} + nP'_{\lambda_n}(|c_k|)\, sgn(c_k)$$
$$= n\lambda_n\left[O_p\left(\frac{1}{\sqrt{n}\lambda_n}\right) + \frac{P'_{\lambda_n}(|c_k|)}{\lambda_n} sgn(c_k)\right].$$

Since $\sqrt{n}\lambda_n \to \infty$ and $\liminf_{n\to\infty, c_k \to 0^+} \frac{P'_{\lambda_n}(|c_k|)}{\lambda_n} > 0$, the sign of $\frac{\partial Q}{\partial c_k}$ is determined by the sign of $c_k$.

It follows from Part (a) that

$$\left.\frac{\partial Q\left((c_1^T, c_2^T)^T, \widehat{g}\right)}{\partial c}\right|_{c = \begin{pmatrix} \widehat{c}_1 \\ 0 \end{pmatrix}} = 0$$

and

$$\frac{1}{2}\frac{\partial \widehat{S}((\widehat{c}_1, 0), h)}{\partial c_1} + n\Delta\Psi_{\lambda}^{d_1} = 0.$$

where $\Delta\Psi_{\lambda_n}^{d_1} = \{P'_{\lambda_n}(|c_1|)\, sgn(c_1), \ldots, P'_{\lambda_n}(|c_{d_1}|)\, sgn(c_{d_1})\}^T$. Note that as $n \to \infty$ and $\lambda_n \to 0, P'_{\lambda_n}(|c_k|) = 0$ for $k = 1, \ldots, d_1$ and

$$\frac{1}{2}\frac{\partial \widehat{S}((\widehat{c}_1, 0), h)}{\partial c_1} = 0,$$

which implies that

$$n\widetilde{V}_{10}(\widehat{c}_1 - c_{10}) - n^{1/2}\sigma \widetilde{V}_{10}^{1/2}\mathbf{e}_1 + o_p(1) = 0$$

$$\sqrt{n}(\widehat{c}_1 - c_{10}) = \widetilde{V}_{10}^{-1}(1/\sqrt{n})V_{1n}$$
$$= \widetilde{V}_{10}^{-1}(1/\sqrt{n})\sum_{i=1}^{n}(Z_{1i} - E(Z_{1i}|c_{10}^T Z_{1i}))\dot{g}^T(c_{10}^T Z_{1i})X_i\varepsilon_{1i}$$

so that

$$\sqrt{n}(\widehat{c}_1 - c_{10}) \to^D N\left(0, \widetilde{V}_{10}^{-1}V_{10}\widetilde{V}_{10}^{-1}\right).$$

where $\widetilde{V}_{10}$ is defined earlier and $V_{10} = \Gamma(0) + 2\sum_{\ell=1}^{\infty}\Gamma(\ell)$ with $\Gamma(\ell) = \text{Cov}(\Gamma_i, \Gamma_{i-\ell})$ and $\Gamma_i = (Z_{1i} - E(Z_{1i}|c_{10}^T Z_{1i}))\dot{g}^T(c_{10}^T Z_{1i})X_i\varepsilon_{1i}$. □

## References


Akdeniz, L., Altay-Salih, A., Caner, M., 2003. Time-varying betas help in asset pricing: the threshold CAPM. Studies in Nonlinear Dynamics & Econometrics 6, 1–16.
Aït-Sahalia, Y., Brandt, M.W., 2001. Variable selection for portfolio choice. Journal of Finance 56, 1297–1351.
Ang, A., Chen, J., 2007. CAPM over the long run: 1926–2001. Journal Empirical Finance 14, 1–40.
Ang, A., Kristensen, D., 2012. Testing conditional factor models. Journal of Financial Economics 106, 132–156.
Bernanke, B., 1990. On the predictive power of interest rates and interest rate spreads. New England Economic Review, 51–68.
Cai, Z., 2002. A two-stage approach to additive time series models. Statistica Neerlandica 56, 415–433.
Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. Journal of American Statistical Association 95, 941–956.
Cai, Z., Hong, Y., 2009. Some recent developments in nonparametric finance. Advances in Econometrics 25, 379–432.
Cai, Z., Tiwari, R.C., 2000. Application of a local linear autoregressive model to BOD time series. Environmetrics 11, 341–350.
Campbell, J., Vuolteenaho, T., 2004. Good beta, bad beta. American Economic Review 94, 1249–1275.
Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.
Connor, G., Hagmann, M., Linton, O., 2012. Efficient semiparametric estimation of the Fama–French model and extensions. Econometrica 80, 713–754.
Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.
Fama, E., French, K., 1997. Industry costs of equity. Journal of Financial Economics 43, 153–193.
Fama, E., French, K., 2005. Financing decisions: who issues stock? Journal of Financial Economics 76, 549–582.
Fama, E., French, K., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and Its Applications. Chapman and Hall, London.
Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of American Statistical Association 96, 1348–1360.
Fan, J., Lv, J., 2008. Sure independence screening for ultra-high dimensional feature space. Journal of Royal Statistical Society B 70, 849–911.
Fan, J., Lv, J., 2009. A selective overview of variable selection in high dimensional feature space. Statistica Sinica 20, 101–148.
Fan, J., Yao, Q., Cai, Z., 2003. Adaptive varying-coefficient linear models. Journal of Royal Statistical Society B 65, 57–80.
Ferreira, E., Gil-Bazo, J., Orbe, S., 2011. Conditional beta pricing models: a nonparametric approach. Journal of Banking and Finance 35, 3362–3382.
Ferson, W., Harvey, C., 1997. Fundamental determinants of national equity market returns: a perspective on conditional asset pricing. Journal of Banking and Finance 21, 1625–1665.
Ferson, W., Harvey, C., 1999. Conditional variables and the cross section of stock returns. Journal of Finance 54, 1325–1360.
Ghysels, E., 1998. On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? Journal of Finance 53, 549–573.
Harvey, C., 2001. The specification of conditional expectations. Journal of Empirical Finance 8, 573–637.
Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: an investment approach. Review of Financial Studies 28, 650–705.
Ichimura, H., 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single index model. Journal of Econometrics 58, 71–120.
Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3–53.
Johnstone, I.M., Silverman, B.W., 1997. Wavelet threshold estimators for data with correlated noise. Journal of the Royal Statistical Society 59, 319–351.
Kreiss, J.-P., Neumann, M.H., Yao, Q., 2008. Bootstrap tests for simple structures in nonparametric time series regression. Statistics and its Interface 1, 367–380.
Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56, 815–849.
Lewellen, J., Nagel, S., 2006. The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics 82, 289–314.
Li, X., Cai, Z., Ren, Y., 2015. A new test on the conditional capital asset pricing model. Applied Mathematics – A Journal of Chinese Universities, Series B 30, 163–186.
Lin, C.F., Teräsvirta, T., 1994. Testing the constancy of regression parameters against continuous structural change. Journal of Econometrics 62, 211–228.
Robinson, P.M., 1997. Large-sample inference for nonparametric regression with dependent errors. The Annals of Statistics 25, 2054–2083.
Wang, K., 2002. Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio. Journal of Empirical Finance 9, 133–169.
Wang, K., 2003. Asset pricing with conditional information: A new test. Journal of Finance 58, 161–196.
Zhang, L., 2005. The value premium. Journal of Finance 60, 67–103.