University of Cincinnati

Date: 4/30/2013

I, Chaojiang Wu , hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Business Administration.

It is entitled:

Essays on High-dimensional Nonparametric Smoothing and Its Applications to Asset Pricing

Student's name: **Chaojiang Wu**

This work and its defense approved by:

Committee chair: Yan Yu, Ph.D.

Committee member: Hui Guo, Ph.D.

Committee member: Martin Levy, Ph.D.

3413

Essays on High-dimensional Nonparametric Smoothing and Its Applications to Asset Pricing

A dissertation submitted to the Graduate School of

the University of Cincinnati in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Department of Operations, Business Analytics and Information Systems of the

Lindner College of Business

2013

by

Chaojiang Wu

M.S. Quantitative Analysis, 2011, University of Cincinnati

M.S. Management Science, 2007, University of Science and Technology of China (USTC)

B.S. Biological Science, 2004, USTC

B.E. Electronic Information Engineering, 2004, USTC

Committe Chair: Dr. Yan Yu

Abstract

Nonparametric smoothing, a method of estimating smooth functions, has gained increasing popularity in statistics and application literature during the last few decades. This dissertation has focused primarily on the nonparametric estimation in quantile regression (Chapter 1) and an application of nonparametric estimation to financial asset pricing (Chapter 2).

In the first essay (Chapter 1), we consider the estimation problem of conditional quantile when multi-dimensional covariates are involved. To overcome the "curse of dimensionality" yet retain model flexibility, we propose two partially linear models for conditional quantiles: partially linear single-index models (QPLSIM) and partially linear additive models (QPLAM). The unknown univariate functions are estimated by penalized splines. An approximate iteratively reweighted penalized least square algorithm is developed. To facilitate model comparisons, we develop effective model degrees of freedom for penalized spline conditional quantiles. Two smoothing parameter selection criteria, Generalized Approximate Cross-validation (GACV) and Schwartz-type Information Criterion (SIC) are studied. Some asymptotic properties are established. Finite sample properties are investigated through simulation studies. Application to the Boston Housing data demonstrates the success of proposed approach. Both simulations and real applications show encouraging results of the proposed estimators.

In the second essay (Chapter 2), we investigate whether the conditional CAPM helps explain the value premium using the single-index varying-coefficient model. Our empirical specification has two novel advantages relative to those commonly used in the previous studies. First, it not only allows for a flexible dependence of conditional beta on state variables but also modeling heteroskedasticity. Second, from a large set of candidate state variables, we identify the most influential ones through an exhaustive variable selection method. We have also developed statistics to test the functional form of conditional beta and alpha, which provides justifications for or against the practices of letting conditional beta depend linearly on state variables and assuming constant alpha. Consistent with the notion that the value premium tends to be riskier during business recessions than during business expansions, we find that its conditional beta co-moves with unemployment and inflation, the two most closely watched gauges of aggregate economy by the Federal Reserve, and the price-earnings ratio. Realized beta does not subsume all the other explanatory variables when we include the realized beta as a state variable. The alpha is smaller for the conditional CAPM than for the unconditional CAPM; nevertheless, neither model fully explains the value premium.

Key Words: Additive Model; Conditional CAPM; Partially Linear Model; Penalized Splines; Single-Index Models; Semiparametric Model; Smoothing Parameter; Value Premium; Variable Selection.

Acknowledgements

I would like to thank faculty, family and friends, without whom this doctoral dissertation would not have been possible.

Above all, I would like to express the deepest appreciation to my committee chair, Professor Yan Yu, for her unfailing guidance, encouragement and patience. I am fortunate to have an adviser and personal friend as great as Yan, who is always available for questions and discussions. I quite simply cannot imagine a better adviser.

In addition, I am very thankful to my committee members, Professor Martin Levy and Professor Hui Guo for their time and valuable feedback on my dissertation. I would also like to extend my appreciation to the Operations and Business Analytics faculty, from whom I learned a great deal, academically and professionally. I would like to further extend my appreciation to the Doctoral students in the Operations and Business Analytics department.

Finally, I would like to thank my family and friends, whose support and encouragement have been indispensable for me to achieve my goals. I am most grateful to have dedicated parents who have sacrificed a lot to put me through school. This dissertation would not have been possible without their lifetime dedication and unconditional love. I am also grateful to have my brother supporting me, especially during hard times. There are numerous individuals that I wish to thank for their friendship through the years. Special thanks go to Junbin Hu, whose friendship and generous financial support are greatly appreciated.

To My Family

Table of Contents

List of Tables

List of Figures

Chapter 1

Dimension Reduction For Quantile Regression

1.1 Introduction

This study is motivated by analyzing the well-known Boston Housing data. The dataset contains 506 observations of the median price of owner-occupied homes of suburban Boston together with 13 variables from the 1970 census. Among the 13 covariates, most variables are continuous in nature while some are categorical. For example, variable *Chas* is a Charles River dummy variable of value 1 if tract bounds river and 0 otherwise. Variables such as *rm*: average number of rooms per dwelling; *tax*: full-value property-tax; *ptratio*: pupil-teacher ratio by town; *lstat*: percentage of lower status of the population; *dis*: weighted distances to five Boston employment centers; however, are continuous in nature. The dataset and detailed description are available through the Stat library at Carnegie Mellon University and are ready to use in R.

The dependent variable of interest the median housing price *medv* and many covariates are left-skewed. Hence, quantile regression or conditional quantile (see the seminal work Koenker and Bassett 1978) can be naturally used to examine this data. Let Y be the response variable of the median housing price *medv*, and $\mathbf{Z}$ be the covariate vector including the variables such as the dummy variable *Chas*. Let covariate vector $\mathbf{X}$ denote the remaining variables, consisting of $X_1 = rm$, $X_2 = log(tax)$, $X_3 = ptratio$, $X_4 = log(lstat)$ etc. for example. To model the τ-th conditional quantile of the response Y, a linear quantile regression model takes the form

$$q_\tau(Y|\mathbf{X}, \mathbf{Z}) = a + \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}.$$

Linear quantile regression is simple in computation and easy-to-interpret. However, important features of nonlinearity that have been clearly observed (e.g. Figure 5) can not be captured. On the other hand, one may model the conditional quantile fully nonparametrically (Chaudhuri 1991a, 1991b). Nevertheless, for practical problems, the fully nonparametric model suffers from the well-known curse of dimensionality.

To overcome the "curse of dimensionality" yet retain model flexibility, we propose two partially linear models for conditional quantiles: partially linear single-index models (QPLSIM) and partially linear additive models (QPLAM). The partially linear single-index models for conditional quantiles takes the form

$$q_\tau(Y|\mathbf{X}, \mathbf{Z}) = g(\mathbf{X}\boldsymbol{\alpha}) + \mathbf{Z}\boldsymbol{\beta}. \tag{1.1}$$

QPLSIM has two components: the linear combination $\mathbf{X}\boldsymbol{\alpha}$ which is often termed as single-index enters the model via a univariate nonparametric link function $g(\cdot)$ and $\mathbf{Z}\boldsymbol{\beta}$ enters the model as a partially linear term. By reducing the dimensionality from that of a general covariate vector $\mathbf{X}$ to a univariate single-index $\mathbf{X}\boldsymbol{\alpha}$, QPLSIM avoids the so-called "curse of dimensionality". Partially linear term also enjoys easier interpretation of the effect of each variable. It is worth noting that when the univariate link function $g(\cdot)$ is monotonic, the single-index coefficient $\boldsymbol{\alpha}$ of model (1.1) retains similar ease-of-interpretation as that in the linear quantile regression while allowing for curvature through the univariate nonparametric link function.

Model (1.1) is quite flexible to include single-index models and partially linear models as special cases. When there are no partially linear term $\mathbf{Z}\boldsymbol{\beta}$, model (1.1) reduces to the single-index quantile models (Wu, Yu and Yu 2010). Based on their estimation, Kong and Xia (2012) investigated the Bahadur representation of single-index parameter estimators. Both papers adopt local linear methods to estimate the univariate link function $g(\cdot)$. Chaudhuri, Doksum and Samarov (1997) proposed average derivative approach, where the single-index coefficient is estimated by taking an expectation of the vector of partial derivatives of the conditional quantile with respect to the covariates. Since this approach requires multivariate kernel regression, it is less appealing in practice with high-dimensional covariates. Adding partially linear terms is more desirable when categorical variables such as *chas* are present. In fact, when the single-index coefficient is known or the covariate $\mathbf{X}$ is univariate, model (1.1) reduces to partially linear models. For example, He and Liang (2000) studied partially linear errors-in-variables quantile models. Chen and Khan (2001) studied partially linear censored quantile regression. Lee (2003) studied efficient estimator for partially linear terms. Most recently, Härdle, Ritov and Song (2012) studied bootstrap approximation for uniform confidence band.

The second approach we propose is partially linear additive models for quantile regression (QPLAM)

$$q_\tau(Y|\mathbf{X},\mathbf{Z}) = a + g_1(X_1) + \cdots + g_d(X_d) + \mathbf{Z}\boldsymbol{\beta}, \tag{1.2}$$

which include additive models and partially linear models as special cases. Additive models are popular for dimension regression which replace the linear component by a sum of univariate nonparametric functions over the components of $\mathbf{X}$; see e.g. De Gooijer and Zerom (2003) on kernel based marginal integration, Yu and Lu (2004) on local linear smoothing, Horowitz and Lee (2005) on two-stage estimator for quantile regression.

Note that QPLAM model (1.2) reduces dimensionality nicely but does not incorporate interactions of $\mathbf{X}$ in the current structure. QPLSIM model (1.1) applies a nonlinear link function to the index $\mathbf{X}\boldsymbol{\alpha}$, hence some interactions between the covariates can be modeled. Computationally QPLSIM are more complex partly due to this nonlinear structure on index parameters where nonlinear optimization has to be involved. In this study, we find that both approaches are appealing alternatives for dimension reduction on conditional quantile regression.

We estimate the nonparametric functions in models (1.1) and (1.2) using penalized splines (P-splines). Penalized splines have gained increasing popularity, especially due to the computational expediency and easy-adaptability to more complex models (see Ruppert, Wand and Carroll 2003 for a review). Yu and Ruppert (2002) show that penalized spline estimation for partially linear single-index models in mean regression are computationally stable and expedient, while local method estimation may become computationally unstable (Carroll, Fan, Gijbels and Wand 1997). To the best of our knowledge, this is the first work to adopt penalized spline estimation for single-index quantile regression.

Furthermore, we study the effective degrees of freedom in P-splines quantile regression models and investigate smoothing parameter selection in detail. First, we develop the measure for effective model degrees of freedom for the proposed partially linear single-index models (QPLSIM) and partially linear additive models (QPLAM). The effective degrees of freedom help to facilitate model comparisons. For QPLAM, the P-spline approach also easily allows different smoothness for different additive functions by assigning different smoothing parameters accordingly. Second, we study in simulations two smoothing parameter selection criteria, Generalized Approximate Cross-validation (GACV) and Schwarz-type information criterion (SIC). Incorporating model degrees of freedom, we compare model goodness-of-fit for a variety of conditional quantile models with Boston Housing data. The findings are interesting. Among the single-index models, QPLSIM models outperform single-index models without partially linear terms. Similar results are observed in the additive models. Both observations suggest that incorporating model degrees of freedom, partially linear terms are useful to include in modeling conditional quantiles. Among the two partially linear models, QPLSIM and QPLAM are comparable in terms of model goodness-of-fit (see Table 1.5 and Table 1.6). Finally, simulation studies with various error distributions show P-splines estimator for QPLSIM performs well. Simulations with identical design as De Gooijer, Zerom (2003) and Horowitz and Lee (2005) show the superiority of P-splines estimator to the existing additive quantile estimators (see Table 1.3).

While it is natural to consider QPLSIM (1.1) for dimension reduction, its estimation is by no means trivial. Note that the single-index parameter $\boldsymbol{\alpha}$ is obtained through a d-dimensional nonlinear optimization which is subject to possibly local optimum. From our personal communication, Carroll et al. (1997), using local linear approach, observe unstable convergence even with a moderate three dimensional single-index covariates in the mean regression context. Quantile regression adds more challenge in the estimation. Due to the non-differentiablity nature of the objective function, the so-called "check" function in quantile regression, typically linear programming is involved even in a simple linear quantile regression. Recently, Wu, Yu, Yu (2010) propose iterative backfitting algorithm through linear programming using local linear approach without considering partially linear terms. Only comparison of model average absolute residuals is considered due to lack of appropriate model complexity measure.

In this paper, we develop an approximate iteratively reweighted least square algorithm that is not only computationally expedient but also allows for the measure of model degrees of freedom and smoothing parameter selection. The main idea of our algorithm is centered on minimizing a check function along with a roughness penalty for models (1) and (2), where the check function is approximated by a differentiable function near a small neighborhood of the origin. This idea can be traced back to Nychka, Gray, Haaland, Martin and O'Connell (1995) and have also been adopted in Yuan (2006) and Li, Liu and Zhu (2007) in different modeling contexts. Based on this approximation, we develop effective model degrees of freedom which facilitate model comparisons using mean absolute deviations.

The rest of the paper is organized as follows. Section 1.2 introduces the partially linear single-index models (QPLSIM) for the conditional quantile, describes the P-splines estimation together with algorithms, establishes some asymptotic properties of the proposed estimators and presents the simulation results. Section 1.3 provides the study of the partially linear additive conditional quantile models (QPLAM). Section 1.4 presents an application to the Boston Housing data. Section 1.5 concludes the paper. Technical proofs are relegated to Section 1.6.

1.2 Partially Linear Single-Index Conditional Quantiles

We first formally introduce the QPLSIM model and outline P-splines estimation, then discuss smoothing parameter selection and algorithm, and finish this section with asymptotic properties and finite sample simulation studies.

1.2.1 The Model

Suppose we have n observations $\left\{(\mathbf{x}_i, \mathbf{z}_i, y_i)\right\}_{i=1}^{n}$ of $(\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, Y = y)$. Given $\tau \in (0, 1)$ and covariates $\mathbf{x}_i$ and $\mathbf{z}_i$, the partially linear single-index model we propose for the τ-th conditional quantile of the i-th observation y_i is

$$q_\tau(y_i|\mathbf{x}_i, \mathbf{z}_i) = g_\tau(\mathbf{x}_i\boldsymbol{\alpha}_\tau) + \mathbf{z}_i\boldsymbol{\beta}_\tau, \tag{1.3}$$

where

- covariates $\mathbf{x}_i$ and $\mathbf{z}_i$ are d and d_z-dimensional row vectors respectively; column vectors of the single index parameter $\boldsymbol{\alpha}_\tau$ in $\mathbf{R}^d$, the partially linear parameter $\boldsymbol{\beta}_\tau$ in $\mathbf{R}^{d_z}$, and the univariate function $g_\tau : \mathbf{R} \to \mathbf{R}$ are subject to different τ;
- for identifiability, the single index parameter $\|\boldsymbol{\alpha}_\tau\| = 1$ and its first non-zero element of $\boldsymbol{\alpha}_\tau$ is positive.

The identifiability constraint of the single-index parameter $\boldsymbol{\alpha}_\tau$ can be handled by reparameterization to a $(d-1)$-dimensional unconstrained parameter by setting its first non-trivial component as a positive normalized constant. Details of the reparameterization are given in Section 1.6. To further simplify our notation, we omit the subscript τ hereafter when no confusion would occur.

1.2.2 P-splines For QPLSIM

Under model (1.3), the primary interest is to estimate the univariate function $g(\cdot)$, the single-index parameter $\boldsymbol{\alpha}$ and the partially linear parameter $\boldsymbol{\beta}$. We estimate the unknown univariate link function $g(\cdot)$ by penalized splines or P-splines. P-spline approach has become an increasingly popular method (Ruppert, Wand and Carroll 2003), largely due to its computational efficiency, its flexibility to capture nonlinearity, and its capability of easily adapting to more complex problems.

Specifically, given the single-index parameter $\boldsymbol{\alpha}$ or single-index $u = \mathbf{X}\boldsymbol{\alpha}$, the unknown univariate link function can be modeled by $g(u) = \mathbf{B}(u)\boldsymbol{\gamma}$, where $\mathbf{B}(u) = (B_1(u), \cdots, B_{d_\gamma}(u))$ is a d_γ-dimensional row vector of spline basis evaluated at the single-index u and $\boldsymbol{\gamma}$ is the spline coefficient column vector. Popular choices of the spline basis are the computational stable B-spline basis and easy-to-understand truncated power basis. For example, the p-degree truncated power basis can be represented by

$$\mathbf{B}(u) = (1,\ u,\ \cdots,\ u^p,\ (u-\kappa_1)^p max(u-\kappa_1, 0),\ \cdots,\ (u-\kappa_K)^p max(u-\kappa_K, 0)),$$

where $(\kappa_1, \cdots, \kappa_K)$ are the spline knots that are often placed at the equally spaced quantiles of u and here $d_\gamma = 1 + p + K$. Define $\boldsymbol{\theta} = \left(\alpha_1, \ldots, \alpha_d, \gamma_1, \ldots, \gamma_{d_\gamma}, \beta_1, \ldots, \beta_{d_z}\right)^\mathsf{T}$ as our column vector of parameters. Combine the partially linear term, our spline model for the τ-th conditional quantile is

$$q_\tau(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}) = \mathbf{B}(\mathbf{x}_i\boldsymbol{\alpha})\boldsymbol{\gamma} + \mathbf{z}_i\boldsymbol{\beta}. \tag{1.4}$$

The parameter vector $\boldsymbol{\theta}$ is estimated by minimizing

$$\begin{aligned} Q_{n,\lambda,\tau}(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i - q_\tau(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta})) + \lambda\boldsymbol{\gamma}^\mathsf{T}\mathbf{D}\boldsymbol{\gamma} \\ &= \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i - \mathbf{B}(\mathbf{x}_i\boldsymbol{\alpha})\boldsymbol{\gamma} - \mathbf{z}_i\boldsymbol{\beta}) + \lambda\boldsymbol{\gamma}^\mathsf{T}\mathbf{D}\boldsymbol{\gamma}, \end{aligned} \tag{1.5}$$

where $\rho_\tau(u) = |u| + (2\tau - 1)u$, $0 < \tau < 1$ is a check function. The second term is a roughness penalty term that alleviates possible overfitting. $\lambda \geq 0$ is a penalty parameter that controls the smoothness of the fitting curve, which will be further discussed in Section 1.2.4. $\mathbf{D}$ is some symmetric positive semi-definite matrix that can be properly chosen to yield the usual quadratic integral penalty. Alternatively, a simple choice of $\mathbf{D}$ is a diagonal matrix with the first $p+1$ elements being 0 and remaining diagonal elements as 1 (Yu and Ruppert 2002; Ruppert, Carroll Wand 2003). That is, $\boldsymbol{\gamma}^\mathsf{T}\mathbf{D}\boldsymbol{\gamma} = \sum_{l=1}^{K}\boldsymbol{\gamma}_{p+l}^2$ when truncated power basis is used.

1.2.3 A Check Function Approximation and Degrees of Freedom

The check function $\rho_\tau(\cdot)$ in the above objective function (1.5) is non-differentiable at the origin. The combination of non-differentiable check functions and the quadratic roughness penalty pose serious computational challenges. One natural idea is to approximate the check function with a differentiable square loss function near a small neighborhood of the origin. The approximated check function takes the form of

$$\rho_{\tau,\epsilon}(u) = \begin{cases} 2\tau u, & u > \epsilon \\ 2\tau u^2/\epsilon, & 0 \leq u < \epsilon \\ 2(1-\tau)u^2/\epsilon, & -\epsilon \leq u < 0 \\ -2(1-\tau)u, & u < -\epsilon \end{cases},$$

which only differs ρ_τ within the region $(-\epsilon, \epsilon)$. By setting ϵ small enough, a good smooth approximation for ρ_τ can be obtained. This idea is originated in Nychka et al. (1995), and adopted in Yuan (2006) and Li, Liu and Zhu (2007) in different contexts. The central idea of such approximation is that as in the least square mean regression an iterative reweighted penalized least square (IRPLS) algorithm can be easily developed to estimate the conditional quantile. The model "degrees of freedom" can be naturally obtained using the trace of "hat matrix" (e.g. Hastie and Tibshirani 1990; Yu and Ruppert 2002). With the degrees of freedom measure, smoothing parameter selection naturally follows using some information criterion or cross validation. It is worth noting that the "degrees of freedom" can be used as a measure of model complexity to facilitate further model comparison. To the best of our knowledge, it is not clear how to obtain model degrees of freedom measure for quantile regression using local methods (e.g. Wu et al. 2010). Alternatively, L_1 type check function along with L_1 type roughness penalty can be minimized using linear programming optimization (see e.g. Koenker, Ng and Portnoy 1994). Using such approach, however, the model degrees of freedom cannot be easily obtained.

In our implementation, when the check function $\rho_\tau(\cdot)$ is approximated with the differentiable function $\rho_{\tau,\epsilon}(\cdot)$, we can then take the first derivative of the objective function (1.5) with respect to $\boldsymbol{\theta}$.

$$\begin{aligned}\frac{\partial Q_{n,\lambda,\tau,\epsilon}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0 \quad &\Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} \rho'_{\tau,\epsilon}(y_i - q_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}))(-\mathbf{q}'_i) + 2\lambda \mathbf{P}\boldsymbol{\theta} = 0 \\ &\Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} 2(y_i - q_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}))\frac{\rho'_{\tau,\epsilon}(y_i - q_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}))}{2(y_i - q_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}))}(-\mathbf{q}'_i) + 2\lambda \mathbf{P}\boldsymbol{\theta} = 0,\end{aligned}$$

where $\mathbf{q}'_i$ is a column vector of the first derivative of $q_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, and $\mathbf{P}$ is a diagonal matrix whose first d_γ diagonal elements are the diagonal elements of $\mathbf{D}$ and the rest are 0's. While in the weighted least square regression with penalty, to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - q_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}))^2 W_i + \lambda \boldsymbol{\theta}^\mathsf{T} \mathbf{P}\boldsymbol{\theta},$$

its first order condition is

$$\frac{1}{n}\sum_{i=1}^{n} 2(y_i - q_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}))W_i(-\mathbf{q}_i') + 2\lambda \mathbf{P}\boldsymbol{\theta} = 0.$$

By comparing the two first order conditions, we can optimize (1.5) by iteratively solving a weighted least square problem with penalty, where the weights are given by

$$W_i = \frac{\rho'_{\tau,\epsilon}(y_i - q_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}))}{2(y_i - q_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}))}. \tag{1.6}$$

The "hat matrix" $\mathbf{H}_\lambda$ can be obtained in the last iteration through

$$\mathbf{H}_\lambda = \mathbf{V}(\mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{V} + n\lambda\mathbf{P})^{-1}\mathbf{V}^\mathsf{T}\mathbf{W}, \tag{1.7}$$

where $\mathbf{V}$ with its i-th row $\mathbf{V}_i = (\mathbf{B}(\mathbf{x}_i\boldsymbol{\alpha}), \mathbf{z}_i)$ is similar as the "design matrix" in the mean regression; and $\mathbf{W} = \mathrm{diag}(W_i)$ is the weight matrix with each diagonal element W_i given by (1.6). The conditional quantile can be estimated by

$$\widehat{\mathbf{q}}_\lambda = \mathbf{H}_\lambda \mathbf{y},$$

where $\mathbf{y}$ is the response vector.

When $\mathbf{X}$ is univariate or the single-index parameter is known, the model degrees of freedom can be obtained by $tr(\mathbf{H}_\lambda)$. Note that a computationally expedient formula is

$$tr(\mathbf{H}_\lambda) = tr(\mathbf{V}(\mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{V} + n\lambda\mathbf{P})^{-1}\mathbf{V}^\mathsf{T}\mathbf{W}) = tr((\mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{V} + n\lambda\mathbf{P})^{-1}\mathbf{V}^\mathsf{T}\mathbf{W}\mathbf{V}), \tag{1.8}$$

where only the trace of matrices of $dim(d_\gamma + d_z) \times dim(d_\gamma + d_z)$ need to be calculated.

1.2.4 Smoothing Parameter Selection

Selecting a suitable value of smoothing parameter is crucial to any good curve fitting. Koenker, Ng and Portnoy (1994) and He, Ng and Portnoy (1998) use Schwarz Information Criterion (SIC) in the smoothing spline quantile regression. In both papers, the number of points interpolated was used as a heuristic estimate of the effective model degree of freedom. It is worth noting that in the implementation of He, Ng and Portnoy (1998), penalized B-splines with smaller number of knots were used for practical computing reasons, which is in the same spirit of P-splines. Another popular smoothing parameter selection criterion in spline context is Generalized Cross Validation (GCV). Yuan (2006) studied Generalized Approximate Cross-validation (GACV) in the univariate smoothing spline quantile context. For our QPLSIM framework, with the effective degrees of freedom obtained in Section 1.2.2, smoothing or penalty parameter λ can be selected by minimizing:

$$\mathrm{SIC}(\lambda) = ln\Big[\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\{y_i - \widehat{q}_\lambda(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta})\}\Big] + \frac{ln(n)}{2n}df, \; or \tag{1.9}$$

$$\mathrm{GACV}(\lambda) = \frac{\sum_{i=1}^{n}\rho_\tau\{y_i - \widehat{q}_\lambda(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta})\}}{n - df}, \tag{1.10}$$

where the df is the effective model degrees of freedom, a measure of complexity of the model. Note that both smoothing parameter selection criteria depend on the degrees of freedom, which in turn depend on which smoothing parameter is selected. We implement and study in detail the degrees of freedom and compare model performances using the two different information criteria, as shown later in the simulation studies.

1.2.5 Algorithm

We develop an iteratively reweighted penalized least squares (IRPLS) algorithm for the proposed QPLSIM. We minimize $Q_{n,\lambda,\tau}(\boldsymbol{\theta})$ for a given λ and the optimum λ is chosen over a grid search by minimizing SIC or GACV. We summarize our estimation algorithm as follows.

Step 0: For a given τ and fixed λ, set an approximation threshold ϵ. In our implementation, ϵ is set to be effectively zero (e.g. $\epsilon = 10^{-4}$) relative to the data values. Initialize $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}^{\mathsf{T}}, \widehat{\boldsymbol{\gamma}}^{\mathsf{T}}, \widehat{\boldsymbol{\beta}}^{\mathsf{T}})^{\mathsf{T}}$.

(A) Initiate $\widehat{\boldsymbol{\alpha}}$. For example, linear quantile regression estimates for the model $q_\tau(y) = \mathbf{x}\boldsymbol{\alpha} + \mathbf{z}\boldsymbol{\beta}$ may be used. Normalize $\widehat{\boldsymbol{\alpha}}$ such that $\|\widehat{\boldsymbol{\alpha}}\| = 1$ and its first element is positive for identifiability.

(B) Given $\widehat{\boldsymbol{\alpha}}$, initialize $(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$. Given preliminary estimates of the index values $\{u_i = \mathbf{x}_i\widehat{\boldsymbol{\alpha}} : \ i = 1, \cdots, n\}$, minimize the penalized sum of squared errors $Q(\boldsymbol{\gamma}, \boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n \{y_i - \mathbf{B}(u_i)\boldsymbol{\gamma} - \mathbf{z}_i\boldsymbol{\beta}\}^2 + \lambda\boldsymbol{\gamma}^{\mathsf{T}}\mathbf{D}\boldsymbol{\gamma}$. This is equivalent to the ridge regression estimates $(\widehat{\boldsymbol{\gamma}}^{\mathsf{T}}, \widehat{\boldsymbol{\beta}}^{\mathsf{T}})^{\mathsf{T}} = (\mathbf{V}^{\mathsf{T}}\mathbf{W}\mathbf{V} + n\lambda\mathbf{P})^{-1}\mathbf{V}^{\mathsf{T}}\mathbf{W}\mathbf{y}$, where $\mathbf{W}$ is taken as the identity matrix. The spline basis $\mathbf{B}(u)$ may use equally spaced knots at sample quantiles of the index $u_1, \cdots, u_n$. Obtain initial quantile estimates $\widehat{\mathbf{q}} = (\widehat{q}_1, \ldots, \widehat{q}_n)^{\mathsf{T}}$ by (1.4).

Step 1: Iteratively reweighted solve for $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\beta}}$. Given $\widehat{\boldsymbol{\alpha}}$ and the current estimates $\widehat{\mathbf{q}}$ and $(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$, find the iteratively reweighted least square estimates by

$$(\widehat{\boldsymbol{\gamma}}^{\mathsf{T}(k+1)}, \widehat{\boldsymbol{\beta}}^{\mathsf{T}(k+1)})^{\mathsf{T}} = (\mathbf{V}^{\mathsf{T}}\mathbf{W}^{(k)}\mathbf{V} + n\lambda\mathbf{P})^{-1}\mathbf{V}^{\mathsf{T}}\mathbf{W}^{(k)}\mathbf{y},$$

where $\mathbf{W}^{(k)}$ is the weight matrix as in (1.6). Update $\widehat{\mathbf{q}}^{(k+1)}$ by (1.4) and $\mathbf{W}^{(k+1)}$ by (1.6). Iterate this step until sequence $(\widehat{\boldsymbol{\gamma}}^{(k)}, \widehat{\boldsymbol{\beta}}^{(k)})$ converges under some criterion.

Step 2: Given $(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$, minimize $Q(\boldsymbol{\alpha}; \boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_{i=1}^n \rho_{\tau,\epsilon}\{y_i - q(\boldsymbol{\alpha}; \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma})\}$ over $\boldsymbol{\alpha}$.

Step 3: Iterate Step 1 and 2 until $\widehat{\boldsymbol{\alpha}}$ and $(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ both converge under some criterion.

Remark:

1. In the initiation Step 0, we recommend trying various initial starting points. For example, an initial estimate can be generated randomly on the unit half sphere with

positive first coefficient. Without partially linear terms, Wu, Yu and Yu (2010) suggest using average derivative estimate (ADE) as the starting point. Alternatively, linear quantile coefficients may be used as starting value. Zhu, Huang and Li (2012) point out that under some linearity conditions, linear quantile estimator is consistent with the single-index quantile estimator.

2. In Step 2, the problem is a d-variate nonlinear minimization problem; we use fminsearch() from MATLAB's optimization toolbox. One could use other optimization software, for example, nls in S-PLUS or PROC NLIN in SAS.

3. Note that the knot locations and $\mathbf{V}$ change when the single-index estimates change between iterations.

4. In the implementation, unless indicated otherwise, we use 51 equally-spaced log-scaled grid points on $\log_{10}(\lambda) \in [-5, 5]$ for the smoothing parameter grid search. Quadratic spline with 20 equally spaced quantile knots are used.

1.2.6 Some Asymptotic Properties

As discussed in Yu and Ruppert (2002), we provide fixed-knot asymptotics for practical reasons. The following large sample properties for QPLSIM are established to facilitate further inferences.

Theorem 1. *If* $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\alpha}}^\top, \widehat{\boldsymbol{\gamma}}^\top, \widehat{\boldsymbol{\beta}}^\top\right)^\top$ *is the minimizer of (1.5), under the assumptions* $(A1)--(A4)$ *in Section 1.6, if the smoothing parameter* $\lambda_n = o(n^{-\frac{1}{2}})$, $\widehat{\boldsymbol{\theta}}$ *is a consistent estimator of* $\boldsymbol{\theta}_0$ *and converges to a normal distribution,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(0, \tau(1-\tau)\mathbf{J}\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Omega}_1^{-1}\mathbf{J}^\top), \tag{1.11}$$

where $\mathbf{J}$ *is the Jacobian matrix for reparameterization of the single-index parameters.* $\mathbf{J}$*,* $\mathbf{\Omega}_0$ *and* $\mathbf{\Omega}_1$ *are defined in detail in Section 1.6.*

1.2.7 Simulation Studies for QPLSIM

We show three simulation examples with various error distributions to study the finite sample properties of the proposed QPLSIM estimators.

Example 1 – homoscedastic normal error

Example 1 is a sine-bump model with symmetric homoscedastic errors. The true quantile function here is sum of a linear term and a nonlinear sine function on an index.

$$y = sin\Big(\frac{\pi(u-A)}{C-A}\Big) + z\beta + E, \tag{1.12}$$

where $u = \mathbf{x}\boldsymbol{\alpha}$; $A = \sqrt{3}/2 - 1.645/\sqrt{12}$, $C = \sqrt{3}/2 + 1.645/\sqrt{12}$; $\mathbf{x}$ i.i.d. $\sim$ trivariate Uniform(0,1), $z \sim$ N(0,1), x's and z are mutually independent; $E \sim$ N(0, 0.1^2); $\boldsymbol{\alpha} = (1,1,1)^{\mathsf{T}}/\sqrt{3}$, $\beta = 0.3$.

For a series of quantiles 10%, 30%, 50%, 70% and 90% respectively, we simulate 100 samples with each sample size $n = 200$. We use 20-knot quadratic splines. One can see from the estimation results in Table 1.1 that the algorithm described in Section 1.2.5 works well. The biases are negligible and the average estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are very close to the true values. This is confirmed in the boxplots of the 100 coefficient estimates for median quantile in Figure 1.1, where most estimates are centered around the true values with small deviations. From Table 1.1, one can also see that the Monte Carlo simulation variances are at similar scale as the estimated asymptotic variances calculated by (1.11).

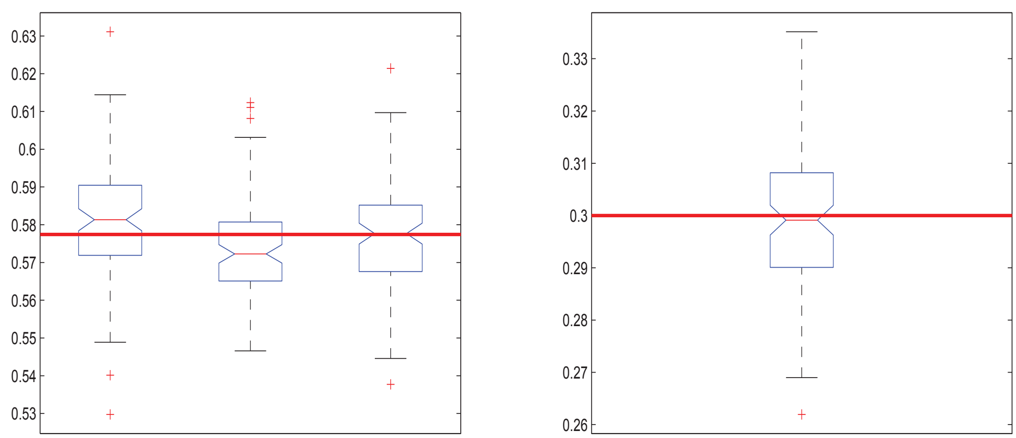
We also study the performance of SIC and GACV in choosing smoothing parameter λ. We consider the representative lower 10%, 30% and 50% quantiles for this symmetric

Table 1.1: Monte Carlo simulation study for partially linear single-index quantile regression

n=200	Estimate	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}$
$\tau = 0.1$	Average	0.5788	0.5775	0.5751	0.2984
	Asym. s.e.	0.0197	0.0228	0.0208	0.0146
	MC s.e.	0.0140	0.0179	0.0143	0.0144
	Bias	0.0014	0.0002	-0.0023	-0.0016
	MSE	0.0002	0.0003	0.0002	0.0002
$\tau = 0.3$	Average	0.5792	0.5749	0.5775	0.3018
	Asym. s.e.	0.0124	0.0129	0.0129	0.0105
	MC s.e.	0.0126	0.0136	0.0124	0.0078
	Bias	0.0018	-0.0024	0.0002	0.0018
	MSE	0.0002	0.0002	0.0002	0.0001
$\tau = 0.5$	Average	0.5764	0.5769	0.5783	0.2986
	Asym. s.e.	0.0142	0.0163	0.0149	0.0104
	MC s.e.	0.0112	0.0137	0.0137	0.0090
	Bias	-0.0009	-0.0005	0.0010	-0.0014
	MSE	0.0001	0.0002	0.0002	0.0001
$\tau = 0.7$	Average	0.5752	0.5769	0.5795	0.3011
	Asym. s.e.	0.0162	0.0160	0.0159	0.0105
	MC s.e.	0.0128	0.0123	0.0141	0.0090
	Bias	-0.0021	-0.0005	0.0022	0.0011
	MSE	0.0002	0.0002	0.0002	0.0001
$\tau = 0.9$	Average	0.5775	0.5750	0.5788	0.2988
	Asym. s.e.	0.0159	0.0170	0.0155	0.0117
	MC s.e.	0.0187	0.0138	0.0157	0.0118
	Bias	0.0002	-0.0023	0.0015	-0.0012
	MSE	0.0004	0.0002	0.0002	0.0001

The table reports Monte Carlo simulation estimates, true parameters, estimation bias, asymptotic standard errors, Monte Carlo standard errors and mean square errors (MSE). True parameter are: $\boldsymbol{\alpha}_0 \approx (0.5774, 0.5774, 0.5774)^\mathsf{T}$ and $\beta_0 = 0.3$.

Figure 1.1: Boxplots of single-index and partially linear coefficient estimates of QPLSIM for Simulation Example 1



(a) α (b) β

The plots report the boxplots of the estimates of single-index parameters ($\boldsymbol{\alpha}$) and partially linear coefficient parameters ($\boldsymbol{\beta}$). True parameters are: $\boldsymbol{\alpha}_0 \approx (0.5774, 0.5774, 0.5774)^{\mathsf{T}}$ and $\beta_0 = 0.3$ (horizontal lines).

distribution. To study the smoothing parameter at a finer scale, for each simulated sample, we find the λ's that minimize each criterion over the 80 equally spaced grids of $log_{10}(-5, 5)$. Define

$$MAD = \frac{\sum_{i=1}^{n} \rho_\tau\{y_i - \hat{q}_\lambda(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta})\}}{n - df} \quad and \quad MSE = \frac{\sum_{i=1}^{n} \{y_i - \hat{q}_\lambda(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta})\}^2}{n - df},$$

where *df* is the effective model degrees of freedom. Based on 100 simulations, we calculate the mean and standard deviations of MAD, MSE, and model degrees of freedom. The results are summarized in Table 2. Based on either MAD or MSE, one can see that GACV performs similar to SIC when sample size is small (n=200) and slightly better than SIC in the larger sample (n=1000) in this simulation example. The model degrees of freedom using GACV are larger than those of using SIC, which is also observed in Li, Liu and Zhu (2007). In another words, GACV tends to select more complex models than SIC does, which agrees with the general notion that SIC tends to select simpler models for its heavier penalty. For illustration purpose, we report results using SIC in the the following simulation studies and the Boston Housing data application.

Example 2 – heteroscedastic normal error

We consider a partially linear single-index model with heteroscedastic normal error. That is,

$$y = 10sin(0.75u) + \mathbf{z}\boldsymbol{\beta} + \sqrt{sin(u) + 1}E, \tag{1.13}$$

where $u = (x_1 + 2x_2)/\sqrt{5}$; x_1 and x_2 are i.i.d. $\sim N(0, 0.25^2)$; $\mathbf{z}\boldsymbol{\beta} = 0.3z_1 + 0.5z_2$; z_1 is i.i.d. $N(0, 1^2)$ and z_2 is binary variable with 0.4 probability of being 0 and 0.6 probability of being 1; $E \sim N(0, 1^2)$, independent of $\mathbf{x}$ and $\mathbf{z}$. We simulate 100 samples with sample size $n = 400$ each. Figure 1.2 shows boxplots of the single-index coefficient estimates for median quantile with 20-knot quadratic P-splines. One can see that the estimates

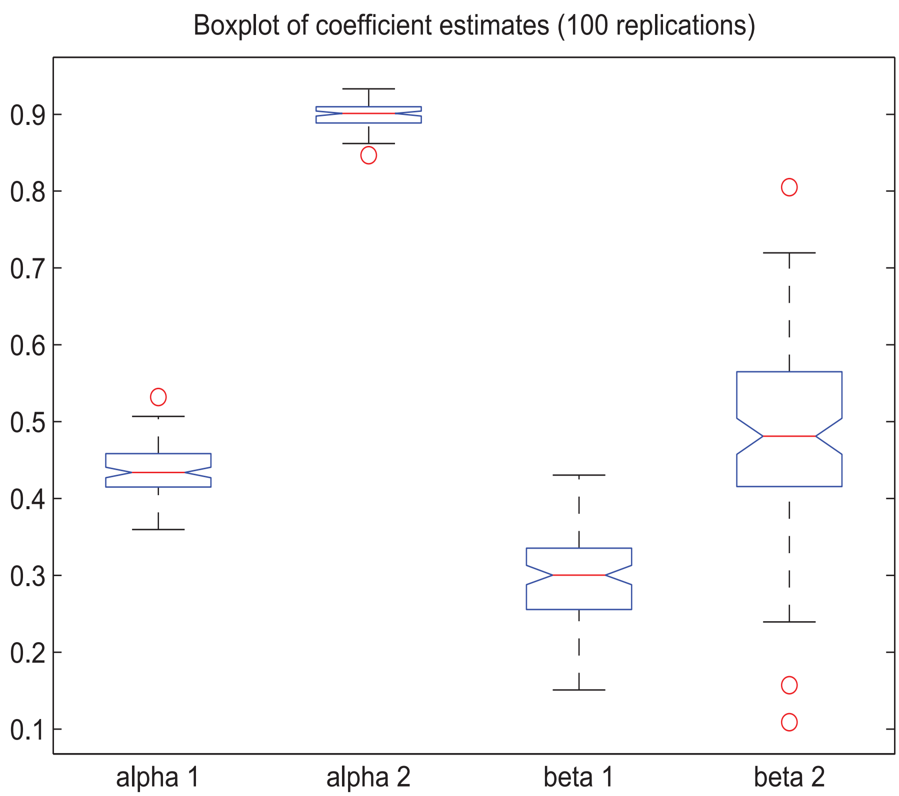
Table 1.2: GACV and SIC performance comparison

τ		$n = 200$		$n = 1000$	
		GACV	SIC	GACV	SIC
0.10	MAD	0.1388	0.1391	0.1349	0.1363
	s.e.(MAD)	(0.0090)	(0.0091)	(0.0044)	(0.0044)
	MSE	0.0257	0.0260	0.0260	0.0263
	s.e.(MSE)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
	d.f.	7.6894	7.0735	20.7290	12.6958
	s.e.(d.f.)	(0.5229)	(0.2442)	(3.8888)	(1.0599)
0.30	MAD	0.0948	0.0954	0.0893	0.0898
	s.e.(MAD)	(0.0059)	(0.0059)	(0.0025)	(0.0026)
	MSE	0.0125	0.0126	0.0126	0.0127
	s.e.(MSE)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
	d.f.	8.6347	7.9978	17.3686	12.5895
	s.e.(d.f.)	(0.3261)	(0.7476)	(4.0641)	(0.6516)
0.50	MAD	0.0840	0.0848	0.0788	0.0795
	s.e.(MAD)	(0.0046)	(0.0046)	(0.0024)	(0.0023)
	MSE	0.0098	0.0099	0.0099	0.0101
	s.e.(MSE)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
	d.f.	8.6311	7.6548	17.6300	12.8069
	s.e.(d.f.)	(0.4880)	(0.8083)	(4.0719)	(0.7601)

The table reports the comparison results on the performance measures of using GACV and SIC as the smoothing parameter selection criterion. MAD is the mean absolute error; MSE is the mean squared error; d.f. are the model degrees of freedom.

are close to and centered around the true values of $(1,2)^{\mathsf{T}}/\sqrt{5} \approx (0.4472, 0.8944)^{\mathsf{T}}$ with small deviations. The partially linear coefficients estimates are close to the true value of $(0.3, 0.5)^{\mathsf{T}}$ with somewhat larger variations.

Figure 1.2: Boxplots of single-index and linear coefficient estimates for Simulation Example 2



The plot reports the boxplots of the estimates of single-index parameters ($\boldsymbol{\alpha}$) and partially linear coefficient parameters ($\boldsymbol{\beta}$). True parameters are: $\alpha_0 \approx (0.4472, 0.8944)^{\mathsf{T}}$ and $\beta_0 = (0.3, 0.5)^{\mathsf{T}}$.

Example 3 – skew distribution

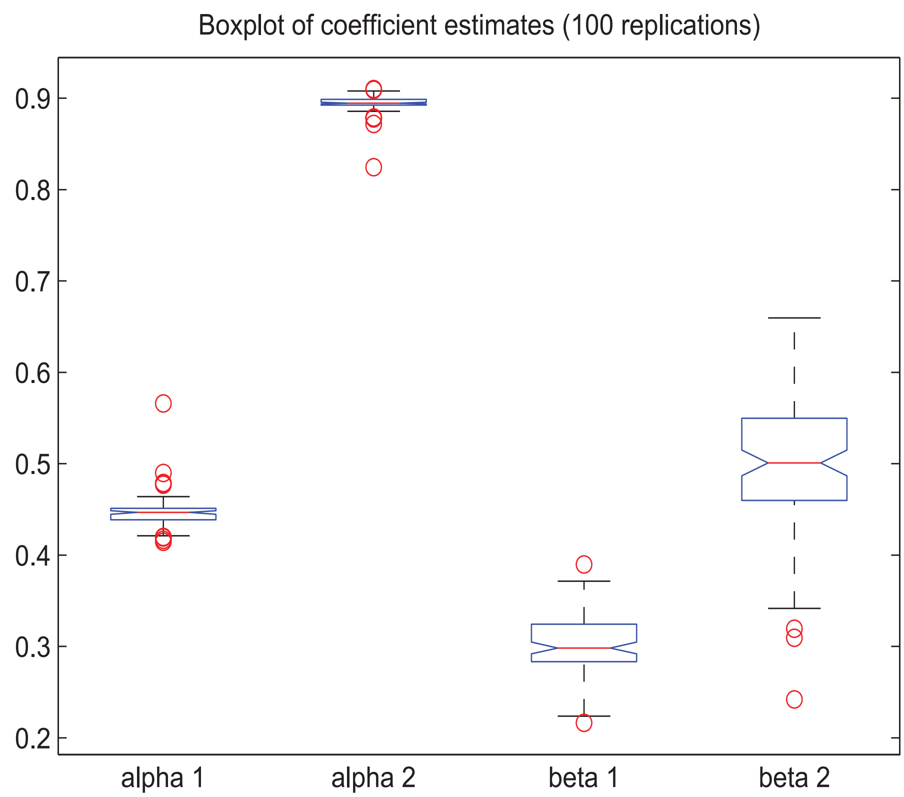
To show quantile regression is robust to skew distributions, we consider an example with asymmetric exponential error. That is,

$$y = 5\cos(u) + \exp(-u^2) + \mathbf{z}\boldsymbol{\beta} + E, \tag{1.14}$$

where $u = (x_1 + 2x_2)/\sqrt{5}$; x_1 and x_2 are i.i.d. $\sim N(0, 1^2)$; $\mathbf{z}\boldsymbol{\beta} = 0.3z_1 + 0.5z_2$; z_1 is i.i.d. $N(0, 1^2)$ and z_2 is binary variable with 0.4 probability of being 0 and 0.6 probability of being 1; $E \sim Exp(1)$, and independent of $\mathbf{x}$ and $\mathbf{z}$. Again we simulate 100 samples with each sample size $n = 400$. Figure 1.3 shows the boxplots of the 100 single-index coefficient estimates for the median quantile. Again one can see that the estimates are close to and centered around the true values of $(1, 2)^\mathsf{T}/\sqrt{5} \approx (0.4472, 0.8944)^\mathsf{T}$ with very small deviations. Again, the partially linear coefficients estimates are close to the true value of $(0.3, 0.5)^\mathsf{T}$ with somewhat larger variations.

We have also conducted simulation studies with different sample size and different noise level, the general observation is similar as above. Overall, we observe that the iteratively reweighted penalized least squares algorithm for the proposed partially linear single-index quantile regression model using P-splines is effective.

Figure 1.3: Boxplots of single-index and linear coefficient estimates for Simulation Example 3



The plot reports the boxplots of the estimates of single-index parameters ($\boldsymbol{\alpha}$) and partially linear coefficient parameters ($\boldsymbol{\beta}$). True parameters are: $\alpha_0 \approx (0.4472, 0.8944)^\top$ and $\beta_0 = (0.3, 0.5)^\top$.

1.3 Partially Linear Additive Conditional Quantile

Naturally, an alternative dimension reduction model to partially linear single-index conditional quantile model (QPLSIM) is partially linear additive conditional quantile models (QPLAM). Both proposed models avoid the "curse-of-dimensionality" by restricting the functional form to some extent, where only univariate unknown function(s) need to be estimated. We use similar notation for simplicity with hope that no confusion would occur.

1.3.1 Model and Estimation for QPLAM

Given $\tau \in (0,1)$ and n observations $\big\{(\mathbf{x}_i, \mathbf{z}_i, y_i)\big\}_{i=1}^{n}$ of $(\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, Y = y)$, the partially linear additive model we propose for the τ-th conditional quantile of y_i is

$$q_\tau(y_i|\mathbf{x}_i, \mathbf{z}_i) = a_\tau + g_{1,\tau}(x_{1,i}) + \cdots + g_{d,\tau}(x_{d,i}) + \mathbf{z}_i\boldsymbol{\beta}_\tau, \tag{1.15}$$

where a_τ is the intercept term; $g_{j,\tau}(\cdot)$ is univariate unknown function of x_j, $j = 1, \cdots, d$; $\boldsymbol{\beta}_\tau$ is unknown partially linear parameters. We need the conditions $E\{g_{j,\tau}(x_j)\} = 0$ $(j = 1, \cdots, d)$ to make each additive component identifiable (Hastie and Tibshirani 1990; De Gooijer and Zerom 2003; Yu and Lu 2004). Again we omit the subscript τ hereafter for clean notation.

Each function $g_j(\cdot)$ is estimated by P-splines, $g_j(x_j) = \mathbf{B}_j(x_j)\boldsymbol{\gamma}_j$, where $\mathbf{B}_j(x_j)$ is the spline basis for the j-th additive component. The flexibility of P-splines allows for different degrees and different numbers of knots of splines for each additive component. For example, the spline model for each additive component can be p_j degree with K_j knots, which will specify spline coefficients of $d_j = p_j + K_j + 1$ dimensions when truncated power basis is used. To further simplify our notation, define the "design matrix" $\mathbf{V} =$

$(\mathbf{V}_1^\mathsf{T}, \ldots, \mathbf{V}_n^\mathsf{T})^\mathsf{T}$, with its i-th row

$$\mathbf{V}_i = \Big(1 \quad \mathbf{B}_1(x_{1,i}) \quad \cdots \quad \mathbf{B}_j(x_{j,i}) \quad \cdots \quad \mathbf{B}_d(x_{d,i}) \quad \mathbf{z}_i\Big). \tag{1.16}$$

and parameter column vector $\boldsymbol{\theta} = \Big(a, \boldsymbol{\gamma}_1^\mathsf{T}, \cdots, \boldsymbol{\gamma}_d^\mathsf{T}, \boldsymbol{\beta}^\mathsf{T}\Big)^\mathsf{T}$. The τ-th conditional quantile can then be represented as

$$q(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}) = \mathbf{V}\boldsymbol{\theta}. \tag{1.17}$$

In QPLAM models, we can use separate smoothing parameters to allow different smoothness for different coefficient functions. Let $\lambda_j \geq 0$ be the penalty parameter for $g_j(\cdot)$. We minimize the penalized average check functions

$$Q_{n,\lambda,\tau}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{V}_i\boldsymbol{\theta}) + \sum_{j=1}^{d} \lambda_j \boldsymbol{\gamma}_j^\mathsf{T}\mathbf{D}_j\boldsymbol{\gamma}_j, \tag{1.18}$$

where $\rho_\tau(u)$ is the check function and $\mathbf{D}_j$ is similarly defined such that $\boldsymbol{\gamma}_j^\mathsf{T}\mathbf{D}_j\boldsymbol{\gamma}_j = \sum_{l=1}^{K_j} \boldsymbol{\gamma}_{p_j+l}^2$. To handle the multiple smoothing parameters, denote

$$\boldsymbol{\lambda} = diag(0, \lambda_1, \cdots, \lambda_1, \cdots, \lambda_d, \cdots, \lambda_d, 0, \cdots, 0)$$

and the corresponding penalty matrix $\mathbf{P} = diag(0, \mathbf{D}_1, \cdots, \mathbf{D}_d, 0, \cdots, 0)$ in such a way that matches the dimensions of the parameter $\boldsymbol{\theta}$. Note that λ_j may differ if we use separate smoothness for different additive components. For computational expediency, instead of conducting a full d-dimensional grid search of smoothing parameters, we select λ_j, $j = 1, 2, \ldots, d$ one at a time. That is, we first choose one common λ that minimizes the smoothing criterion (SIC or GACV), then λ_j is chosen with the rest of λ's are fixed.

Approximation to the above objective function and effective degrees of freedom calculations are carried out using the iteratively reweighted penalized least squares algorithm in the same manner as in QPLSIM. We propose the following simple algorithm for QPLAM.

Step 0: For given τ and $\boldsymbol{\lambda}$, set an approximation threshold ϵ. Start with an initial estimator $\widehat{\boldsymbol{\theta}}^{(0)} = (\hat{a}, \widehat{\boldsymbol{\gamma}}_1^\mathsf{T}, \cdots, \widehat{\boldsymbol{\gamma}}_d^\mathsf{T}, \widehat{\boldsymbol{\beta}}^\mathsf{T})^\mathsf{T}$ and obtain initial quantile estimates of $\widehat{\mathbf{q}}^{(0)}$

by (1.17). For example, estimates from mean partially linear additive model or linear quantile model can be used.

Step 1: Given the current estimate $\widehat{\mathbf{q}}^{(k)}$ and $(\hat{a}^{(k)}, \widehat{\boldsymbol{\gamma}}_1^{(k)}, \cdots, \widehat{\boldsymbol{\gamma}}_d^{(k)}, \widehat{\boldsymbol{\beta}}^{(k)})$, find the iteratively reweighted penalized least square estimates $(\hat{a}^{(k+1)}, \widehat{\boldsymbol{\gamma}}_1^{\mathsf{T}(k+1)}, \cdots, \widehat{\boldsymbol{\gamma}}_d^{\mathsf{T}(k+1)}, \widehat{\boldsymbol{\beta}}^{\mathsf{T}(k+1)})^{\mathsf{T}} = (\mathbf{V}^{\mathsf{T}}\mathbf{W}^{(k)}\mathbf{V} + n\boldsymbol{\lambda}\mathbf{P})^{-1}\mathbf{V}^{\mathsf{T}}\mathbf{W}^{(k)}\mathbf{y}$, where the weight matrix $\mathbf{W}^{(k)}$ is given by (1.6). Update $\widehat{\mathbf{q}}^{(k+1)}$ by (1.17).

Step 2: Iterate Step 1 until sequence $(\hat{a}^{(k)}, \widehat{\boldsymbol{\gamma}}_1^{(k)}, \cdots, \widehat{\boldsymbol{\gamma}}_d^{(k)}, \widehat{\boldsymbol{\beta}}^{(k)})$ converges under some criterion.

Effective degrees of freedom for P-spline QPLAM models are defined as $df = tr(\mathbf{H}_\lambda)$, where $\mathbf{H}_\lambda = \mathbf{V}(\mathbf{V}^{\mathsf{T}}\mathbf{W}\mathbf{V} + n\boldsymbol{\lambda}\mathbf{P})^{-1}\mathbf{V}^{\mathsf{T}}\mathbf{W}$ is from the last iteration of Step 1, and $\boldsymbol{\lambda}$ is chosen to minimize GACV or SIC. Again we use the computationally expedient formula for the trace calculation $tr(\mathbf{H}_\lambda) = tr((\mathbf{V}^{\mathsf{T}}\mathbf{W}\mathbf{V} + n\boldsymbol{\lambda}\mathbf{P})^{-1}\mathbf{V}^{\mathsf{T}}\mathbf{W}\mathbf{V})$.

For QPLAM, we also establish the following fixed-knot asymptotic results for inferences.

Theorem 2. *If* $\widehat{\boldsymbol{\theta}} = \left(\widehat{\alpha}, \widehat{\boldsymbol{\gamma}}_1^{\mathsf{T}}, \cdots, \widehat{\boldsymbol{\gamma}}_d^{\mathsf{T}}, \widehat{\boldsymbol{\beta}}^{\mathsf{T}}\right)^{\mathsf{T}}$ *is the minimizer of (1.18), under the assumptions* (A2) *and* (A4) *in Section 1.6, if the smoothing parameter* $\lambda_n = max\{\lambda_1, \cdots, \lambda_d\} = o(n^{-\frac{1}{2}})$, $\widehat{\boldsymbol{\theta}}$ *is a consistent estimator of the true* $\boldsymbol{\theta}_0$ *and converges to a normal distribution,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(0, \tau(1-\tau)\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Omega}_1^{-1}), \tag{1.19}$$

where $\boldsymbol{\Omega}_0$ *and* $\boldsymbol{\Omega}_1$ *are defined in Section 1.6.*

Note that computationally QPLAM is easier to implement in the sense that QPLAM is linear in its parameters of spline coefficients and partially linear terms. On the other hand, QPLSIM is nonlinear on the single-index parameters $\boldsymbol{\alpha}$. However, QPLAM assumes no interactions between covariates while in QPLSIM some interactions may be modeled.

1.3.2 A Simulation Study for QPLAM

We consider the model

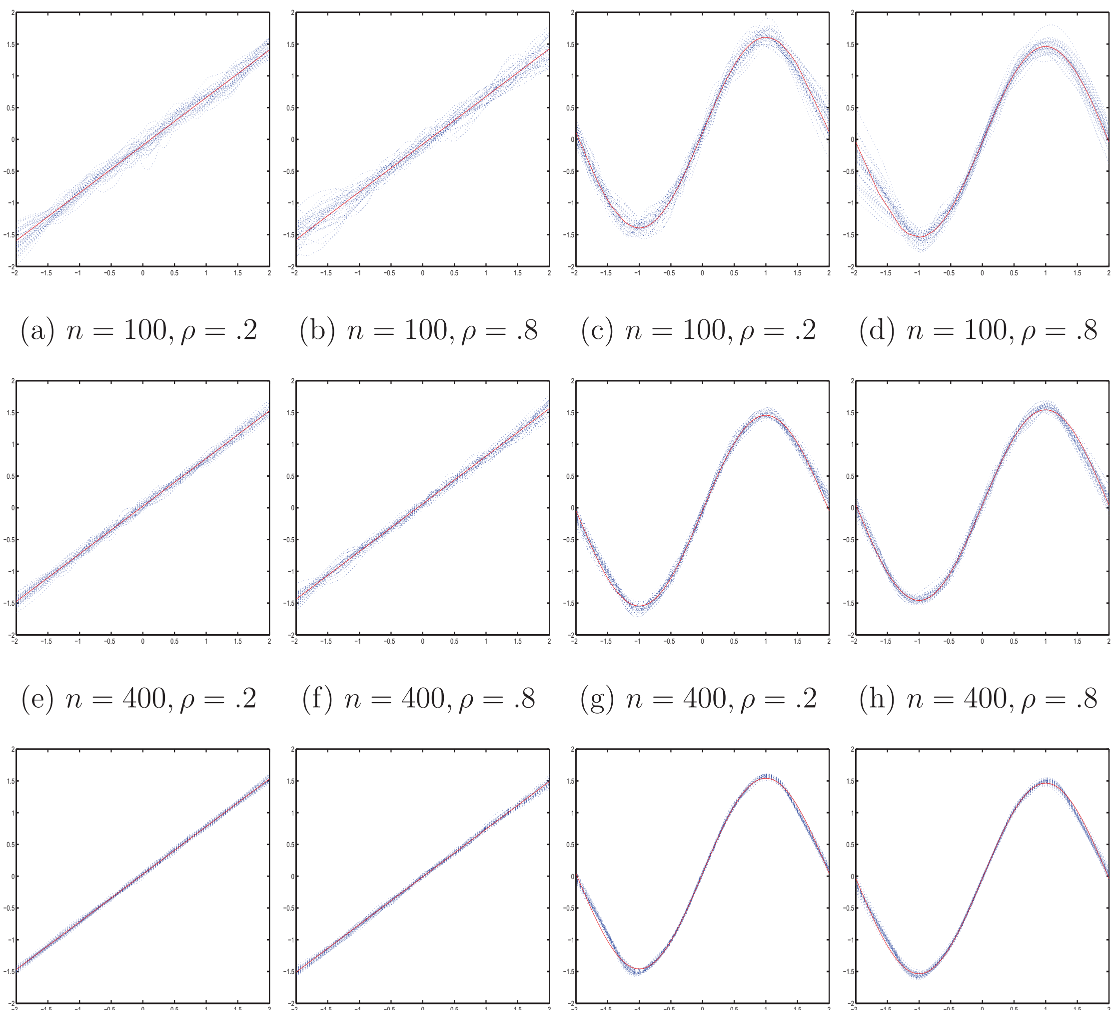
$$y = 0.75x_1 + 1.5sin(0.5\pi x_2) + E, \tag{1.20}$$

where covariates (x_1, x_2) are bivariate normal with mean 0, variance 1 and covariance ρ; $E \sim N(0, 0.5^2)$ and independent of (x_1, x_2). We conduct the simulation at two correlation levels between the covariates, low correlation of $\rho = 0.2$ and high correlation of $\rho = 0.8$. For comparison purpose, this example is intentionally designed to be identical to the simulation example originally conducted in De Gooijer and Zerom (2003) using marginal integration and backfitting and benchmarked in following studies, e.g. Horowitz and Lee (2005) with two-stage estimation. Note that additive models without partially linear terms were considered in De Gooijer and Zerom (2003) and Horowitz and Lee (2005), where the first linear term was treated as an additive component, even though it is in fact linear. We conduct Monte Carlo simulation experiments 41 times as the same in aforementioned studies for median quantile with sample sizes n=100, 200, 400, 800, and 1600 for comparison.

For comparison, we first fit our proposed P-spline additive quantile model treating both terms as additive components. The estimated median quantile curves are presented in Figure 1.4. One can see that the estimated curves are around the true median curves except for boundary points of g_2, where the bias is caused by fewer observations available. When sample size increases, the estimated curves center around the true median curves with less variations.

Naturally, we then fit our proposed P-spline partially linear additive quantile model (QPLAM) treating the first component as a partially linear term. The estimating performances of average absolute deviation errors (AADE) measure, defined as the average of

Figure 1.4: Estimated median quantiles



(a) $n = 100, \rho = .2$ (b) $n = 100, \rho = .8$ (c) $n = 100, \rho = .2$ (d) $n = 100, \rho = .8$

(e) $n = 400, \rho = .2$ (f) $n = 400, \rho = .8$ (g) $n = 400, \rho = .2$ (h) $n = 400, \rho = .8$

(i) $n = 1600, \rho = .2$ (j) $n = 1600, \rho = .8$ (k) $n = 1600, \rho = .2$ (l) $n = 1600, \rho = .8$

The plots report the estimated median quantiles for additive component $g_1(x_1)$ (first two columns), $g_2(x_2)$ (last two columns). The first and third columns are for $\rho = 0.2$ and second and fourth columns for $\rho = 0.8$. Solid lines are the true quantile curves and dotted lines are estimates.

absolute deviation errors (ADE, De Gooijer and Zerom 2003):

$$ADE_r(k) = \frac{1}{|\mathcal{N}'|} \sum_{i \in \mathcal{N}'} \left| \hat{g}_j^r(x_{j,i}^r) - g_j(X_{j,i}^r) \right|, j = 1, 2; r = 1, 2, \ldots, 41,$$

where $\mathcal{N}'$ is set of observations such that $x_{j,r} \in [-2, 2]$ to control for boundary effect.

The results shown in Table 1.3 indicate that our proposed P-spline estimating method has the most favorable performance in terms of AADE. P-spline QPLAM with the first component as a partially linear term yields the best AADE. Among the additive estimators, our P-splines estimators have slightly smaller AADE than the two-stage estimator (Horowitz and Lee 2005) for g_1 and significantly smaller AADE for g_2. Another important feature is that the P-splines estimator performs equally well for high correlation covariates. This feature is not observed in the marginal integration estimator. We have also noticed fast computation and rapid convergence of P-spline in this experiment.

Table 1.3: Monte Carlo simulation study for additive quantile regression

ρ	n	g_1 $_{(s.e.)}$ PS	g_2 $_{(s.e.)}$ PS	g_1 MI	g_2 MI	g_1 BF	g_2 BF	g_1 2S	g_2 2S	g_1 PS(lin)	g_2 PS
0.2	100	$0.0613_{(0.0277)}$	$0.0728_{(0.0227)}$	0.1374	0.1818	0.0597	0.1425	0.0783	0.1497	$0.0214_{(0.0145)}$	$0.0706_{(0.0242)}$
	200	$0.0465_{(0.015)}$	$0.0507_{(0.0149)}$	0.1066	0.1272	0.0511	0.1120	0.0519	0.1125	$0.0155_{(0.0082)}$	$0.0503_{(0.0147)}$
	400	$0.0362_{(0.0131)}$	$0.0387_{(0.0124)}$	0.0734	0.0936	0.0430	0.0889	–	–	$0.0090_{(0.0081)}$	$0.0350_{(0.0096)}$
	800	$0.0270_{(0.0075)}$	$0.0336_{(0.0078)}$	0.0625	0.0703	0.0264	0.0704	–	–	$0.0057_{(0.0043)}$	$0.0325_{(0.0072)}$
	1600	$0.0188_{(0.0044)}$	$0.0309_{(0.0061)}$	0.0523	0.0546	0.0196	0.0572	–	–	$0.0040_{(0.0035)}$	$0.0305_{(0.0050)}$
0.8	100	$0.0696_{(0.0292)}$	$0.0733_{(0.0301)}$	0.1365	0.4865	0.1124	0.1783	0.0920	0.1620	$0.0206_{(0.0217)}$	$0.0692_{(0.0281)}$
	200	$0.0580_{(0.0202)}$	$0.0576_{(0.0166)}$	0.1093	0.4350	0.1263	0.1767	0.0621	0.1146	$0.0207_{(0.0151)}$	$0.0555_{(0.0205)}$
	400	$0.0444_{(0.0103)}$	$0.0424_{(0.0093)}$	0.0985	0.4009	0.1099	0.1467	–	–	$0.0184_{(0.0126)}$	$0.0423_{(0.0113)}$
	800	$0.0277_{(0.0075)}$	$0.0418_{(0.0087)}$	0.0882	0.3690	0.0780	0.1124	–	–	$0.0097_{(0.0066)}$	$0.0398_{(0.0076)}$
	1600	$0.0222_{(0.0045)}$	$0.0313_{(0.0048)}$	0.0870	0.3330	0.0630	0.0908	–	–	$0.0084_{(0.0062)}$	$0.0307_{(0.0054)}$

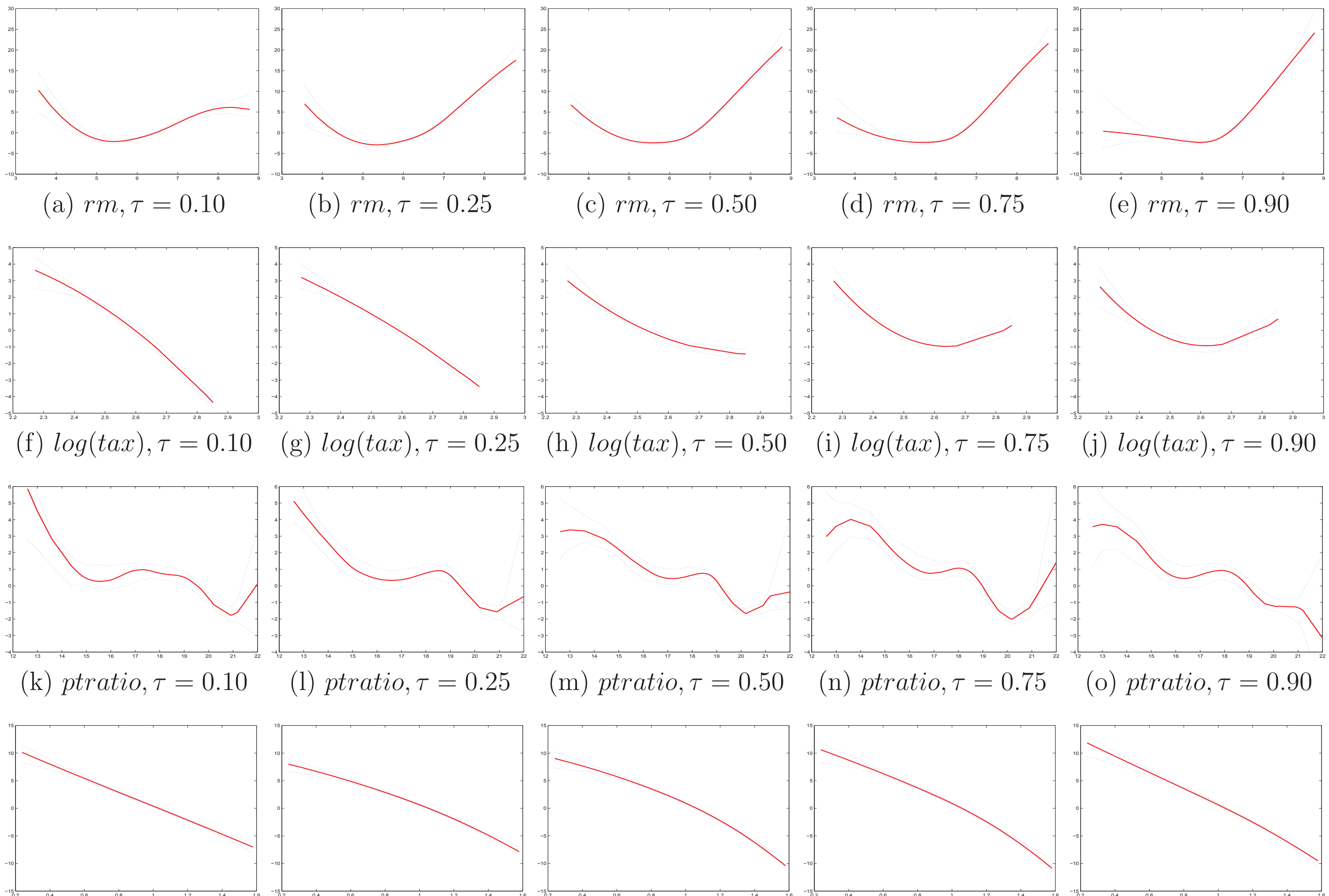
The values shown in the table are the average absolute deviation errors (AADE's) for penalized spline (PS), kernel based marginal integration (MI) and backfitting (BF) of De Gooijer and Zerom (2003), Two-stage estimators (2S) of Horowitz and Lee (2005). The last two columns are from penalized spline (PS), but with the first component g_1 specified as linear.

1.4 Boston Housing Data Application

We apply the partially linear single-index models (QPLSIM) and partially linear additive models (QPLAM) for conditional quantiles to the Boston housing data using P-splines. The dataset can be found on the site: http://lib.stat.cmu.edu/datasets/boston. The 506 observations measure the dependent variable *medv*, median value of owner occupied homes in \$1,000s in Boston suburb area in 1970's, and the other 13 independent variables.

Figure 1.5 displays the P-spline estimates of conditional quantile curves along with the 90% pointwise confidence bands for each component of the additive model with four covariates: *rm, log(tax), ptratio, log(lstat)* at τ = 10%, 25%, 50%, 75%, and 90%. To calculate standard errors for the estimates, we recommend using a simple bootstrap instead of the asymptotic results in real data applications because the asymptotic estimates are hard to calculate (De Gooijer and Zerom 2003; Wu, Yu and Yu 2010). Figure 1.5 indicates that there are clear nonlinearity in covariates across quantiles. The trend for each component is similar for all quantiles. For example, increasing pupil-teacher ratio ($ptratio$) or population of lower economical status ($lstat$) tends to have negative effect on housing price ($medv$) for all quantiles.

Figure 1.5: Conditional quantile curve fits using additive models for Boston Housing Data



(a) $rm, \tau = 0.10$ (b) $rm, \tau = 0.25$ (c) $rm, \tau = 0.50$ (d) $rm, \tau = 0.75$ (e) $rm, \tau = 0.90$

(f) $log(tax), \tau = 0.10$ (g) $log(tax), \tau = 0.25$ (h) $log(tax), \tau = 0.50$ (i) $log(tax), \tau = 0.75$ (j) $log(tax), \tau = 0.90$

(k) $ptratio, \tau = 0.10$ (l) $ptratio, \tau = 0.25$ (m) $ptratio, \tau = 0.50$ (n) $ptratio, \tau = 0.75$ (o) $ptratio, \tau = 0.90$

(p) $log(lstat), \tau = 0.10$ (q) $log(lstat), \tau = 0.25$ (r) $log(lstat), \tau = 0.50$ (s) $log(lstat), \tau = 0.75$ (t) $log(lstat), \tau = 0.90$

The plots report the estimated quantile curves of the additive components. rm average number of rooms per dwelling; $log(tax)$ is the log of full-value property-tax rate per USD 10,000; $ptratio$ is pupil-teacher ratio by town; $log(lstat)$ is the log of percentage of lower status of the population. Estimated τ-th (left to right: $\tau = .1, .25, .5, .75, .9$) quantile for additive component $g_j(x_j)$ versus x_j. Solid lines are the estimated quantile curves and dotted lines are the 90% confidence bands.

This Boston Housing data have been extensively studied in the quantile regression literature. For example, Chaudhuri et al. (1997) use average derivative estimates with three covariates: *rm, lstat, dis*. Wu, Yu and Yu (2010) consider single-index quantile regression using local linear method with four covariates: *rm, log(tax), ptratio, log(lstat).* Yu and Lu (2004) consider additive model with local linear method using the same four covariates. The aforementioned approaches exclude the remaining covariates in the models, which we refer as 3QSIM ADV, 4QSIM LL, and 4QAM respectively in the comparison Tables 1.5 and 1.6.

Following previous literature, we study various model fitting and conduct model comparisons using the same variables in nonlinear part. For example, adding partially linear terms to 4QSIM and 4QAM models results in 4QPLSIM and 4QPLAM respectively:

$$\begin{aligned} q_\tau\Big(medv\Big) &= g\Big(\alpha_1 rm + \alpha_2 log(tax) + \alpha_3 ptratio + \alpha_4 log(lstat)\Big) + \beta_1 crim + \beta_2 zn + \\ &\quad \beta_3 indus + \beta_4 chas + \beta_5 nox + \beta_6 age + \beta_7 dis + \beta_8 rad + \beta_9 black, \end{aligned} \tag{1.21}$$

$$\begin{aligned} q_\tau\Big(medv\Big) &= g_1(rm) + g_2(log(tax)) + g_3(ptratio) + g_4(log(lstat)) + \beta_1 crim + \beta_2 zn + \\ &\quad \beta_3 indus + \beta_4 chas + \beta_5 nox + \beta_6 age + \beta_7 dis + \beta_8 rad + \beta_9 black. \end{aligned} \tag{1.22}$$

The single-index coefficient estimates and bootstrap standard errors for model (1.21) are presented in Table 1.4. For the P-spline QPLSIM implementation, 20-knot quadratic P-splines are used. To save space, partially linear coefficients are not reported here. The signs of single-index coefficients are similar to Wu, Yu and Yu (2010), rm coefficients being positive while the other three coefficients being mostly negative. $log(tax)$ is not significant at $\tau = 0.1$ and 0.9, but is significant in middle quantiles.

Finally, we conduct comparisons among various models with different estimation methods studied in the literature. Due to lack of model complexity measure from the previous

Table 1.4: Single-index coefficient estimates of QPLSIM for the Boston Housing data

τ	$0.10_{(s.e.)}$	$0.25_{(s.e.)}$	$0.50_{(s.e.)}$	$0.75_{(s.e.)}$	$0.90_{(s.e.)}$
rm	$0.8680_{(0.0207)}$	$0.7318_{(0.0101)}$	$0.6717_{(0.0122)}$	$0.8962_{(0.0649)}$	$0.6509_{(0.0116)}$
log(tax)	$0.0431_{(0.0372)}$	$-0.4608_{(0.0317)}$	$-0.4830_{(0.0152)}$	$-0.0290_{(0.0557)}$	$0.0153_{(0.0097)}$
ptratio	$-0.0438_{(0.0096)}$	$-0.1041_{(0.0044)}$	$-0.1029_{(0.0030)}$	$-0.0703_{(0.0134)}$	$-0.0464_{(0.0078)}$
log(lstat)	$-0.4928_{(0.0449)}$	$-0.4913_{(0.0290)}$	$-0.5523_{(0.0217)}$	$-0.4370_{(0.0936)}$	$-0.7576_{(0.0297)}$

The table reports single-index coefficient estimates of QPLSIM and their standard errors for the Boston Housing data. rm average number of rooms per dwelling; $log(tax)$ is the log of full-value property-tax rate per USD 10,000; $ptratio$ is pupil-teacher ratio by town; $log(lstat)$ is the log of percentage of lower status of the population.

research, the average sum of check-function-based absolute residuals

$$R_\tau = \frac{\sum_{i=1}^{n} \rho_\tau\{y_i - \hat{q}(\mathbf{x}_i; \mathbf{z}_i; \boldsymbol{\theta})\}}{n}$$

is used for model comparison (Table 1.5). Our proposed P-spline approach with iteratively reweighted penalized least square algorithm naturally provides the effective model degrees of freedom. Hence, we can use the mean absolute deviations (MAD)

$$MAD = \frac{\sum_{i=1}^{n} \rho_\tau\{y_i - \hat{q}(\mathbf{x}_i; \mathbf{z}_i; \boldsymbol{\theta})\}}{n - df}$$

to incorporate model complexity measure in Table 1.6.

Table 1.5 shows that the proposed penalized spline approach for partially linear single-index models (QPLSIM) and partially linear additive models (QPLAM) is effective for modeling conditional quantiles of median housing price. Naturally, two partially linear models 4QPLSIM and 3QPLSIM with P-splines estimates seem to yield better results than models without partially linear terms (4QSIM and 3QSIM), regardless of estimation methods. This indicates that partially linear terms are useful to include in predicting median housing prices. Another observation is that QPLSIM and QPLAM models are comparable in terms of R_τ. Here R_τ provides a straightforward measure of goodness-of-fit across different models but it fails to consider the model complexity. Table 1.6 shows that

Table 1.5: Average sum of absolute residuals R_τ from various model fitting for Boston Housing data

τ	0.10	0.25	0.50	0.75	0.90
3QPLSIM PS	0.99	2.05	2.66	2.45	1.49
4QPLSIM PS	1.04	1.92	2.64	2.47	1.64
3QSIM PS	1.38	2.63	3.37	2.79	1.68
4QSIM PS	1.27	2.17	3.04	2.73	1.72
3QSIM LL	1.23	2.23	2.87	2.49	3.32
4QSIM LL	1.10	2.10	2.84	2.58	1.75
3QSIM ADV	1.56	2.70	3.04	2.43	3.13

The table reports the average sum of absolute residuals for various models with various estimation methods: PS=P-Splines, LL=Local Linear, ADV=Average Derivative.

incorporating the model degrees of freedom, models with partially linear terms (QPLSIM and QPLAM) have smaller mean absolute deviations than models without partially linear terms (QSIM and QAM). While adding partially linear terms results in more predictors and hence more complex models, the smaller mean absolute deviation indicates partially linear models have better fit even after model complexity is considered. QPLSIM and QPLAM models are comparable in terms of mean absolute deviations. Again, partially linear terms are shown to be useful to include in predicting median housing prices.

Table 1.6: Mean absolute deviations (MAD) for model goodness-of-fit comparison for Boston Housing data using P-Splines.

τ	0.10	0.25	0.50	0.75	0.90
3QPLSIM	1.04	2.12	2.76	2.45	1.56
4QPLSIM	1.07	1.97	2.70	2.54	1.69
3QSIM	1.40	2.65	3.44	2.86	1.72
4QSIM	1.28	2.20	3.08	2.76	1.75
3QPLAM	1.12	2.07	2.84	2.53	1.23
4QPLAM	0.99	2.03	2.69	2.38	1.52
3QAM	1.27	2.43	3.33	2.82	1.36
4QAM	1.11	2.25	3.03	2.72	1.65

The table reports mean absolute deviations for each penalized spline estimator.

1.5 Conclusion

This paper proposes partially linear single-index and partially linear additive modeling for conditional quantiles, both of which are flexible and effectively avoid the "curse of dimensionality". The unknown function is estimated using P-splines. We develop an iteratively reweighted penalized least squares algorithm for estimation. Model degrees of freedom and the performance of two smoothing parameter selection methods, namely GACV and SIC, are studied via a simulation. Some asymptotic properties are established. Simulation studies and application to Boston Housing data have shown the success of the two proposed models. Furthermore, it has been shown that partially linear terms are useful to include in modeling conditional quantiles.

1.6 Proofs

We present the proofs for QPLSIM in this section; proofs for the theorem for QPLAM are a combination of linear quantile and this proof.

We reparameterize the constrained $\boldsymbol{\alpha}$ to unconstrained $\boldsymbol{\psi}$, defined as $\boldsymbol{\psi} = (\psi_1, \cdots, \psi_{d-1})^\mathsf{T}$ such that $\boldsymbol{\alpha}(\boldsymbol{\psi}) = \left(1, \boldsymbol{\psi}^\mathsf{T}\right)^\mathsf{T} / \sqrt{1 + \|\boldsymbol{\psi}\|^2}$. Let $\boldsymbol{\theta}_{\boldsymbol{\psi}} = \left(\boldsymbol{\psi}^\mathsf{T},\ \boldsymbol{\gamma}^\mathsf{T},\ \boldsymbol{\beta}^\mathsf{T}\right)^\mathsf{T}$, which is one dimension lower than $\boldsymbol{\theta}_{\boldsymbol{\alpha}} = \left(\boldsymbol{\alpha}^\mathsf{T},\ \boldsymbol{\gamma}^\mathsf{T},\ \boldsymbol{\beta}^\mathsf{T}\right)^\mathsf{T}$. Then the τ-th quantile function becomes $q(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}_{\boldsymbol{\psi}}) = \mathbf{B}\Big(\mathbf{x}_i \boldsymbol{\alpha}(\boldsymbol{\psi})\Big)\boldsymbol{\gamma} + \mathbf{z}_i \boldsymbol{\beta}$. The Jacobian matrix of $\boldsymbol{\alpha}_{\boldsymbol{\psi}}$ is

$$\boldsymbol{\alpha}'(\boldsymbol{\psi}) = -(1+\|\boldsymbol{\psi}\|^2)^{-\frac{3}{2}} \begin{bmatrix} \psi_1 & \psi_2 & \cdots & \psi_{d-1} \\ -(1+\|\boldsymbol{\psi}\|^2) + \psi_1^2 & \psi_2\psi_1 & \cdots & \psi_{d-1}\psi_1 \\ \psi_1\psi_2 & -(1+\|\boldsymbol{\psi}\|^2) + \psi_2^2 & \cdots & \psi_{d-1}\psi_2 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1\psi_{d-1} & \psi_2\psi_{d-1} & \cdots & -(1+\|\boldsymbol{\psi}\|^2) + \psi_{d-1}^2 \end{bmatrix}$$

and the first derivative of $q_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is a column vector

$$\mathbf{q}'_i := \mathbf{q}'(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}_{\boldsymbol{\psi}}) = \Big[\mathbf{B}'\Big(\mathbf{x}_i\boldsymbol{\alpha}(\boldsymbol{\psi})\Big)\boldsymbol{\gamma} \cdot \mathbf{x}_i\boldsymbol{\alpha}'(\boldsymbol{\psi}),\ \mathbf{B}(\mathbf{x}_i\boldsymbol{\alpha}), \mathbf{z}_i\Big]^{\mathsf{T}}. \tag{1.23}$$

Note that for QPLAM, $\mathbf{q}'_i$ is defined by (1.16). The Jacobian matrix of $\boldsymbol{\theta}_{\boldsymbol{\alpha}}$ is

$$\mathbf{J} = \boldsymbol{\theta}'_{\boldsymbol{\alpha}}(\boldsymbol{\theta}_{\boldsymbol{\psi}}) = \begin{bmatrix} \boldsymbol{\alpha}'(\boldsymbol{\psi}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_\gamma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{d_z} \end{bmatrix},$$

which is a matrix of dimension $\dim(\boldsymbol{\theta}_{\boldsymbol{\alpha}}) \times (\dim(\boldsymbol{\theta}_{\boldsymbol{\alpha}}) - 1)$. After reparameterization, the estimation procedure can be done in the space of $\boldsymbol{\theta}_{\boldsymbol{\psi}}$ without the identifiability constraints.

We need the following assumptions to prove the theorems.

(A1) The parameter space Θ is compact, and the function $g(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta})$ is continuous on Θ for each fixed $(\mathbf{x}, \mathbf{z})$.

(A2) y_i are independently distributed with cumulative distribution function F_i. F_i is absolutely continuous with continuous density $f_i(\xi_i)$ uniformly bounded away from 0 and ∞ at the points $q(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}_0), i = 1, \ldots, n$.

(A3) There exist constants k_0 and k_1 such that when n is sufficiently large, for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ which is a compact subset in $\mathcal{R}^{dim(\boldsymbol{\theta})}$,

$$k_0\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \le \left(\frac{1}{n}\sum_{i=1}^{n}(q(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}_1) - q(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}_2))^2\right)^{1/2} \le k_1\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

(A4) There exist positive definite matrices $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$ such that

(a) $\boldsymbol{\Omega}_0 = \lim_{n\to\infty} \frac{1}{n}\sum \mathbf{q}'_i\mathbf{q}'^{\mathsf{T}}_i$.

(b) $\boldsymbol{\Omega}_1 = \lim_{n\to\infty} \frac{1}{n}\sum f_{y_i}(q(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}_0))\mathbf{q}'_i\mathbf{q}'^{\mathsf{T}}_i$, where $f_{y_i}(q(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}_0))$ is the density function for the i-th observation, evaluated at the τ-th conditional quantile.

(c) $\max_{i=1,\cdots,n} \|\mathbf{q}_i'\|/\sqrt{n} \to 0$.

Proof of Consistency: If $\lambda \to 0$, the bias tends to be 0 as $n \to 0$. The consistency follows from the combination of similar proof of Yu and Ruppert (2002) and results in §4.1 (117) of Koenker (2005).

Proof of Asymptotic Normality: If $\lambda_n = o(n^{-1/2})$, both theorems similarly follow from the combination of similar proof of Yu and Ruppert (2002) and the result in §4.4 (page 123) of Koenker (2005). The proof of Theorem 1 relies on Bahadur representation for the nonlinear parameter estimator and quadratic approximation to the minimization of objective function, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \boldsymbol{\Omega}_1^{-1} n^{-1/2} \sum_{i=1}^n \mathbf{q}_i' \phi_\tau(r_i) + o_P(1)$, where $r_i = y_i - q(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}_0)$ and $\phi_\tau(r) = \tau - I(r < 0)$. Variations of Bahadur representations by different authors are summarized in Koenker (2005).

Chapter 2

Time-Varying Beta and the Value Premium: Evidence from the Single-Index Varying-Coefficient Model

2.1 Introduction

Value stocks have substantially higher average returns than do growth stocks, although their loadings on the market risk are similar to, or smaller than, those of growth stocks. The value premium (e.g., Fama and French 1992; 1993 and 2006), the return difference between value and growth stocks, is one of the most prominent and challenging CAPM-related anomalies documented in the empirical asset pricing literature. An important risk-based explanation, as first advanced by Lettau and Ludvigson (2001), is that value stocks are riskier during business recessions when conditional risk premium is higher than during business expansions when conditional risk premium is low. Zhang (2005) elaborates on the conditional CAPM explanation of the value premium using a dynamic rational expectations model, in which the value premium has a countercyclical market beta because investment adjustment costs affect value and growth stocks differently across business cycles. In general, because there is no compelling reason for assuming constant betas, conditional factor models are arguably more appropriate than are their unconditional counterparts.

Extant studies commonly assume conditional beta as a linear function of some state variable(s).[1] Let $\{R_i\}_{i=1}^{n+1}$ and $\{R_{m,i}\}_{i=1}^{n+1}$ denote respectively the series of value premium and excess market return observed at $n+1$ time points, $0 < t_1 < t_2 < \cdots < t_{n+1}$. The value premium is modeled (e.g. Petkova and Zhang 2005) as

$$R_{i+1} = \alpha + \beta_{i+1} R_{m,i+1} + \epsilon_{i+1}, i = 1, \ldots, n \tag{2.1}$$

[1]There are a few notable exceptions, such as Wang (2003), Li and Yang (2011) and Ang and Kristensen (2011). Wang (2003) assumes fully nonparametric conditional beta on four state variables; Li and Yang (2011) and Ang and Kristensen (2011) both assume time varying beta, where conditional beta varies on time but not on conditional variables. The latter two studies can be viewed as a parallel approach to ours where the conditional beta is a function of state variable.

where $\beta_{i+1} = \mathbf{z}_i\boldsymbol{\gamma} = z_{1,i}\gamma_1 + \cdots + z_{d,i}\gamma_d$ is a linear combination of d-dimensional state variables $\mathbf{z} = (z_1, \ldots, z_d)$ at time i. In some other instances, α may be a linear combination of state variables (e.g. Lettau and Ludvigson 2001) as well.

In this paper, we set up a general framework in investigating whether the conditional CAPM helps explain the value premium using the single-index varying-coefficient model (Xia and Li 1999; Fan, Yao and Cai 2003; Wu, Lin and Yu 2011). Our empirical specification has two novel advantages relative to those commonly used in the previous studies, e.g., Shanken (1990), Ferson and Harvey (1991, 1999), Jagannathan and Wang (1996), Lettau and Ludvigson (2001), and Petkova and Zhang (2005). First, earlier authors assume that conditional beta is a linear function of some predetermined state variables, although such a relation could be rather complex (e.g., Ghysel (1998)). To avoid such a problem, we allow for a flexible dependence of conditional beta on an index that is a linear combination of state variables, i.e., $\beta_{i+1} = \beta(\mathbf{z}_i\boldsymbol{\gamma})$, where the functional form of $\beta(\cdot)$ is not specified *a priori* but estimated from data. We further allow our single-index varying-coefficient model to account for heteroskedasticity, where the varying error volatility is modeled by a flexible function of single-index of state variables through log P-splines. To the best of our knowledge, this is novel in terms statistical methodology in this application content. Take into consideration of heteroskedasticity is often necessary because the financial returns are well known to have such features. Single-index varying-coefficient modeling nests previous studies as special cases. This general setup allows us to test the validity of linear specification, and more importantly, to avoid misspecifications when the function is not linear. This single-index varying-coefficient setup also enjoys great interpretability. The index also has a straightforward economic interpretation; for example, it can be thought of as a composite measure of business conditions. The index coefficients share similar explanations as that in linear regressions when the nonparametric function is monotonic.

Second, we try to address the issue of variables selection in empirical CAPM research, by including a more comprehensive variables pool and conducting variable selection procedure in identifying the most significant variables. Earlier authors use standard stock market predictors as the state variables. Such a choice is arguably ad hoc because many of these variables have rather weak predictive power, especially in the out of sample context (Welch and Goyal (2008)). Specifically, many authors, Ghysels (1998) and Harvey (2001), caution that the estimation of conditional factor models is quite sensitive to the choice of state variables. Because economic theories emphasize a strong correlation of conditional beta with business cycles, we include closely watched gauges of aggregate economy as potential conditioning variables, in addition to commonly used stock market return predictors. Our comprehensive selection of state variables contains nine variables (economic variables and stock market return predictors, see Section 2.2 for details). With such moderate number of variables, our fast algorithm allows us to use a preferred exhaustive search for identifying the state variables that have most significant influence on conditional beta. Similar exhaustive variable selection approach can be found in Cremer (2002) where 14 predictor variables are included in trying to characterize stock return predictability.

Recent studies, e.g., Lewellen and Nagel (2006), Ang and Kristensen (2011), and Li and Yang (2011), have advocated using realized beta estimated from high-frequency return data as a measure of conditional beta. Because the estimation of realized beta does not depend on the choice of state variables, robustness is the main advantage of the realized beta model. The realized beta model, however, is not as efficient as an appropriately specified single-index varying-coefficient model because we can always include realized beta as a potential conditioning variable. Moreover, we find that some theoretically motivated state variables remain significant predictors of conditional beta when we include realized beta as a conditioning variable. This result is analogous to the stylized fact that realized volatility is not an efficient estimate of conditional volatility (e.g., Christensen and Prabhala (1998) and Guo and Whitelaw (2006)). Therefore, our specification allows for an improvement for the realized beta model because it provides a flexible way to incorporate information of both realized beta and state variables. Moreover, because realized beta is estimated using high-frequency, e.g., daily data, it is practically infeasible when factors are available at a lower frequency. For example, we cannot use realized beta to estimate the conditional consumption-based CAPM because consumption data are available at the quarterly frequency. In contrast, as we explore in this paper, it is straightforward to estimate the conditional consumption-based CAPM using the single-index varying-coefficient model.

We estimate the single-index varying-coefficient model using the monthly data spanning the July 1963 to December 2012 period. Following existing studies, we consider commonly used market return predictors, i.e., the default premium, the term premium, the stochastically de-trended risk-free rate, the dividend-price ratio, the realized market volatility and the price-earnings ratio, as potential conditioning variables. We include the industry production, the inflation rate, and the unemployment rate in the instrumental variable set as well to capture the dependence of conditional beta on business cycles.

Consistent with the risk-based explanation, we document strong countercyclical variation in the conditional beta for the value premium. Interestingly, the unemployment rate and inflation rate, the two macroeconomic variables watched most closely by the Federal Reserve, are identified as the significant conditioning variables for the conditional beta. More importantly, the conditional beta correlates negatively with the pro-cyclical inflation rate but positively with the countercyclical unemployment rate[2]. The pro-cyclical price-earnings ratio, which is also identified as a significant conditioning variable, correlates negatively with the conditioning beta. Other financial and macroeconomic variables, however, provide no additional information about conditional beta beyond these three variables. Overall, we reject the null hypothesis that the value premium has a constant conditional beta at the 1% level, and find that the fitted conditional beta increases sharply during economic recessions. The conditional CAPM does help explain the value premium. In our sample period, the value premium has an unconditional CAPM alpha of 5.6% per annum; however, when we take into account time-varying beta, the alpha decreases to 4.4% per annum. Nevertheless, the conditional CAPM does not fully explain the value premium because the alpha is economically large and statistically significant.

The conditional CAPM fails to fully explain the value premium possibly because some important conditioning variables are omitted from the instrumental variable set. As a robustness check, we include the monthly realized beta as a potential state variable because, as we discussed above, it is a robust measure of conditional beta. While realized beta is identified to be a significant conditioning variable, it does not subsume the information content of all financial and macroeconomic state variables that we consider. In particular, the inflation rate and price earning ratios remain significant conditioning variables and correlate negatively with conditional beta. Nevertheless, we find again that the conditional CAPM does not fully explain the value premium. As another robustness check, we also consider a smoothed realized beta, the average of realized monthly betas in the previous twelve months, and the results are found to be qualitatively robust. Similar findings are observed using quarterly data. Moreover, the value premium seems to vanish in the pre-1963 sample (January 1927 – June 1963). As a result, when we use the full monthly

[2]To the best of our knowledge, the finding of a close relation between the value premium's conditional beta and the key gauges of macroeconomic activity is novel.

sample from January 1927 to December 2012, the value premium is still significantly positive but reduces to 3.7% per annum, versus 4.4% in the post-1963 sample.

Recent studies, e.g., Petkova and Zhang (2005), Lewellen and Nagel (2006), Ang and Kristensen (2011), and Li and Yang (2011), find that the conditional CAPM does not fully explain the value premium. Interestingly, we confirm their findings using a more flexible empirical specification. Moreover, our results also shed some new light on the countercyclical variation in the conditional beta for the value premium. Overall, existing studies provide compelling evidence that alternative hypotheses are needed for the value premium, and our empirical findings also shed some light on potential explanations. We document a strongly countercyclical variation in the conditional beta for the value premium; specifically, it moves closely with the key gauges of aggregate economic activity. The finding, which casts doubt that behavioral explanations provide a whole explanation for the value premium, suggests risk-based explanations remain a viable alternative. The failure of the conditional CAPM reflects the fact that excess market returns are not an adequate measure of risk. For example, Campbell and Vuolteenaho (2004) show that discount-rate risk and cash-flow risk are priced differently. Similarly, investors might want to hedge against human capital risk, illiquidity risk, distress risk, and volatility risk, in addition to market risk. To the extent that these risks tend to move more closely to market risk during business downturns than during business upturns, a multi-factor model might provide a better explanation for the value premium than the CAPM.

The remainder of the paper proceeds as follows. We first discuss data in Section 2.2. We then explain the single-index varying-coefficient model in Section 2.3. We present empirical findings in Section 2.4. We offer some concluding remarks in Section 2.5.

2.2 Data

We obtain the monthly risk-free rate, excess market return (R_m), and value premium (R) data from Ken French at Dartmouth College. Following earlier studies, e.g., Ferson and Harvey (1999) and Petkova and Zhang (2005), we use standard stock market predictors as conditioning variables. The default premium (DEF) is the yield spread between Baa- and Aaa-rated corporate bonds. The term premium (TERM) is the yield spread between 10-year Treasury bonds and 3-month Treasury bills. The stochastically de-trended risk-free rate (RREL) is the difference between the risk-free rate and its average in the previous 12 months. The dividend-price ratio (DP) is the dividend paid in the most recent 12 months divided by the end-of-month stock market price. The price-earnings ratio (PE) is the end-of-month stock market priced divided by the earnings in the most recent 12 months. The realized market volatility (VOL) is the sum of squared daily excess market returns in a month. We obtain these data up to 2008 from Amit Goyal at Emory University and update them to 2012 using the data obtained from Ken French at Dartmouth College, Robert Shiller at Yale University, and the Federal Reserve Bank at St. Louis.

Because theoretical models suggest a close relation between the conditional beta of the value premium and business cycles, we also consider several major economic indicators as potential conditioning variables. The unemployment (UE) and inflation (INF) are arguably the most important gauges of aggregate economic activity. Specifically, because by legal mandates, the Federal Reserve is required to maintain full employment and price stability, it monitors closely and reacts promptly to the development in the job market and aggregate price indices. We obtain the civilian unemployment rate and the consumer price index for all urban consumers (all items) from the St. Louis Fed. Because the unemployment rate and the consumer price index are quite persistent, we use their year-over-year log changes as conditioning variables. Chen, Roll, and Ross (1986) and others find that industrial production (IP) is a priced state variable. For comparison, we include its year-over-year log change obtained from the St. Louis Fed as a conditioning variable. Lewellen and Nagel (2006) advocate the use of realized beta (BETAT) estimated from daily data as a proxy for conditional beta. Ang and Kristensen (2011) and Li and Yang (2011) offer some more elaborate estimators of the realized beta. For comparison, we follow Lewellen and Nagel (2006) and estimate realized beta of a month by regressing daily value premium on a constant and the excess market return in that month using the ordinary least squares (OLS) regression.

Table 2.1 provides summary statistics of the state variables. Interestingly, consistent with the notion that conditional beta of the value premium is countercyclical, we find that the realized beta, BETAT, a proxy for the conditional beta, correlates positively with the IP and UE and correlates negatively with INF. On the other hand, except for PE and TERM, the correlation of BETAT with stock market return predictors is rather weak. Similarly, while the macroeconomic variables correlate with the stock market return predictors, the correlation is not very strong. These results highlight the importance of including macroeconomic variables as the conditioning variables because they may provide additional information about conditional beta beyond financial variables.

Table 2.1: Summary Statistics and Correlations of Conditioning Variables

Panel A: Descriptive statistics of conditioning variables.

Variable	BETAT	DEF	TERM	RREL	DP	IP	INF	VOL	PE	UE	CYCLE
Mean	-0.19	0.01	0.02	0.00	0.03	0.03	0.04	0.00	0.20	0.01	0.15
Standard Error	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
Median	-0.21	0.01	0.02	0.00	0.03	0.03	0.03	0.00	0.20	-0.04	0.00
Standard Deviation	0.26	0.00	0.02	0.00	0.01	0.05	0.03	0.00	0.08	0.17	0.36
Sample Variance	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.13
Kurtosis	1.27	3.63	-0.38	2.52	-0.61	2.18	1.93	144.29	0.61	1.31	1.80
Skewness	0.50	1.60	-0.29	-0.11	0.37	-1.19	1.39	10.59	0.77	1.24	1.95
Range	1.98	0.03	0.08	0.01	0.05	0.28	0.16	0.07	0.38	0.93	1.00
Minimum	-1.10	0.00	-0.04	0.00	0.01	-0.16	-0.02	0.00	0.07	-0.34	0.00
Maximum	0.88	0.03	0.05	0.00	0.06	0.12	0.14	0.08	0.44	0.59	1.00

Panel B: Correlation matrix

	BETAT	DEF	TERM	RREL	DP	IP	INF	VOL	PE	UE	CYCLE
BETAT	1.00										
DEF	0.07	1.00									
TERM	0.23	0.23	1.00								
RREL	-0.05	-0.37	-0.53	1.00							
DP	-0.08	0.43	-0.22	0.05	1.00						
IP	-0.11	-0.63	-0.19	0.48	-0.15	1.00					
INF	-0.24	0.28	-0.42	0.19	0.71	-0.17	1.00				
VOL	0.02	0.25	0.07	-0.09	0.02	-0.19	0.03	1.00			
PE	-0.10	-0.51	-0.02	0.00	-0.90	0.18	-0.58	0.00	1.00		
UE	0.14	0.57	0.17	-0.51	0.16	-0.88	0.19	0.19	-0.19	1.00	
CYCLE	-0.05	0.43	-0.08	-0.33	0.30	-0.52	0.37	0.24	-0.23	0.54	1.00

The table reports descriptive statistics of conditioning variables (Panel A) and their correlation matrix (Panel B). The data sample is at a monthly frequency and spans July 1963 to December 2012. The NBER business recession indication (CYCLE) is also included.

2.3 The Single-Index Varying-Coefficient Model

In this paper, we propose an alternative model to (2.1), namely the single-index varying-coefficient model, where conditional beta can flexibly depend on state variables:[3]

$$R_{i+1} = \alpha(\mathbf{z}_i\boldsymbol{\gamma}) + \beta(\mathbf{z}_i\boldsymbol{\gamma})R_{m,i+1} + \sigma(\mathbf{z}_i\boldsymbol{\gamma})\epsilon_{i+1}, i = 1, \dots, n \tag{2.2}$$

where $\alpha(\cdot)$ and $\beta(\cdot)$ are both varying functions of the previous period single-index $\mathbf{z}_i\boldsymbol{\gamma}$; $\boldsymbol{\gamma}$ is the single-index parameters; ϵ_{i+1} is standard normal and $\sigma(\mathbf{z}_i\boldsymbol{\gamma})$ is the volatility, which is a smooth function of single-index and allows for heteroskedasticity; ϵ_{i+1} is independent of $\mathbf{z}_{s+1}$ and $R_{m,s+1}$ where $s < i$ and ϵ_{i+1} is independent of $\epsilon_{i'+1}$, $i \neq i'$. We need to estimate unknown functions $\alpha(\cdot)$, $\beta(\cdot)$ and $\sigma(\cdot)$, as well as single-index parameter $\boldsymbol{\gamma}$. To our best knowledge, the volatility specification in model (2.2) is novel. Because of the nonlinear dependence of the volatility on single-index parameters, the estimation can be quite challenging. We use a two step estimation algorithm where log-spline is utilized in estimating the volatility function. The estimation details are described later. In statistical literature on single-index and related models, to ensure the identifiability of single-index coefficients, the constraints of unit norm and positive first component are commonly imposed, i.e., $\|\boldsymbol{\gamma}\| = 1$ and $\gamma_1 > 0$ (see e.g. Carroll, Fan, Gijbels and Wand 1997; Xia and Li 1999; Yu and Ruppert 2002).

[3] Another alternative model specification to (2.1) is a fully nonparametric one. To allow maximum amount of flexibility, conditional beta can be modeled as a fully nonparametric function of a number of state variables:

$$R_{i+1} = \alpha(\mathbf{z}_i) + \beta(\mathbf{z}_i)R_{m,i+1} + \epsilon_{i+1}, i = 1, \dots, n,$$

where $\alpha(\cdot)$ and $\beta(\cdot)$ are fully nonparametric functions of d-dimensional state variable vector $\mathbf{z}_i$. This fully nonparametric specification allows for maximum flexibility. However the estimation of the above model suffers from the "curse of dimensionality" when the dimension of $\mathbf{z}_i$ is moderately high (four or higher). Even though estimation of fully nonparametric function might be feasible, the interpretation suffers from such multi-dimensional dimensional specification.

Model (2.2) is called single-index varying-coefficient model because single-index functions $\alpha(\cdot)$ and $\beta(\cdot)$ are acting like varying coefficients. Modeling conditional beta as a single-index function of state variables has certain desirable properties, both in statistics and finance. First of all, single-index models overcome the "curse of dimensionality" and have easy interpretability. Estimation of single-index models is considerably easier than fully nonparametric models because one-dimension nonparametric functions need to be estimated. The single-index varying-coefficient model also enjoys straightforward interpretation: single-index may be thought of a composite measure of business conditions; in the presence of a monotonic function (e.g. $\beta(\cdot)$), as pointed out by Li (1991) and summarized by Carroll, Fan, Gijbels and Wand (1997), the single-index coefficients share similar interpretations as that of the coefficients in ordinary linear regressions. Moreover, single-index varying-coefficient modeling of conditional beta follows the same rationale of previous studies (Lettau and Ludvigson 2001; Petkova and Zhang 2005), where conditional beta is directly conditioned on state variables. Last but not least, this model is quite general in the sense that some previous models on conditional CAPM can be viewed as the special case. For example, when $\alpha(\cdot)$ is constant and $\beta(\cdot)$ is identity functions, model (2.2) reduces to previous models such as Lettau and Ludvigson (2001) and Petkova and Zhang (2005).[4]

We estimate the univariate varying coefficients $\alpha(\cdot)$ and $\beta(\cdot)$ in model (2.2) using pe-

[4]Ang and Kristensen (2011) also uses nonparametric approach to model conditional beta where beta is modeled as time varying. Our approach of modeling beta as functions of state variables at the previous time period is consistent with previous literature. Our estimation method can be used to estimate the conditional beta in Ang and Kristensen (2011), but not vice versa. Although time can be treated as a state variable, thus our model reduces to Ang and Kristensen (2011), there is a fundamental difference in the approaches of modeling conditional beta. Our approach assumes there are economic or financial or any other measurable variables driving the conditional beta, while the approach of Ang and Kristensen (2011) does not explicitly take driving variables into consideration.

nalized splines (P-splines). Because the volatility function $\sigma(\cdot)$ is always positive, we use penalized splines to estimate the logarithmic transformation of the volatility, or equivalently use log-spline to estimate the volatility function. P-splines have gained popularity recently because of its expedient and stable computations, natural link to ridge regression, mixed models and Bayesian models, and the ability to widely extend to more complicated studies; see Yu and Ruppert (2002), Crainiceanu, Ruppert, Claeskens and Wand (2003), Ruppert, Wand and Carroll (2003), and Jarrow, Ruppert and Yu (2004) for examples. In the nonparametric literature, local methods for single-index models (e.g., Carroll, Fan, Gijbels and Wand, 1997) have been extensively studied but the computation may become unstable while Penalized splines are computationally stable and expedient (Yu and Ruppert 2002). Another reason why P-spline is the preferred estimation method for $\alpha(\cdot)$ and $\beta(\cdot)$ is because the computational expedience of P-splines allows for exhaustive variable selection, which is another major issue (Ghysels 1998; Harvey 2001) we are trying to address. Fast estimation is desirable because the number of combinations of state variables will grow exponentially with the number of state variables (2^d combinations for d-variate state variables, to be exact). In our experience, the computational expediency of penalized spline estimation allows us to find the optimal model in a relatively short time, even if the dimension of potential state variables is relatively large.

Once the best subset of state variables is identified by exhaustive variable selection procedure, the model with the best set of state variables is fitted and tested. The test statistic is established based on the asymptotic properties of the estimates, which is discussed in section 2.3.3. There are a few hypotheses worth testing. Is β constant or varying? Constant beta indicates static CAPM while varying beta suggest conditional beta model. If β is varying, then is β linear in single-index? Linear functional form provides justifications in modeling approach in Lettau and Ludvigson (2001) and Petkova and Zhang (2005), while nonlinear functional form of beta warns that linear specification should be used in caution. We have found mixed results in the functional form of conditional beta, dependent on the choice of state variables, which will be discussed later. Another interesting test is the functional form and positivity of α. We find the functional form of α is often constant, which justifies the practice in Petkova and Zhang (2005). However, conditional alpha together with long-run alpha, which is the average conditional alpha are both significantly positive, which suggests conditional CAPM model cannot fully explain value premium. Although earlier sample (before 1963 period) results in conditional alpha or long-run alpha not significant different from zero, consistent with existing findings (e.g. Ang and Chen 2005), full sample from 1927 to 2012 leads qualitatively robust results.

Our approach is most related to Wang (2003), Ang and Kristensen (2011) and Li and Yang (2011), where nonparametric methods are also used. There are major differences in our approach. Wang (2003) assumes conditional beta as a fully nonparametric function of four-dimensional state variables while our single-index approach reduces the nonparametric dimensions to just one. The advantage of dimension reduction is the gain in easier interpretation and estimation robustness. The higher dimension state variables in our study also requires us to practically reduce the dimensionality rather than to assume fully nonparametric form. Ang and Kristensen (2011) and Li and Yang (2011) assumes conditional beta as a time varying function, hence no state variables are involved. Moreover, our estimation method differs from the aforementioned methods, where local polynomials (or kernel methods) are used. The penalized spline approach in this paper is computationally expedient and robust, suitable for the exhaustive variable selection procedure. We first introduce our estimation method with smoothing parameter selection, then proceed with variable selection and statistical tests in the following.

2.3.1 Estimation

There are three types of parameters to be estimated: varying coefficient functions $\alpha(\cdot)$ and $\beta(\cdot)$, volatility function $\sigma(\cdot)$, and the single-index parameters $\boldsymbol{\gamma}$. We introduce a an iterative algorithm in finding all these estimators. We first introduce the estimation of flexible functions $\alpha(\cdot)$, $\beta(\cdot)$ and $\sigma(\cdot)$, assuming known single-index parameter $\boldsymbol{\gamma}$. We then introduce a fixed point algorithm in updating the single-index parameters.

P-splines Estimation of Flexible Functions

Assuming known single-index parameter $\boldsymbol{\gamma}$ or equivalently single-index $u = \mathbf{z}\boldsymbol{\gamma}$, both varying coefficient $\alpha(u)$ and $\beta(u)$ are univariate flexible function, which can be estimated using penalized splines (P-splines). To handle the positivity constraint, the volatility function $\sigma(u)$ is estimated using log penalized splines–the logarithmic transformation of the volatility function is modeled using P-splines. P-splines are employed for their fast computation and robustness. We use truncated power functions for its easy interpretation and straightforward testing. The truncated power basis of degree p, with knots at $v_1, \dots, v_k$, is $\mathbf{B}(u) = (1, u, u^2, \dots, u^p, (u - v_1)_+^p, \dots, (u - v_k)_+^p)$, where $(u - v)_+^p = (u - v)^p I(u - v)$ is the truncated polynomial function with a break (also known as a knot in P-splines) at v. Any function $f(u)$ with $p - 1$ continuous derivatives can be approximated by

$$f(u) = \delta_0 + \delta_1 u + \delta_2 u^2 + \ldots + \delta_p u^p + \delta_{p+1}(u - v_1)_+^p + \ldots + \delta_{p+k}(u - v_k)_+^p = \mathbf{B}\boldsymbol{\delta}_f,$$

where $\boldsymbol{\delta}_f = (\delta_0, \delta_1, \dots, \delta_{p+k})^\mathsf{T}$ is the spline coefficient vector. The shape of $f(u)$ is determined by degree p, number and locations of knots. Higher degrees P-splines result in smoother curves but more computational requirements. In practice, degree of two or three are used for estimating functions. Various studies (e.g. Ruppert 1998; Ruppert, Wand and Carroll 2003) have also shown that as long as enough number of knots are used, the number and location of knots are no long crucial. For example, knots are selected at the equidistant quantiles of the predictor variable. The possible overfitting will be avoided by adding some proper roughness penalty, where one single smoothing parameter can control the smoothness of the curve. In particular, adding a square roughness penalty to the usual least square loss function result in a penalized least squares objective function, $\sum_{i=1}^{n}\big(f(u) - \mathbf{B}\boldsymbol{\delta}_f\big)^2 + \lambda n(\delta_{p+1}^2 + \ldots + \delta_{p+k+1}^2)$, where λ controls the tradeoff of goodness of fit and the roughness penalty.

To estimate $\alpha(u)$ and $\beta(u)$ in (2.2), denote the basis $\mathbf{B}_a$ and $\mathbf{B}_b$ and the parameter vectors $\boldsymbol{\delta}_a$ and $\boldsymbol{\delta}_b$ for conditional alpha and beta respectively, then i-th conditional alpha is $\alpha(u) = \mathbf{B}_{a,i}\boldsymbol{\delta}_a$ and conditional beta $\beta(u) = \mathbf{B}_{b,i}\boldsymbol{\delta}_b$. Combine the scalar 1 and $R_{m,i+1}$ with spline bases, we can write the i-th row of "design" matrix $\mathbf{X}$ as $\mathbf{X}_i = (\mathbf{B}_{a,i}, \mathbf{B}_{b,i}R_{m,i+1})$. Define spline coefficient parameters $\boldsymbol{\delta}_1 = (\boldsymbol{\delta}_a^{\mathsf{T}}, \boldsymbol{\delta}_b^{\mathsf{T}})^{\mathsf{T}}$, the mean function of R_{i+1} $(i = 1, 2, \ldots, n)$ can be written as

$$m_i \equiv E(R_{i+1}) = \mathbf{B}_{a,i}\boldsymbol{\delta}_a + \mathbf{B}_{b,i}R_{m,i+1}\boldsymbol{\delta}_b = \mathbf{X}_i\boldsymbol{\delta}_1.$$

Similarly, the varying volatility function can be approximated using $log\sigma(u_i) = \mathbf{B}_2(u_i)\boldsymbol{\delta}_2$, or equivalently $\sigma(u_i) = exp\{\mathbf{B}_2(u_i)\boldsymbol{\delta}_2\}$ (Yu, Yu, Wang and Li 2009). The log likelihood function, excluding constants, can be written as the negative

$$\sum_{i=1}^{n}\Big(exp\{-2\mathbf{B}_2(u_i)\boldsymbol{\delta}_2\}\{R_{i+1} - \mathbf{B}_{a,i}\boldsymbol{\delta}_a - \mathbf{B}_{b,i}R_{m,i+1}\boldsymbol{\delta}_b\}^2 + 2\mathbf{B}_2(u_i)\boldsymbol{\delta}_2\Big).$$

For notation convenience, we reserve subscript 1 for (flexible) mean functions and 2 for (flexible) volatility function. The above P-spline estimation allows for different knots and degrees for different flexible functions. Different smoothness could also be reached by using separate smoothing parameters. One popular way of adding separate penalties (e.g. Ruppert 2002; Yu and Ruppert 2002; Ruppert, Wand and Carroll 2003) is the so-called

L_2 penalty which will yield penalized log likelihood function

$$\sum_{i=1}^{n}\Big(exp\{-2\mathbf{B}_2(\mathbf{z}_i\boldsymbol{\gamma})\boldsymbol{\delta}_2\}\{R_{i+1}-\mathbf{B}_{a,i}(\mathbf{z}_i\boldsymbol{\gamma})\boldsymbol{\delta}_a-\mathbf{B}_{b,i}(\mathbf{z}_i\boldsymbol{\gamma})\boldsymbol{\delta}_b R_{m,i+1}\}^2+2\mathbf{B}_2(\mathbf{z}_i\boldsymbol{\gamma})\boldsymbol{\delta}_2\Big)$$
$$+\frac{n}{2}\lambda_a\boldsymbol{\delta}_a^{\mathsf{T}}\mathbf{D}_a\boldsymbol{\delta}_a+\frac{n}{2}\lambda_b\boldsymbol{\delta}_b^{\mathsf{T}}\mathbf{D}_b\boldsymbol{\delta}_b+\frac{n}{2}\lambda_2\boldsymbol{\delta}_2^{\mathsf{T}}\mathbf{D}_2\boldsymbol{\delta}_2, \tag{2.3}$$

where $\lambda_a > 0$, $\lambda_b > 0$ and $\lambda_2 > 0$ are separate penalty parameters for varying coefficients functions and volatility function respectively; $\mathbf{D}_\xi$ is some appropriate 0-1 diagonal matrices such that $\boldsymbol{\delta}_\xi^{\mathsf{T}}\mathbf{D}_\xi\boldsymbol{\delta}_\xi = \sum_{j=1}^{k_\xi}\delta_{p_\xi+j}^2$, where k_ξ and p_ξ ($\xi = a, b, 2$) are the number and degree of knots for different spline coefficients respectively. One could optimize the penalized log likelihood function (2.3) in one step. However, the number of parameters could be large and the estimation algorithm may not be efficient. Instead, we propose an two-step algorithm which reweights the mean function using the estimates of volatility function, in a fashion similar to weighted least squares. The algorithm first estimates the mean function then the volatility is estimated after. Parameters in mean function are thereafter recalculated using weighted least squares where the weights are the inverse of the estimated volatility. We find two or three iterations are sufficient. Similar treatments of iteratively estimating mean and volatility can be found in Carroll, Wu and Ruppert (1988) and Yu et al. (2009). The two-step estimation procedure is described as follows:

Step 1: Mean Estimation.

First to get an initial estimation of the mean function,[5] the spline coefficients can be calculated using a linear shrinkage estimator $\widehat{\boldsymbol{\delta}}_1 = (\mathbf{X}^{\mathsf{T}}\mathbf{X} + n\boldsymbol{\lambda}\mathbf{D}_1)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{R}$, where $\boldsymbol{\lambda}_1 = blockdiag(\boldsymbol{\lambda}_a, \boldsymbol{\lambda}_b) = diag(\lambda_a, \ldots, \lambda_a, \lambda_b, \ldots, \lambda_b)$ is penalty parameter matrix and $\mathbf{D}_1 = blockdiag\{\mathbf{D}_a, \mathbf{D}_b\}$ and $\mathbf{R} = (R_2, \ldots, R_{n+1})^{\mathsf{T}}$ is the vector of value premiums. Estimated value premium is $\widehat{\mathbf{R}} = \mathbf{X}\widehat{\boldsymbol{\delta}}_1 = \mathbf{H}\mathbf{R}$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathsf{T}}\mathbf{X} + n\boldsymbol{\lambda}_1\mathbf{D}_1)^{-1}\mathbf{X}^{\mathsf{T}}$ is the smoothing matrix.

[5] Alternatively, in the presence of homoscedasticity, the weighting procedure is unnecessary. The second step and the following reweighting can be omitted altogether.

Step 2: Volatility Estimation.

After the mean function is estimated, the parameter $\boldsymbol{\delta}_2$ can be estimated by minimizing the negative penalized likelihood

$$\sum_{i=1}^{n}\Big(e^2(\mathbf{z}_i\boldsymbol{\gamma})exp\{-2\mathbf{B}_2\boldsymbol{\delta}_2\} + 2\mathbf{B}_2(\mathbf{z}_i\boldsymbol{\gamma})\boldsymbol{\delta}_2\Big) + \frac{n}{2}\boldsymbol{\lambda}_2\boldsymbol{\delta}_2^{\mathsf{T}}\mathbf{D}_2\boldsymbol{\delta}_2, \tag{2.4}$$

where $e(\mathbf{z}_i\boldsymbol{\gamma}) = R_{i+1} - \mathbf{X}_i\widehat{\boldsymbol{\delta}}_1 \approx \sigma(\mathbf{z}_i\boldsymbol{\gamma})\epsilon_{i+1}$ is the residual from the estimation of mean function. The volatility function is approximated using

$$\widehat{\sigma}(\mathbf{z}_i\boldsymbol{\gamma}) = exp\{\mathbf{B}_2(\mathbf{z}_i\boldsymbol{\gamma}) * \widehat{\boldsymbol{\delta}}_2\}. \tag{2.5}$$

The new estimator $\widehat{\boldsymbol{\delta}}_1$ can be obtained through (penalized) weighted least squares where the weight is the inverse of the estimated volatility. More specifically, the new estimator $\widehat{\boldsymbol{\delta}}_1$ is the minimizer of the penalized log likelihood function

$$\sum_{i=1}^{n}\frac{1}{\widehat{\sigma}^2(\mathbf{z}_i\boldsymbol{\gamma})}\Big(R_{i+1} - \mathbf{X}_i\boldsymbol{\delta}_1\Big)^2 + \frac{n}{2}\boldsymbol{\lambda}_1\boldsymbol{\delta}_1^{\mathsf{T}}\mathbf{D}_1\boldsymbol{\delta}_1,$$

which has the analytical solution $\widehat{\boldsymbol{\delta}}_1 = (\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{X} + n\boldsymbol{\lambda}_1\mathbf{D})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{R}$, where $\mathbf{W} = diag\{1/\widehat{\sigma}^2(\mathbf{z}_i\boldsymbol{\gamma})\}$ is a diagonal weight matrix with each diagonal element calculated by (2.5).

Smoothing Parameter Selection

The selection of smoothing parameter is crucial in any smoothing problems, as evidenced by the dependence of the smoothing matrix on the smoothing parameter $\boldsymbol{\lambda}$. In practice, the optimal smoothing parameter is selected using a criterion of some sort on a grid values of the smoothing parameter. In the above iterative algorithm, grid search is done iteratively as well, i.e., smoothing parameter search accompanies spline coefficients estimation. In the two-dimensional $\boldsymbol{\lambda}_1$, to effectively search the optimal amount of smoothing, we recommend a straightforward grid searching scheme. First we fix both smoothing parameter as the same and conduct a one-dimensional grid search to find the optimal smoothing parameter. Then we fix one smoothing parameter and grid search for the other parameter. We have studied two-dimensional separate smoothing parameters and found no significant improvement for the fit using separate smoothing parameters. Also, the statistical test suggests constant alpha component and a smoothing function is unnecessary. Therefore, we present our results using a single smoothing parameter $\boldsymbol{\lambda}_1$ in this paper.

The most common criteria used to select smoothing parameter include Cross Validation (CV), Generalized Cross Validation (GCV), Mallow's C_p, and Akaike's Information Criterion (AIC) and some other variations. GCV is particularly popular in P-splines applications. In this paper, we choose the smoothing parameter λ_1 which minimizes the GCV criterion

$$GCV(\boldsymbol{\lambda}_1) = \frac{\sum_{i=1}^{n}(R_{i+1} - \mathbf{X}_i\boldsymbol{\delta}_1)^2}{\{1 - n^{-1}tr(\mathbf{H})\}^2},$$

where $tr(\mathbf{H})$ is the trace of the smoothing matrix, which is often called the degree of freedom of the fit (see Hastie and Tibshirani 1990). Similarly, the GCV for smoothing parameter λ_2 minimizes

$$GCV(\lambda_2) = \frac{Deviance(\lambda_2)}{\{1 - n^{-1}tr(\mathbf{B}_2(\mathbf{B}_2^{\mathsf{T}}\mathbf{B}_2 + n\lambda_2\mathbf{D}_2)^{-1}\mathbf{B}_2^{\mathsf{T}})\}^2},$$

where the numerator is the deviance of the model for a given value of λ_2.

Estimation of Single-index Parameter $\boldsymbol{\gamma}$

In this section we first introduce the estimation of single-index parameters $\boldsymbol{\gamma}$ when the varying functions $\alpha(\cdot)$ and $\beta(\cdot)$ are known, and then present the two-phase algorithm. First of all, to handle to constraints of $||\boldsymbol{\gamma}|| = 1$ and the first element $\gamma_1 > 0$, we let $\gamma_1 = \sqrt{1 - (\gamma_2^2 + \ldots + \gamma_d^2)}$, where $\sum_{j=2}^{d}\gamma_j^2 < 1$, thus the parameters become $\boldsymbol{\varphi} = (\gamma_2, \gamma_3, \ldots, \gamma_d)^{\mathsf{T}}$. Denote $\boldsymbol{\theta} = (\boldsymbol{\gamma}^{\mathsf{T}}, \boldsymbol{\delta}_a^{\mathsf{T}}, \boldsymbol{\delta}_b^{\mathsf{T}})^{\mathsf{T}}$ and $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\varphi}^{\mathsf{T}}, \boldsymbol{\delta}_a^{\mathsf{T}}, \boldsymbol{\delta}_b^{\mathsf{T}})^{\mathsf{T}}$ as the parameter vector before and after parameterization, the Jacobian matrix of this transformation is

$$\mathbf{J}(\tilde{\boldsymbol{\theta}}) = \begin{bmatrix} -(1-\|\boldsymbol{\varphi}\|^2)^{-1/2}\boldsymbol{\varphi}^T & \mathbf{0} \\ \mathbf{I}_{d-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{dim(\boldsymbol{\delta})} \end{bmatrix}. \tag{2.6}$$

After the reparameterization,[6] we need only estimate $\tilde{\boldsymbol{\theta}}$. From this point on, if not mentioned explicitly, $\boldsymbol{\theta}$ refers to reparametrized $\tilde{\boldsymbol{\theta}}$ for notational simplicity. Let $\dot{\mathbf{m}}_i(\boldsymbol{\theta})$ denote the first derivative with respect to $\boldsymbol{\theta}$, which is vector of $dim(\boldsymbol{\delta}) + d - 1$ dimension. Since there is no explicit solution for $\boldsymbol{\theta}$, we have to rely on numerical method. Some nonlinear optimization routines exist for this purpose, such as *lsqnonlin* and *fminunc* from Matlab, *nls* in S-PLUS. However, based on our experience, in high dimension $\boldsymbol{\theta}$, the convergence of the estimate is slow and resulting estimates are very sensitive to starting values. Instead we propose an iterative fixed point algorithm which is less sensitive to starting point and suitable for high dimension $\boldsymbol{\gamma}$. Fixed point algorithm was introduced in Cui, Härdle and Zhu (2011) for local linear estimation for single-index models, where fixed point algorithm works extremely well for high dimensional single-index models. We adapt the algorithm to penalized spline estimation for single-index varying-coefficient models and find similar results.

Before introducing the two-phase algorithm, we need to find an explicit expression of $\boldsymbol{\gamma}$ to iterate in the fixed point algorithm. The treatment here is in spirit similar to Cui, Härdle and Zhu (2011), although the estimation of smoothing components is quite different and more expedient in our case. To find the explicit expression of $\boldsymbol{\gamma}$, use the

[6] An alternative way of reparameterization is letting $\gamma_1 = 1$, and the new parameter $\boldsymbol{\varphi} = (1, \gamma_2, \gamma_3, \ldots, \gamma_d)/\sqrt{1 + \sum_{j=2}^{d} \gamma_j^2}$. Both reparameterizations would satisfy the constraints of norm equals one and first element positive. However, this reparametrization may not be suitable for the fixed point algorithm proposed here.

chain rule $\frac{\partial Q}{\partial \boldsymbol{\varphi}} = \frac{\partial Q}{\partial \boldsymbol{\gamma}} \frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\varphi}} = 0$, which result in $d-1$ equations

$$\begin{cases} \varphi_2 \frac{\partial Q}{\partial \boldsymbol{\gamma}_1} & = & \gamma_1 \frac{\partial Q}{\partial \boldsymbol{\gamma}_2} \\ & \vdots & \\ \varphi_d \frac{\partial Q}{\partial \boldsymbol{\gamma}_1} & = & \gamma_1 \frac{\partial Q}{\partial \boldsymbol{\gamma}_d}. \end{cases}$$

Define $M_s = \sum_{i=1}^{n} (R_{i+1} - \mathbf{B}_{a,i}\boldsymbol{\delta}_a - \mathbf{B}_{b,i}\boldsymbol{\delta}_b R_{m,i+1}) \cdot (\mathbf{B}_{a,i}\boldsymbol{\delta}_a + \mathbf{B}_{b,i}\boldsymbol{\delta}_b R_{m,i+1} \cdot Z_{s,i})$ and $\mathbf{M} = (M_1, \ldots, M_d)$. Then the above equations become

$$\begin{cases} \varphi_2 M_1 & = & \gamma_1 M_2 \\ & \vdots & \\ \varphi_d M_1 & = & \gamma_1 M_d, \end{cases}$$

which have the solution

$$\begin{cases} \gamma_1 & = & \frac{||M_1||}{||\mathbf{M}||} \\ \gamma_s^2 & = & \frac{||M_s||^2}{||\mathbf{M}||^2}, 2 \le s \le d \\ sign(\gamma_s M_1) & = & sign(M_s), 2 \le s \le d. \end{cases}$$

The above equals $\boldsymbol{\gamma}\frac{M_1}{||\mathbf{M}||} = \frac{|M_1|}{||\mathbf{M}||} \cdot \frac{\mathbf{M}}{||\mathbf{M}||}$, which automatically handles the constraints $||\boldsymbol{\gamma}|| = 1$ and $\gamma_1 > 0$. For fast convergence and robustness of the fixed algorithm, some constant C is added to $||\mathbf{M}||$ to avoid dividing by a small value. To achieve this, add $C\boldsymbol{\gamma}$ to both sides and after transformation we obtain

$$\boldsymbol{\gamma} = \frac{C}{M_1/||\mathbf{M}|| + C}\boldsymbol{\gamma} + \frac{M_1/||\mathbf{M}||^2}{M_1/||\mathbf{M}|| + C}\mathbf{M}. \tag{2.7}$$

The constant C is properly chosen to avoid dividing by zero. Further discussions on choosing the constant C are referred to Cui, Härdle and Zhu (2011).

The two-phase iterative algorithm for iteratively estimating the functions $\alpha(\cdot)$ and $\beta(\cdot)$ and single-index coefficient parameters $\boldsymbol{\gamma}$ are described as the following.

0) Initiate single-index parameter $\boldsymbol{\gamma}$. The initial value is crucial in any nonlinear optimization problem. This is especially true when the nonlinear optimization is

high dimensional. Our experience together with Cui, Härdle and Zhu (2011) suggest that fixed point algorithm works favorably over standard nonlinear programming routines, especially in high dimensional single-index coefficient models. For example, the usual ordinary linear squares (OLS) estimates for linear regression $R_{i+1} = (\mathbf{z}_i + \mathbf{z}_i R_{m,i+1})\boldsymbol{\gamma} + \epsilon_{i+1}$ can be used. Alternatively, random initial points from unit sphere can be used. Normalize $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma} = sign(\gamma_1)\boldsymbol{\gamma}/||\boldsymbol{\gamma}||$.

1) With fixed $\widehat{\boldsymbol{\gamma}}$, update $\widehat{\boldsymbol{\delta}}_1$ and $\widehat{R}_{i+1}$. Obtain spline bases with knots placed on equidistant quantiles of the single-index $\mathbf{z}_i\widehat{\boldsymbol{\gamma}}$ and calculate $\widehat{\boldsymbol{\delta}}_1 = (\mathbf{X}^\top\mathbf{X} + n\boldsymbol{\lambda}_1\mathbf{D}_1)^{-1}\mathbf{X}^\top\mathbf{R}$ and $\widehat{\mathbf{R}} = \mathbf{X}\widehat{\boldsymbol{\delta}}_1$

2) Update $\widehat{\boldsymbol{\gamma}}_{old}$ with $\boldsymbol{\gamma}_{new} = \frac{C}{M_1(old)/||\mathbf{M}||(old)+C}\boldsymbol{\gamma}_{old} + \frac{M_1(old)/||\mathbf{M}(old)||^2}{M_1(old)/||\mathbf{M}||(old)+C}\mathbf{M}(old)$. Normalize $\boldsymbol{\gamma}_{new}$ such that $\boldsymbol{\gamma}_{new} = sign(\gamma_{1,new})\boldsymbol{\gamma}_{1,new}/||\boldsymbol{\gamma}_{new}||$.

3) Repeat 1) & 2) until $max_{1\le s\le d}|\boldsymbol{\gamma}_{s,new} - \boldsymbol{\gamma}_{s,old}| \le tol$, where tol is the prescribed tolerance level.

4) Given $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\delta}}_1$, estimate $\boldsymbol{\delta}_2$ by minimizing the negative penalized likelihood function in (2.4). Use the new volatility estimates $\widehat{\sigma}^2(\mathbf{z}_i\widehat{\boldsymbol{\gamma}})$ to update the mean spline coefficients estimator by $\widehat{\boldsymbol{\delta}}_1 = (\mathbf{X}^\top\mathbf{W}\mathbf{X} + n\boldsymbol{\lambda}_1\mathbf{D})^{-1}\mathbf{X}^\top\mathbf{W}\mathbf{R}$, where $\mathbf{W} = diag\{1/\widehat{\sigma}^2(\mathbf{z}_i\widehat{\boldsymbol{\gamma}})\}$ is the diagonal weight matrix.

5) Repeat 1) - 4) a few times or until convergence.

2.3.2 Variable Selection Procedure

As some authors have pointed out that estimation of conditional factor model is quite sensitive to the choice of state variables (e.g., Ghysels 1998; Harvey 2001), it is necessary to investigate what variables should be included in explaining conditional beta. In this paper we have conducted variable selection procedure for a more comprehensive choice of state variables. The variables set not only includes standard stock return predictors but also macroeconomic indicators. The literature on variable selection for single-index varying-coefficient models is extremely limited because of the difficulty of the problem itself. Fan, Yao and Cai (2003) use backward stepwise deletion in combination with modified t-statistic and Akaike Information Criterion (AIC) type criterion for variable selection. However, such variable selection procedure is ad hoc and may not guarantee to select the best predictors. More straightforward way, however, is to evaluate all possible models, which is an exhaustive approach in variable selection. Exhaustive methods fit all the combinations of predictor variables and choose the best one based on certain criterion. The drawback of such exhaustive method is that it is time consuming to evaluate all possible models.In this study, there are two favorable factors for such exhaustive search: reasonably large number of state variables and fast computation of the estimation method. Based on the number of state variables described in Section 2.2, the number of all possible combinations is manageable. Moreover, due to the computational expediency of P-splines and fixed point algorithm, the computation is fast and variable selection can be done in a relatively short time.

Best model is selected by minimizing some criterion such as adjusted R^2, Mallow's C_p, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), among others. Each criterion balances goodness of fit and simplicity of the model. We choose the best model using the BIC, which tends to select more parsimonious and thus more explainable models among several criteria. BIC in our context is defined as

$$\sum_{i=1}^{n}\Big(exp\{-2\mathbf{B}_2(\mathbf{z}_i\boldsymbol{\gamma})\boldsymbol{\delta}_2\}\{R_{i+1}-\mathbf{B}_{a,i}(\mathbf{z}_i\boldsymbol{\gamma})\boldsymbol{\delta}_a-\mathbf{B}_{b,i}(\mathbf{z}_i\boldsymbol{\gamma})\boldsymbol{\delta}_b R_{m,i+1}\}^2+2\mathbf{B}_2(\mathbf{z}_i\boldsymbol{\gamma})\boldsymbol{\delta}_2\Big)+qln(n),$$

where the first term is the maximum likelihood function for the model, and q is the model complexity parameter. In P-splines single-index varying-coefficient model, $q = tr(\mathbf{H}) + d$, where $tr(\mathbf{H})$, trace of the smoothing matrix $\mathbf{H}$, is often called the effective number of parameters in nonparametric estimation.

2.3.3 Tests on Alpha and Beta

We first establish consistency and asymptotic normality, and then construct the Wald statistic for inferences on the alpha and beta. Two types of asymptotics can be used: fixed knots and increasing number of knots. Fixed knots asymptotics are most practical and relevant in applications, where the fixed number of knots are used (Yu and Ruppert 2002; Jarrow, Ruppert and Yu 2004). Also, the bias due to fixed knots approximation of P-spline of smooth function is negligible compared to the standard deviation of the function estimates and the bias due to penalty. The asymptotics introduced here are in spirit similar to the Theorem 2 in Yu and Ruppert (2002) and Theorem 2 in Wu, Lin and Yu (2011).

Theorem 3. *Under mild regularity conditions, if the smoothing parameter $\lambda_n \sim o(n^{-1/2})$, then a sequence of estimators $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}} = (\widehat{\gamma}^{\mathsf{T}}, \widehat{\boldsymbol{\delta}}_a^{\mathsf{T}}, \widehat{\boldsymbol{\delta}}_b^{\mathsf{T}}, \widehat{\boldsymbol{\delta}}_2^{\mathsf{T}})^{\mathsf{T}}$ is root-n consistent and converges to a normal distribution,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N(0, \mathbf{I}^{-1}(\boldsymbol{\theta})), \tag{2.8}$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the usual Fisher Information.

The proof of Theorem above is similar to Yu et al. (2009). The penalty parameter is assumed to vanish fast enough as n goes to infinity to ensure the result given in (2.8) involves no penalty parameter. For finite sample inference, asymptotic results with fixed penalty parameter would be preferred to avoid overestimating the variance of $\boldsymbol{\theta}$ as in (2.8). The sandwich formula will be given by

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda}) - \boldsymbol{\theta}(\boldsymbol{\lambda})) \xrightarrow{D} N\Big(0, \boldsymbol{\Gamma}_n^{-1}\big(\boldsymbol{\theta}(\boldsymbol{\lambda})\big)\boldsymbol{\Lambda}_n\big(\boldsymbol{\theta}(\boldsymbol{\lambda})\big)\boldsymbol{\Gamma}_n^{-T}\big(\boldsymbol{\theta}(\boldsymbol{\lambda})\big)\Big), \tag{2.9}$$

where $\boldsymbol{\Gamma}_n\big(\boldsymbol{\theta}(\boldsymbol{\lambda})\big) = \sum_{i=1}^{n}(\partial/\partial\boldsymbol{\theta}^{\mathsf{T}})\Phi_{\mathbf{z}_i}\boldsymbol{\gamma}(\boldsymbol{\theta})$, $\boldsymbol{\Lambda}_n\big(\boldsymbol{\theta}\big) = \sum_{i=1}^{n}\Phi_{\mathbf{z}_i}\boldsymbol{\gamma}(\boldsymbol{\theta})\Phi_{\mathbf{z}_i}^{\mathsf{T}}\boldsymbol{\gamma}(\boldsymbol{\theta})$, $\Phi_{\mathbf{z}_i}\boldsymbol{\gamma}(\boldsymbol{\theta}) = -(\partial/\partial\boldsymbol{\theta}^{\mathsf{T}})l_n(\boldsymbol{\theta};\mathbf{z}_i\gamma) + \boldsymbol{\lambda}\mathbf{D}\boldsymbol{\theta}$ (see Yu and Ruppert 2002 and Yu et a. 2009). The sandwich estimator of the covariance matrix is justified by $\boldsymbol{\Omega}_n(\widehat{\boldsymbol{\theta}}(\boldsymbol{\lambda})) = \boldsymbol{\Gamma}_n^{-1}\big(\boldsymbol{\theta}(\boldsymbol{\lambda})\big)\boldsymbol{\Lambda}_n\big(\boldsymbol{\theta}(\boldsymbol{\lambda})\big)\boldsymbol{\Gamma}_n^{-T}\big(\boldsymbol{\theta}(\boldsymbol{\lambda})\big)$.

The asymptotic properties in (2.9) can be conveniently used for joint inferences concerning the spline coefficients. In general, if we are testing the null hypothesis $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{c}$, where $\mathbf{L}$ is a r by $dim(\boldsymbol{\theta})$ matrix of full row rank, we construct the Wald statistic

$$W = (\mathbf{L}\boldsymbol{\theta} - \mathbf{c})^{\mathsf{T}}\big(\mathbf{L}\boldsymbol{\Omega}_n(\widehat{\boldsymbol{\theta}}\mathbf{L})\big)^{-1}(\mathbf{L}\boldsymbol{\theta} - \mathbf{c}), \tag{2.10}$$

where $\boldsymbol{\Omega}_n(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\Gamma}_n^{-1}\big(\boldsymbol{\theta}\big)\boldsymbol{\Lambda}_n\big(\boldsymbol{\theta}\big)\boldsymbol{\Gamma}_n^{-T}\big(\boldsymbol{\theta}\big)$ for sufficiently large n. Under H_0, the Wald statistic W is a limiting χ^2 distribution with r degrees of freedom. In cases of inferences concerning the single-index parameter $\boldsymbol{\gamma}$, the upper left $(d-1)\times(d-1)$ block matrix of $\boldsymbol{\Omega}_n(\widehat{\boldsymbol{\theta}})$ can be used to calculate the Wald statistics. The covariance matrix estimator is given by $\boldsymbol{\Omega}_n(\widehat{\tilde{\boldsymbol{\theta}}}) = \widehat{\mathbf{J}}\boldsymbol{\Omega}_n(\widehat{\boldsymbol{\theta}})\widehat{\mathbf{J}}$, where $\mathbf{J} = \mathbf{J}(\tilde{\boldsymbol{\theta}})$ is the Jacobian matrix define in (2.6).

In conditional CAPM, we are interested in testing several hypotheses on α and β. In terms of alpha, we are interested in testing whether alpha component is zero; zero alpha suggests conditional CAPM can fully explain value premium. If alpha is significantly nonzero, then we would like to know if alpha is constant; constant alpha suggests time varying modeling for alpha is unnecessary. In terms of beta, we are interested in testing the functional form. First, constant beta is tested; failure to reject the null hypothesis indicates irrelevance of any state variables. Otherwise linear beta (in single-index) is tested; rejection of null hypothesis suggests nonlinear specification and nonparametric approach should be preferred. On the contrary, failure of rejecting linear hypothesis provides justification of using linear specification such as Lettau and Ludvigson (2001), and Petkova and Zhang (2005).

In each test, with the truncated power basis, we can conveniently formulate the hypothesis in the form of spline coefficients. For example, testing $H_0 : \alpha \equiv 0$ is equivalent to test all its spline coefficients are simultaneously zeros, i.e., $H_0 : \boldsymbol{\delta}_a \equiv 0$ or $\mathbf{L} = blockdiag\{\mathbf{0}_d, \mathbf{1}_{dim(\boldsymbol{\delta}_a)}, \mathbf{0}_{dim(\boldsymbol{\delta}_b)}\}$ in (2.10). The null hypothesis $H_0 : \beta\ is\ linear\ in\ \mathbf{z}\boldsymbol{\gamma}$ is equivalent to test all higher order spline coefficients are simultaneously zeros, i.e., $H_0 : \boldsymbol{\delta}_{b,2} = \cdots = \boldsymbol{\delta}_{b,p} = \boldsymbol{\delta}_{b,p+k+1} = 0$ or $\mathbf{L} = blockdiag\{\mathbf{0}_d, \mathbf{0}_{dim(\boldsymbol{\delta}_a)}, \mathbf{0}_2, \mathbf{1}_{dim(\boldsymbol{\delta}_b)-2}\}$ in (2.10).

The other problem of interest is to test the significance of long-run alpha (Ang and Kristensen 2011), which is defined as the average of conditional alphas. The long-run alpha is estimated using the average of the conditional alphas in the sample period. The hypothesis test on the significance of long-run alpha can be expressed as

$$H_0 : \alpha_{LR} \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \alpha(\mathbf{z}_i\boldsymbol{\gamma}) = 0, \tag{2.11}$$

which is not as a strong hypothesis as the that conditional alpha being zero over any values of single-index.[7] Simple t test is used to test the above hypothesis on long-run alpha. The standard errors can be obtained from the first d diagonal elements of $\boldsymbol{\Omega}_n(\widehat{\widehat{\boldsymbol{\theta}}}) = \widehat{\mathbf{J}}\boldsymbol{\Omega}_n(\widehat{\boldsymbol{\theta}})\widehat{\mathbf{J}}$, or alternatively through bootstrap resamples.

The covariance matrix $\boldsymbol{\Omega}_n(\widehat{\boldsymbol{\theta}})$ in the Wald statistic in (2.10) and the standard error for the long-run alpha can be difficult to calculate using the delta method in the finite samples. Instead we use bootstrapping resamples to construct the covariance matrices in (2.10) and standard errors in testing (2.11). Because of the presence of heteroskedasticity, we use a specialized wild bootstrap procedure for this purpose. Wild bootstrap procedure, originally proposed by Wu (1986) and extended by Mammen (1993) and others, is particularly useful in the presence of heteroskedasticity and small sample sizes.

[7] The hypothesis of conditional alpha being zero over any values of single-index, or $H_0 : \alpha \equiv 0$, is a stronger hypothesis. It holds true that if the conditional alpha is zero over any values, the long-run alpha or the average of conditional alphas is also zero. However, the opposite is not true.

Unlike the usual residual bootstrap where random residuals are drawn and added to the estimates to construct resamples, the wild bootstrap accounts for heteroskedasticity by creating weighted residuals where the weight is a random variable with zero mean and unit variance. The wild bootstrap procedure we use in this paper is described as follows:

(1) Fit single-index varying-coefficient model and find $\widehat{R}_{i+1} = \mathbf{B}_a(\mathbf{z}_i\widehat{\boldsymbol{\gamma}})\widehat{\boldsymbol{\delta}}_a + \mathbf{B}_b(\mathbf{z}_i\widehat{\boldsymbol{\gamma}})\widehat{\boldsymbol{\delta}}_b R_{m,i+1}$, with residual $\widehat{\epsilon}_{i+1} = R_{i+1} - \widehat{R}_{i=1}$ $(i = 1, \ldots, n)$.

(2) Center the residual $\widehat{\epsilon}_{i+1} - \bar{\widehat{\epsilon}}_{i+1}$ $(i = 1, \ldots, n)$, where $\bar{\widehat{\epsilon}}_{i+1}$ is the mean of residuals.

(3) Draw the bootstrap error ϵ^*_{i+1} from centered residuals $\epsilon_2, \ldots, \epsilon_{n+1}$ with replacement. The bootstrap sample is created using $R_{i+1} + v_{i+1}\epsilon^*_{i+1}$ $(i = 1, \ldots, n)$, where v_{i+1} is a random variable of standard normal distribution.[8]

Repeat (3) to create N resamples. Sample covariance matrix and standard errors of long-run alpha can be calculated based on the N estimates.

2.4 Conditional CAPM

2.4.1 Conditional Beta as a Function of Macroeconomic and Financial Variables

We first estimate the conditional CAPM with macroeconomic variables and financial variables as the conditioning variables. For the monthly data, we find that INF, PE, and UE

[8]We have also implemented the wild bootstrap weighting scheme of Mammen (1993) where the random weight takes the form

$$v_{i+1} = \begin{cases} -(\sqrt{5}-1)/2 & \text{with probability } (\sqrt{5}+1)/(2\sqrt{5}) \\ (\sqrt{5}+1)/2 & \text{with probability } (\sqrt{5}-1)/(2\sqrt{5}). \end{cases}$$

The testing results agree with that of using standard normal weight as in Wu (1986).

are significant conditioning variables. This finding is consistent with the results reported in Table 2.1, which shows that BETAT, a proxy for conditional beta have relatively strong correlation with INF, PE, and UE. IP also correlates negatively with BETAT; however, it is not selected as a significant conditioning variable likely because of its strong negative correlation with UE. Similarly, Stock and Watson (2003) find that TERM has significant predictive power for output and inflation; nevertheless, TERM is not selected as a significant conditioning variable. Because the unemployment and inflation are the two most closely watched gauges of aggregate economic activity, our results indicate that the value premium is indeed sensitive to business conditions. Moreover, PE is also a significant conditioning variable, while the other financial variables provide little additional information about the conditional beta. The latter result is in contrast with the specification adopted in the previous studies, which assume that conditional beta depends on only commonly used stock market return predictors.

The estimated conditional beta suggests there is a monotonically negative relation between conditional beta and the index (untabulated). In Table 2.2, we find that the index correlates positively with INF and PE, and the relations are all statistically significant at the 1% level. The relationship between the index and UE is negative and marginally insignificant at the 5% level. Overall, our results suggest that the conditional beta of the value premium changes countercyclically across time because it increases with UE and decreases with INF and PE. To investigate formally this issue, in Figure 2.1, we plot the estimated conditional beta across time, with shaded areas indicating business recession periods dated by the National Bureau of Economic Research (NBER). For each of 7 business recessions in our sample, we observe a sharp spike in the conditional beta. Moreover, the conditional beta tends to decrease during business expansions. Therefore, our results reveal a strongly countercyclical variation in the conditional beta of the value premium.

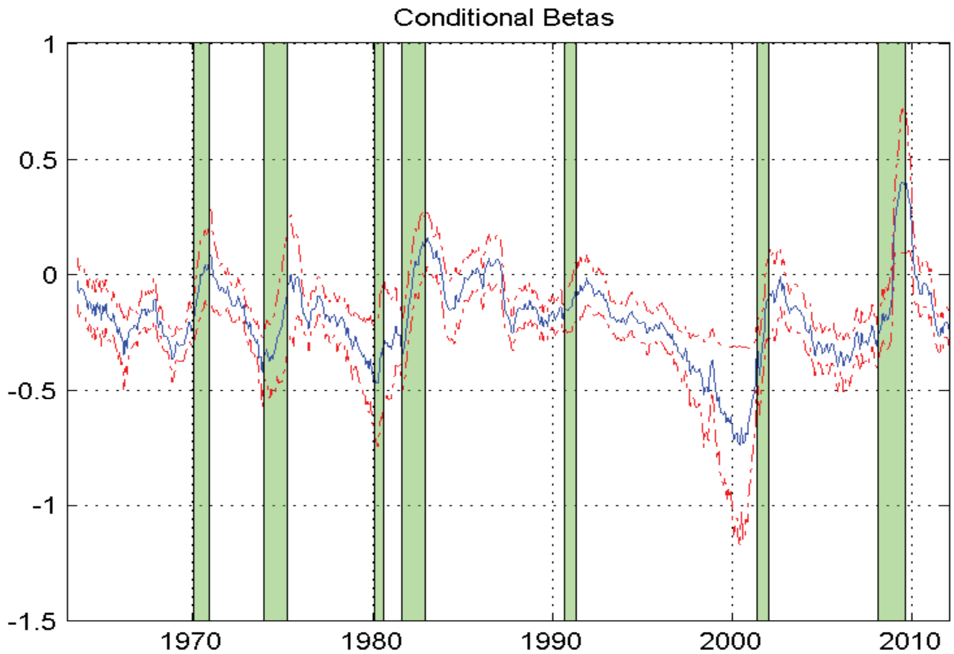
Table 2.2: Tests on Alphas and Betas with INF, PE and UE as conditioning variables

Panel A: Tests on alpha				
Hypothesis		Wald Statistic		P Value
$H_o : \alpha = zero$		27.93		0.0005
$H_o : \alpha = Constant$		13.41		0.0627
Panel B: Test on long-run alpha				
Hypothesis	Estimate	s.e.	t Value	P Value
$H_o : \alpha_{LR} = 0$	0.0038	0.0010	3.82	0.0001
Panel C: Tests on beta				
Hypothesis		Wald Statistic		P Value
$H_o : \beta = 0$		75.32		0.0000
$H_o : \beta = Constant$		52.87		0.0000
$H_o : \beta = Linear$		6.28		0.3924
Panel D: Tests on beta when alpha is modeled as constant				
Hypothesis		Wald Statistic		P Value
$H_o : \beta = 0$		79.29		0.0000
$H_o : \beta = Constant$		44.27		0.0000
$H_o : \beta = Linear$		0.14		0.9999
Panel E: Tests on single-index coefficients				
	Estimate	s.e.	t Value	P Value
$\gamma_1(INF)$	0.9574	.0412	23.21	0.0000
$\gamma_2(PE)$	0.2524	.0891	2.83	0.0000
$\gamma_3(UE)$	-0.1402	.0708	-1.98	0.0503

The table reports statistical test results on conditional and long-run alphas and betas in the single-index long-run model. Tests on single-index coefficients parameters are also reported. Tests on conditional alphas and betas are based on the Wald statistics in equation 2.10. Test on long-run alpha is based on equation 2.11. All covariance matrices and standard errors are calculated from bootstrap samples. The monthly data sample is from July 1963 to December 2012.

In panel D of Table 2.2, we test the hypotheses about the time variation in the conditional beta. We overwhelmingly reject the null hypothesis that the beta is zero. Moreover, we find strong evidence against the null hypothesis that the conditional beta is constant. Nevertheless, we fail to reject the null hypothesis that the conditional beta is a linear function of INF, PE, and UE. This result shows that the relationship between the conditional beta and the index is essentially linear.

Figure 2.1: Conditional beta on INF, PE, and UE



Note: the figure shows monthly estimates of conditional betas of the value premium. The conditioning variables are INF, PE and UE, the best subset from variable selection procedure. We plot the conditional betas in blue line along with 95% confidence bands in red line. We also shade the NBER recession periods in green horizon bars.

We fail to reject the null hypothesis that alpha is constant. Unreported results show that the value premium has a significantly positive alpha of 0.47% per month when we use the unconditional CAPM. In panel B of Table 2.2, we find that the alpha becomes noticeably smaller to 0.37% per month in the conditional CAPM. The 1.2% annual reduction in the alpha is consistent with the notion that the conditional CAPM helps explain the value premium because its risk exposure is larger during business recessions when conditional equity premium is higher than during business explanations when conditional equity premium is low. Nevertheless, consistent with the recent studies, e.g., Petkova and Zhang (2005), Lewellen and Nagel (2006), Ang and Kristensen (2011), and Li and Yang (2011), we find that the alpha remains significant positive even when we control for time-varying conditional beta.

For comparison, we estimate the single-index varying-coefficient model using all the macroeconomic and financial variables as the conditioning variables. We find that TERM, RREL, INF, VOL and UE are statistically significant at the 5% level. The statistical significance in these five variables does not give the best state variable set. The long-run alpha is significantly positive but conditional alpha is not significant from zero. The latter result is due to large variance in the conditional alpha component. These results highlight the importance of variable selection, which allow us to precisely estimate the conditional beta. For brevity, they are not reported in the paper but are available upon request.

2.4.2 Realized Beta as the Conditioning Variable

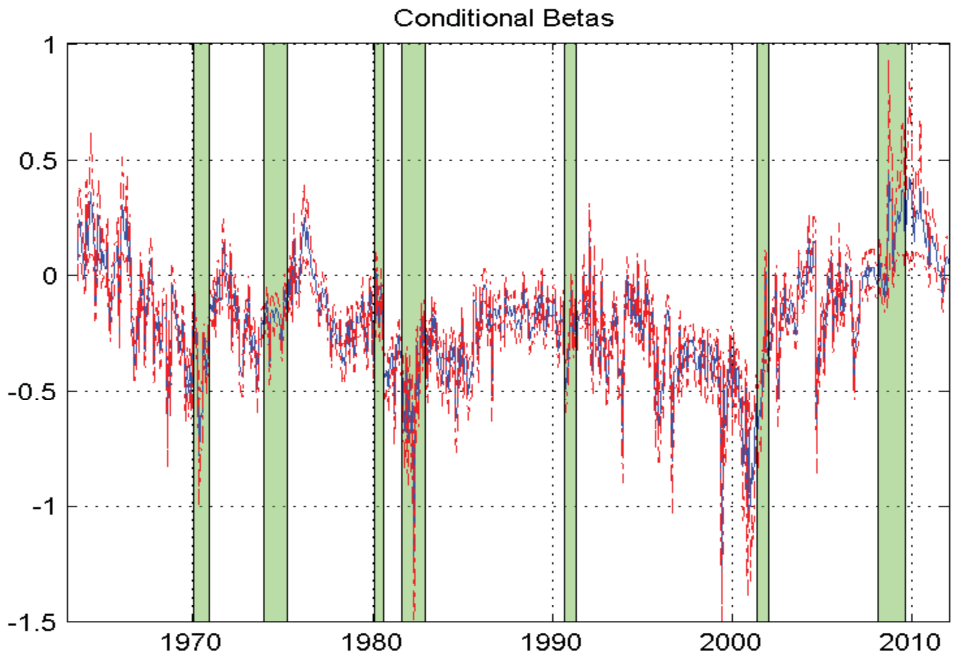
The conditional CAPM fails to explain the value premium possibly because we omit some important conditioning variables. If conditional beta does not change quickly across time, Lewellen and Nagel (2006) argue that the realized beta is a good proxy for the conditional beta. As a robustness check, in this subsection, we include only the realized beta as the conditioning variable. Table 2.3 reports the main estimation results of the single-index varying-coefficient model. Panel A shows that we cannot reject the null hypothesis of constant alpha. In panel C, we reject the null hypothesis zero or constant conditional beta, but cannot reject the hypothesis that conditional beta is linear in realized beta, which is consistent with Lewellen and Nagel (2006), where realized beta is used as a proxy for conditional beta. The results are qualitatively similar when we assume a constant alpha, as reported in panel D. Moreover, in Figure 2.2, we plot the fitted conditional beta across time, while conditional alpha is modeled as constant. Again, we find that conditional beta tends to increase during business recessions, indicating that conditional beta change countercyclically across time, although this relationship is less obvious due to the highly erratic realized beta. Nevertheless, consistent with findings by Lewellen and Nagel (2006), Ang and Kristensen (2011), and Li and Yang (2011), we confirm that the conditional CAPM does not fully explain the value premium. Panel B shows that the long-run alpha is 0.44% per month and is statistically significant at the 1% level.

Table 2.3: Tests on Alphas and Betas with realized beta (BETAT) as conditioning variables

Panel A: Tests on alpha				
Hypothesis		Wald Statistic		P Value
$H_o : \alpha = zero$		30.45		0.0002
$H_o : \alpha = Constant$		12.73		0.0791
Panel B: Tests on long-run alpha				
Hypothesis	Estimate	s.e.	t Value	P Value
$H_o : \alpha_{LR} = 0$	0.0044	0.0010	4.24	0.0000
Panel C: Tests on beta				
Hypothesis		Wald Statistic		P Value
$H_o : \beta = 0$		127.40		0.0000
$H_o : \beta = Constant$		74.47		0.0000
$H_o : \beta = Linear$		5.73		0.1255
Panel D: Tests on beta when alpha is modeled as constant				
Hypothesis		Wald Statistic		P Value
$H_o : \beta = 0$		116.16		0.0000
$H_o : \beta = Constant$		92.62		0.0000
$H_o : \beta = Linear$		7.41		0.2844

The table reports statistical test results on conditional and long-run alphas and betas in the single-index long-run model. Tests on conditional alphas and betas are based on the Wald statistics in equation 2.10. Test on long-run alpha is based on equation 2.11. All covariance matrices and standard errors are calculated from bootstrap samples. The monthly data sample is from July 1963 to December 2012.

Figure 2.2: Conditional beta on BETAT



Note: the figure shows monthly estimates of conditional betas of the value premium. The conditioning variable is realized beta (BETAT) only. We plot the conditional betas in blue line along with 95% confidence bands in red line. We also shade the NBER recession periods in green horizon bars.

2.4.3 Conditional Beta as a Function of Macroeconomic and Financial Variables and Realized Beta

Because realized beta is not necessarily an efficient estimate of conditional beta, the macroeconomic and financial variables may provide additional information beyond BETAT about the conditional beta. An improved measure of conditional beta might provide a better explanation for the value premium. To address this issue, we add BETAT to the instrument variable set. Consistent with the existing studies, e.g., Lewellen and Nagel (2006), Ang and Kristensen (2011), and Li and Yang (2011), we find that BETAT is a significant conditioning variable, even when we control for the macroeconomic and financial variables. On the other hand, as conjectured, BETAT does not fully subsume the information in the macroeconomic and financial variables either. In particular, INF and PE remain significant conditioning variables, although UE becomes insignificant when we control for BETAT.

In Table 2.4, we report the main estimation results of the single-index varying-coefficient model. Because conditional beta increases linearly with the index (figure not shown), Panel E shows that the index correlates positively with BETAT, confirming the argument by Lewellen and Nagel (2006), Ang and Kristensen (2011), and Li and Yang (2011) that realized beta provides important information about conditional beta. We also find that the index correlates negatively with INF and PE, supporting the notion that conditional beta moves countercyclically across time. To illustrate this point further, in Figure 2.3, we plot the estimated conditional beta across time. Similar to the results reported in Figures 2.1 and 2.2, we find that realized beta increases sharply during business recessions and decreases during business recessions. Again, this result is less obvious due to the erratic realized beta. In panel C and D, we overwhelmingly reject the null hypotheses that conditional beta is zero and the null hypothesis that conditional beta is constant. Moreover, we fail to reject the null hypothesis of a linear relation between conditional beta and the conditioning variables, i.e., BETAT, INF and PE. Moreover, adding realized beta as an additional conditioning variable does not change our main finding in any qualitative manner. Panel B shows that the long-run alpha is significantly positive at the 1% level.

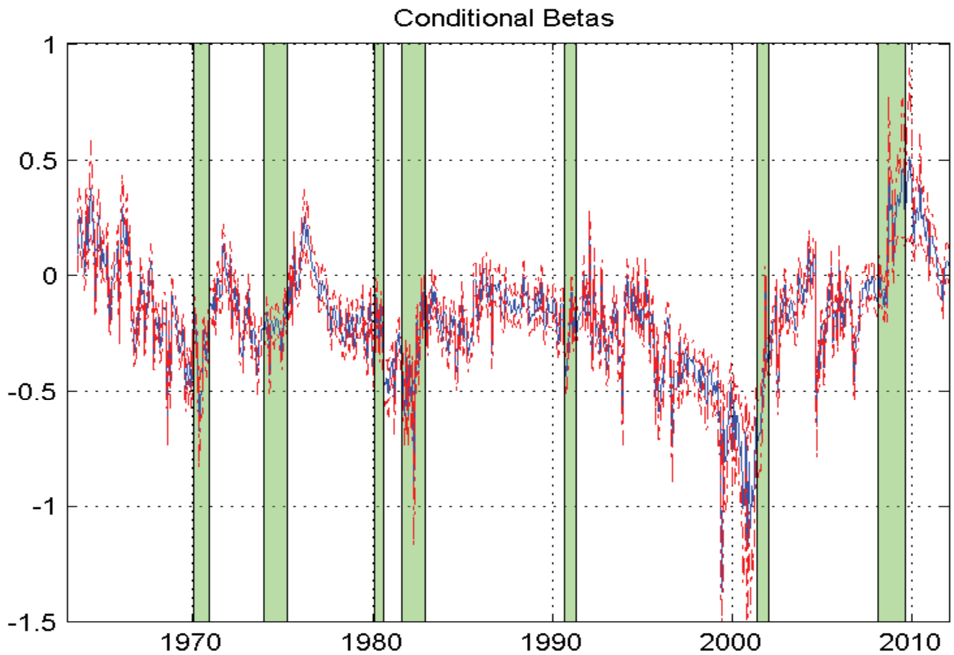
Table 2.4: Tests on Alphas and Betas with BETAT, INF and PE as conditioning variables

Panel A: Tests on alpha				
Hypothesis		Wald Statistic		P Value
$H_o : \alpha = zero$		26.20		0.0010
$H_o : \alpha = Constant$		14.07		0.0400
$H_o : \alpha = Linear$		12.11		0.0595
Panel B: Tests on long-run alpha				
Hypothesis	Estimate	s.e.	t Value	P Value
$H_o : \alpha_{LR} = 0$	0.0040	0.0011	3.80	0.0001
Panel C: Tests on beta				
Hypothesis		Wald Statistic		P Value
$H_o : \beta = 0$		118.12		0.0000
$H_o : \beta = Constant$		88.26		0.0000
$H_o : \beta = Linear$		3.72		0.7144
Panel D: Tests on beta when alpha is modeled as constant				
Hypothesis		Wald Statistic		P Value
$H_o : \beta = 0$		135.49		0.0000
$H_o : \beta = Constant$		94.98		0.0000
$H_o : \beta = Linear$		5.03		0.5402
Panel E: Tests on single-index coefficients				
	Estimate	s.e.	t Value	P Value
$\gamma_1(BETAT)$	0.2352	.1430	1.65	0.1031
$\gamma_2(INF)$	-0.8977	.2291	-3.92	0.0002
$\gamma_3(PE)$	-0.3725	.1475	-2.53	0.0131

The table reports statistical test results on conditional and long-run alphas and betas in the single-index long-run model. Tests on single-index coefficients parameters are also reported. Tests on conditional alphas and betas are based on the Wald statistics in equation 2.10. Test on long-run alpha is based on equation 2.11. All covariance matrices and standard errors are calculated from bootstrap samples. The monthly data sample is from July 1963 to December 2012.

As a robustness check, we estimate the single-index varying-coefficient model using all

Figure 2.3: Conditional beta on BETAT, INF and PE



Note: the figure shows monthly estimates of conditional betas of the value premium. The conditioning variables are realized beta (BETAT), inflation (INF) and unemployment rate (UE). We plot the conditional betas in blue line along with 95% confidence bands in red line. We also shade the NBER recession periods in green horizon bars.

instrumental variables. We find that RREL, IP, UE and BETAT are significant conditioning variables at the 5% level. The other testing results on conditional alpha and beta, and long-run alpha are qualitatively in agreement with those using the model selected based on the best BIC criterion (BETAT, INF and PE). Again, these variables differ from the best subset state variables. These results again highlight the importance of the variable selection, which allows us to estimate the conditional beta more precisely. For brevity, these results are not reported but are available upon request.

2.4.4 Smoothed Realized Beta

Monthly realized beta is quite erratic, which is witnessed in Figure 2.2 and 2.3. As a robustness check, we use a smoothed realized beta measure, BETAT12–the average of the monthly realized betas in the most recent twelve months. In this case, we identify IP, DP and the smoothed realized beta (BETAT12) as the significant conditioning variables. There is a monotonically decreasing relation between the conditional beta and the index (figure not shown), and Table 2.5 shows that the index correlates positively with IP and negatively with DP and BETAT12.

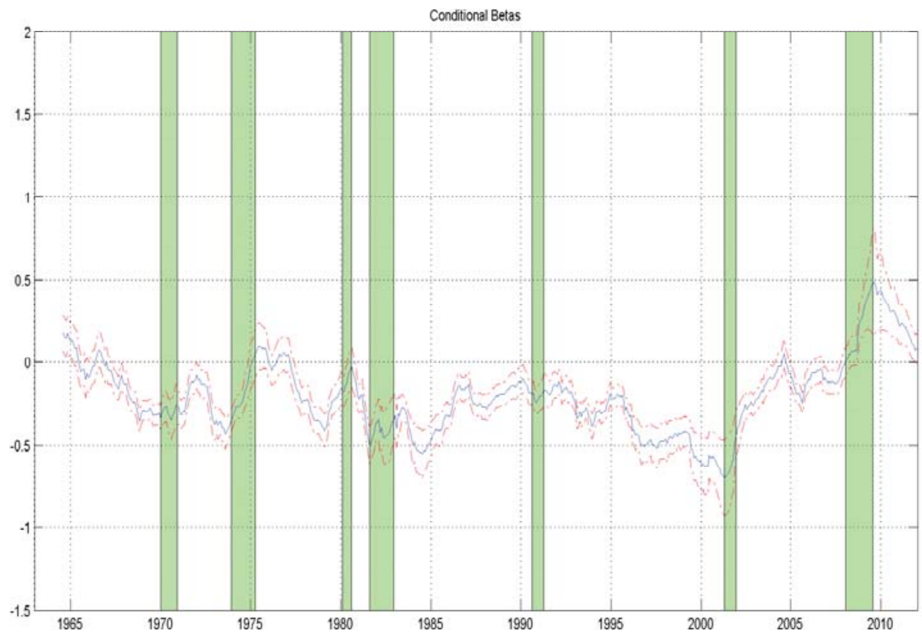
Again, our results suggest that the conditional beta of the value premium is countercyclical. We further illustrate this point in Figure 2.4, in which the conditional beta increases sharply during business recessions. The results in Table 2.5 are congruent with those reported in Table 2.4. Specifically, we find that the conditional CAPM does not fully explain the value premium. Smoothed realized beta using three months average yields qualitatively analogous results. For brevity, these results are omitted but are available upon request.

Table 2.5: Tests on Alphas and Betas with IP, DP and BETAT12 as conditioning variables

Panel A: Tests on alpha				
Hypothesis		Wald Statistic		P Value
$H_o : \alpha = zero$		25.42		0.0013
$H_o : \alpha = Constant$		10.41		0.1667
Panel B: Tests on long-run alpha				
Hypothesis	Estimate	s.e.	t Value	P Value
$H_o : \alpha_{LR} = 0$	0.0040	0.0011	3.68	0.0002
Panel C: Tests on beta				
Hypothesis		Wald Statistic		P Value
$H_o : \beta = 0$		135.86		0.0000
$H_o : \beta = Constant$		112.33		0.0000
$H_o : \beta = Linear$		3.63		0.7260
Panel D: Tests on beta when alpha is modeled as constant				
Hypothesis		Wald Statistic		P Value
$H_o : \beta = 0$		165.16		0.0000
$H_o : \beta = Constant$		79.15		0.0000
$H_o : \beta = Linear$		5.96		0.4282
Panel E: Tests on single-index coefficients				
	Estimate	s.e.	t Value	P Value
$\gamma_1(IP)$	0.6831	.1929	3.54	0.0006
$\gamma_2(DP)$	-0.6181	.5164	-1.20	0.2342
$\gamma_3(BETAT12)$	-0.3890	.1124	-3.46	0.0008

The table reports statistical test results on conditional and long-run alphas and betas in the single-index long-run model. Tests on single-index coefficients parameters are also reported. IP is the industrial production, DP is the dividend-price ratio, and BETA12 is smoothed realized beta constructed by averaging the previous 12 month realized beta. Tests on conditional alphas and betas are based on the Wald statistics in equation 2.10. Test on long-run alpha is based on equation 2.11. All covariance matrices and standard errors are calculated from bootstrap samples. The monthly data sample is from July 1963 to December 2012.

Figure 2.4: Conditional beta on smoothed BETAT12, DP and IP



Note: the figure shows monthly estimates of conditional betas of the value premium. The conditioning variables are 12-month average realized beta (BETAT12), dividend-price ratio (DP) and industrial production (IP). We plot the conditional betas in blue line along with 95% confidence bands in red line. We also shade the NBER recession periods in green horizon bars.

2.4.5 Time-Varying Beta of Ang and Kristensen (2011)

Ang and Kristensen (2011) also use nonparametric approach to model conditional beta but find weak effect of business cycle on conditional beta, which seems to contradict our findings. As another robust check, we model conditional beta as time varying–modeling approach corresponding to Ang and Kristensen (2011) but estimating conditional beta using method in this paper.[9] The plot of conditional beta over time in Figure 2.5 resembles the conditional beta in Ang and Kristensen (2011). This time varying beta can be always included in our single-index varying-coefficient framework as a conditioning variable. Thus we can investigate if this time-varying realized beta is an efficient estimate of conditional beta, the way similar to the treatment in Section 2.4.2-2.4.4 where the realized beta in Lewellen and Nagel (2006) is treated as a conditioning variable. The statistical test results (Panel A-D) on conditional alphas and betas do not qualitatively differ from what obtained in previous sections. Conditional beta is found to be countercyclical because conditional beta based on Panel E. To further illustrate conditional beta comoves with business cycle, following Ang and Kristensen (2011), we regress time-varying realized beta onto state variables such as the significant variables we have found in Section 2.4.1 (i.e., INF, PE and UE). The results are reported in Table 2.6 Panel F. Conditional beta is negatively related to INF and PE and positively related to UE. All state variables are statistically highly significant at the 1% level. This result is consistent with the results from singe-index varying coefficient models in Table 2.2. The differences between our

[9] It is worth noting that our approach is more general than Ang and Kristensen (2011) in the way that our model can nest their model as a special case. When conditional beta is modeled as only time-varying, then our model is conceptually the same as Ang and Kristensen (2011). The functional form of conditional beta is not linear function but a complicated function of time t (see Figure 2.5). This highlights the the importance of modeling conditional beta as some unspecified function rather than linear function in the first place.

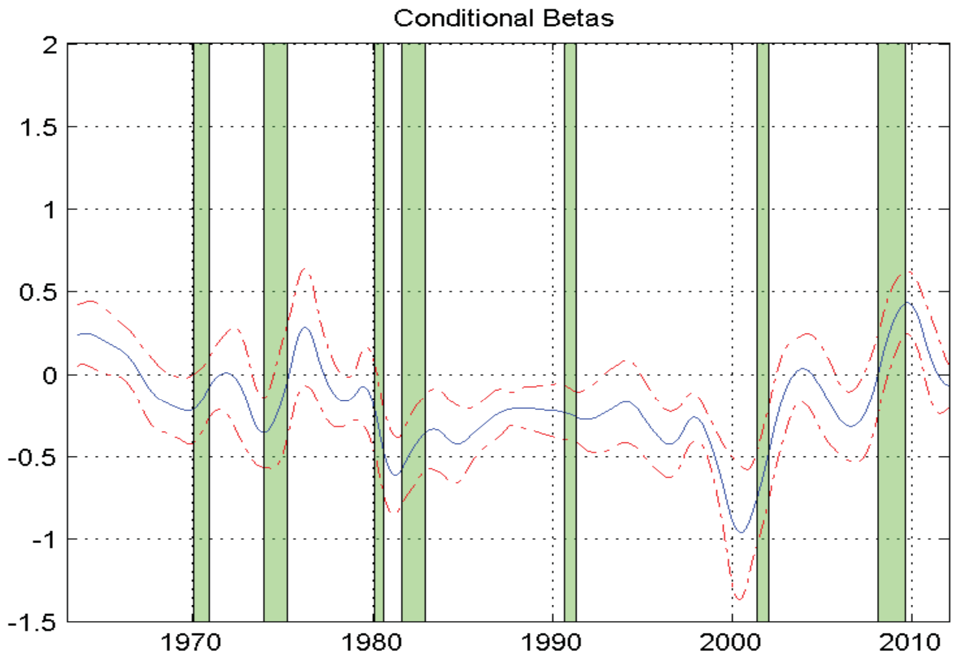
results in Table 2.6 and those of Ang and Kristensen (2011) may be due to estimation or data difference. To address this issue, we have implemented a local polynomial smoothing algorithm, similar to Ang and Kristensen (2011). Smoothing parameter is chosen by visually checking conditional beta curve to resemble Ang and Kristensen (2011). The regression results are similar with that of Table 2.6. We conclude that the difference is not caused by estimation methods such as spline estimation but by the data. Note that we used monthly data throughout our study while Ang and Kristensen (2011) used daily return data. Since daily data are high volatile with significant level of noises, even with smoothing techniques such as kernel smoothing, the estimates of conditional beta may be sensitive to noise.

Our method can be viewed as a generalization of Ang and Kristensen (2011) because not only we can estimate their model but also we can always include the conditional beta estimated from their model in our state variable pool. To illustrate the latter, we treat the estimated conditional beta from Ang and Kristensen (2011) as realized beta (BETAT) and conduct variable selection the same way as in Section 2.4.3. We identify the model with BETAT, DEF, TERM and INF in the state variables as the best model. The testing results are reported in Table 2.6 Panel A-D. Both functional form of alpha and beta are linear in single-index while the long-run alpha is significantly positive. The tests on the single-index coefficients indicate that besides BETAT, the default premium and term premium remain significant. That is to say, realized beta from Ang and Kristensen (2011) does not subsume all the other state variables.

2.4.6 Quarterly Data and Consumption-based CAPM

As a robustness check, we also investigate the conditional CAPM using quarterly data. We convert monthly instrumental variables into quarterly data by using the last month obser-

Figure 2.5: Estimated time-varying realized beta



Note: the figure shows monthly estimates of time-varying conditional betas of the value premium. The conditioning beta is modeled as purely time-varying as in Ang and Kristensen (2011). We plot the conditional betas in blue line along with 95% confidence bands in red line. We also shade the NBER recession periods in green horizon bars.

Table 2.6: Tests on Alphas and Betas with time-varying realized beta (BETAT), DEF, TERM and INF as conditioning variables

Panel A: Tests on alpha				
Hypothesis		Wald Statistic		P Value
$H_o : \alpha = zero$		44.27		0.0000
$H_o : \alpha = Constant$		13.62		0.0583
Panel B: Tests on long-run alpha				
Hypothesis	Estimate	s.e.	t Value	P Value
$H_o : \alpha_{LR} = 0$	0.0042	0.0010	4.61	0.0001
Panel C: Tests on beta				
Hypothesis		Wald Statistic		P Value
$H_o : \beta = 0$		153.71		0.0000
$H_o : \beta = Constant$		112.92		0.0000
$H_o : \beta = Linear$		0.60		0.9964
Panel D: Tests on beta when alpha is modeled as constant				
Hypothesis		Wald Statistic		P Value
$H_o : \beta = 0$		153.49		0.0000
$H_o : \beta = Constant$		108.34		0.0000
$H_o : \beta = Linear$		0.85		0.9906
Panel E: Tests on single-index coefficients				
	Estimate	s.e.	t Value	P Value
$\gamma_1(BETAT)$	0.0897	.0383	2.34	0.0212
$\gamma_2(DEF)$	0.9536	.1022	9.33	0.0000
$\gamma_3(TERM)$	-0.2507	.1412	-1.78	0.0789
$\gamma_4(INF)$	-0.1402	.1053	-1.33	0.1861
Panel F: OLS regression of BETAT onto INF, PE and UE				
	Estimate	s.e.	t Value	P Value
Intercept	0.32	.04	7.95	0.0000
INF	-4.74	.42	-11.42	0.0000
PE	-1.64	.14	-11.89	0.0000
UE	0.27	.06	4.88	0.0000

Panel A-D report statistical test results on conditional and long-run alphas and betas in the single-index long-run model. Tests on single-index coefficients parameters are also reported in Panel E. Panel F presents the regression results of time-varying realized beta specified in Ang and Kristensen (2011) onto the best subset variables that we have found through variable selection. Time-varying realized beta (BETAT) is estimated from model (2.2) where single-index is replaced by time, which reduces to the model specification of Ang and Kristensen (2011). Penalized splines are used to estimate the flexible functions of conditional alpha and beta, where in Ang and Kristensen (2011) local polynomials are utilized. Tests on conditional alphas and betas are based on the Wald statistics in equation 2.10. Test on long-run alpha is based on equation 2.11. All covariance matrices and standard errors are calculated from bootstrap samples. The monthly data sample is from July 1963 to December 2012.

vations in each quarter. We convert monthly returns into quarterly returns through simple compounding. Following Lettau and Ludvigson (2001), we also include the consumption-wealth ratio, CAY, as a candidate conditioning variable.[10] Note that, because CAY is available only at the quarterly frequency, we cannot use it in our monthly analysis. When we exclude the realized beta as a conditioning variable, we identify DP, VOL and UE as significant conditioning variables. When we allow the realized beta as a candidate conditioning variable, both realized beta (BETAT) and realized market volatility (VOL) are identified as significant conditioning variables. Again, while we find strong countercyclical variations in conditional beta with the business cycle, the conditional CAPM does not fully explain the value premium.

Lettau and Ludvigson (2001) find that the conditional consumption-based CAPM helps explain the value premium using the Fama and MacBeth (1973) cross-sectional regression. To address this issue, we investigate whether the consumption risk accounts for the value premium in the time-series regression. Specifically, we regress the value premium on the contemporaneous consumption growth and allow the coefficient to be a nonlinear function of the instrumental variables, including CAY. We identify DEF, RREL, INF and VOL as the significant conditioning variables. Again, we find that the conditional beta comoves with the business cycle. More importantly, the long-run alpha is significantly positive at the 1% level. To summarize, we find that neither the CAPM nor the consumption-based CAPM explain the value premium using the quarterly data. For brevity, the results for quarterly and consumption data are not reported but are available upon request.

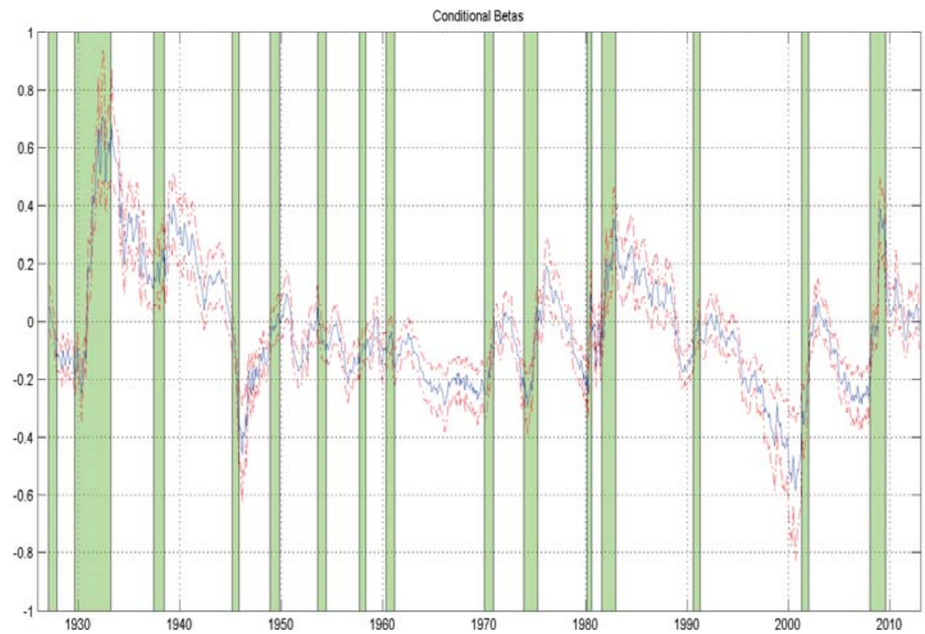
[10]We obtain the CAY variable from Martin Lettau at the University of California at Berkeley.

2.4.7 Different Samples

To check the robustness of our results, we consider different samples in this paper: January 1927 – June 1963, July 1963 – December 2012, and January 1927 – December 2012. Since unemployment rate is not always available before 1963,[11] we exclude UE from our state variable pool in the pre-1963 sample and the full sample. For comparison, we also exclude UE in the post-1963 sample as another robust check. We summarize our results in Table 2.7. Recall that the best subset variables are INF, PE and UE in the post-1963 sample (INF, PE and BETAT when include realized beta). Without the UE in the candidate variable pool, the variable selection results are somewhat consistent. For example, PE is consistently identified as significant variable regardless sample periods. Besides PE, other common variables consistently identified as significant is DEF and IP in full sample and pre-1963 sample, INF in the full sample and post-1963 sample. Another observation is, when realized beta is dropped from the pool, TERM becomes significant. In terms of long-run alpha, the full sample and post-1963 sample has positive alpha while the pre-1963 sample long-run alpha is not significantly different from zero. This is consistent with the general consensus that CAPM can explain the value premium before 1963 but not after (Ang and Chen 2005). However, we do observe strong countercyclical variation in the conditional beta of the value premium. We plot the estimated conditional beta across time in Figure 2.6. Once again, we observe conditional beta rises sharply in business recessions. Although this observation is not as consistent as the sample after 1963 as observed in Figure 2.1, the deviations are only obvious in 1927 recession and 1945 post second world war recession.

[11]Unemployment rate is available after January 1949.

Figure 2.6: Conditional beta in the full sample (1927–2012)



Note: the figure shows monthly estimates of conditional betas of the value premium in the full sample, which is from January 1927 to December 2012. The conditioning variables are default premium (DEF), industrial production (IP), inflation (INF), Price-Earning ratio (PE) and term premium (TERM). We plot the conditional betas in blue line along with 95% confidence bands in red line. We also shade the NBER recession periods in green horizon bars.

Table 2.7: Conditional CAPM–Different samples

Sample	Variable Pool	Best Variable Set	α_{LR}
Full Sample	No BETAT	DEF, INF, IP, PE, TERM	0.0031*
	Include BETAT	DEF, INF, IP, PE, BETAT	0.0025**
Pre-1963 Sample	No BETAT	DEF, IP, PE, TERM	0.0014
	Include BETAT	DEF, IP, PE, TERM	0.0014
Post-1963 Sample	No BETAT	INF, PE, TERM	0.0043**
	Include BETAT	INF, PE, BETAT	0.0040**

Note: this table presents the variable selection results for different samples and their long-run alpha estimates. Full sample stands for the monthly sample January 1927 – December 2012; Pre-1963 stands for the monthly sample January 1927 – June 1963; Post-1963 stands for the monthly sample July 1963 – December 2012. α_{LR} is the long-run alpha defined in (2.11). We mark rejection at the 95% level with * and 99% level with **.

2.5 Conclusion

We revisit whether the conditional CAPM helps explain the value premium using the single-index varying-coefficient model. Our setup provides a general framework that has two innovative features, compared with those adopted in the previous studies. First, it allows for a nonlinear dependence of conditional beta on the state variables and provides formal tests on the nonlinearity. Second, we can use an exhaustive variable selection method to choose the significant conditioning variables from a large set of potential state variables. To take advantage of the second feature, we consider several important measures of business cycles, in addition to the stock market return predictors commonly used in the previous studies. For comparison, we also include realized beta as a conditioning variable. Other robustness checks with smoothed realized beta, quarterly data, or different samples yield qualitatively robust results.

Consistent with risk-based explanations of the value premium, we find strongly countercyclical variation in its conditional beta. However, the conditional CAPM does not fully explain the value premium. Our results suggest that we should not rule out the risk-based explanation; rather, they indicate that the market return is not a sufficient statistic for the risk. Existing studies have found that other risks matter as well. For example, Fama and French (1996) suggest that value stocks are more vulnerable to distress risk than are growth stocks. Pastor and Stambaugh (2003) show that the illiquidity risk is significantly priced in the cross-section of stock returns. Campbell and Vuolteenaho (2004) find that cash-flow shocks are riskier than discount-rate shocks. Basal and Yaron (2004) argue that long-run cash flow risk accounts for a large portion of the observed equity premium. Kuehn, Petrosky-Nadeau, and Zhang (2012) show that unexpected changes in labor income are a potentially important risk that investors want to hedge against. Because these risks tend to move closely with the market risk, a countercyclical market beta might suggest that the value premium has larger exposures to these risks during business downturns than during business upturns.

Bibliography

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

Ang, A., D. Christensen. 2011. Testing Conditional Factor Models. *Journal of Financial Economics*, 106(1):132-156.

Campbell, J. Y., T. Vuolteenaho. 2004. Inflation illusion and stock prices. *American Economic Review*, 94(2):19-23.

Carroll, R.J., J. Fan, I. Gijbels, and M.P. Wand. 1997. Generalized Partially Linear Single-Index Models. *Journal of the American Statistical Association*, 92: 477-489.

Carroll, R. J., C. J. F. Wu, D. Ruppert. 1988. The effect of estimating weights in weighted least squares. *Journal of the American Statistical Association*, 83:1045-1054.

Chaudhuri, P. 1991a. Global Nonparametric Estimation of Conditional Quantile Functions and Their Derivatives. *Journal of Multivariate Analysis*, 39: 246-269.

Chaudhuri, P. 1991b. Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation. *Annals of Statistics*, 19: 760-777.

Chaudhuri, P., K. Doksum, A. Samarov. 1997. On Average Derivative Quantile Regression. *Annals of Statistics*, 25: 715-744.

Chen, S., S. Khan. 2001. Semiparametric Estimation of a Partially Linear Censored Regression Model, *Econometric Theory*, 17: 567-590.

Chen, N.-F., R. Roll, S. A. Ross. 1986. Economic Forces and the Stock Market. *The Journal of Business*, 59(3):383-403.

Christensen B.J., N.R. Prabhala. 1998. The relation between implied and realized volatility. *Journal of Financial Economics*, 50(2):125-150.

Crainiceanu, C., D. Ruppert, G. Claeskens, M. Wand. 2005. Exact Likelihood ratio tests for penalized splines. *Biometrika*, 92(1):91-103.

Cremers, K. M. 2002. Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies*, 15(4):1223-1249.

Cui, X., W. Härdle, L. Zhu. 2011. The EFM approach for single-index models. *Annals of Statistics*, 39(3):1658-1688.

De Gooijer, J.G., D. Zerom. 2003. On Additive Conditional Quantiles With High Dimensional Covariates. *Journal of the American Statistical Association*, 98: 135-146.

Fama, E. F., K. R. French. 1992. The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427-465.

Fama, E. F., K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3-56.

Fama, E. F., K. R. French. 1996. Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, 51(1):55-84.

Fama, E. F., K. R. French. 2006. The value premium and the capm. *Journal of Finance*, 61(5):2163-2185.

Fama, E. F., J. D. MacBeth. 1973. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81(1973): 607-636.

Fan, J., Q. Yao, Z. Cai. 2003. Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(1):57-80.

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

Ferson, W. E., C. R. Harvey. 1999. Conditioning variables and the cross section of stock returns. *The Journal of Finance*, 54(4):1325-1360.

Ghysels, E. 1998. On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *The Journal of Finance*, 53(2):549-573.

Guo, H., R. F. Whitelaw. 2006. Uncovering the Risk-Return Relation in the Stock Market. *The Journal of Finance*, 61(3):1433-1463.

Härdle, W., Y. Ritov, S. Song. 2010. Bootstrap Confidence Bands and Partial Linear Quantile Regression. *Journal of Multivariate Analysis*, 107: 244-262.

Hastie, T., R. Tibshirani. 1990. Generalized Additive Models, Chapman and Hall.

He, X., P. Ng. 1999. Quantile Splines With Several Covariates, *Journal of Statistical Planning and Inference*, 75(2): 343-352.

He, X., P. Ng, S. Portnoy. 1998. Bivariate Quantile Smoothing Splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3): 537-550.

He, X., H. Liang. 2000. Quantile Regression Estimates for a Class of Linear and Partially Linear Errors-in-variables Models. *Statistica Sinica*, 10(1): 129-140.

Horowitz, J. L., S. Lee. 2005. Nonparametric Estimation of an Additive Quantile Regression Model. *Journal of American Statistical Association*, 100(472): 1238.

Jagannathan R., Z. Wang. 1996. The conditional capm and the cross-section of expected returns. *The Journal of Finance*, 51(1):3-53.

Jarrow, R., D. Ruppert, Y. Yu. 2004. Estimating the term structure of corporate debt with a semiparametric penalized spline model, *Journal of the American Statistical Association*, 99: 57-66.

Koenker, R., G. Bassett. 1978. Regression Quantiles. *Econometrica*, 46: 33-50.

Koenker, R. 2005. Quantile Regression. Cambridge University Press.

Koenker, R., P. Ng., S. Portnoy. 1994. Quantile Smoothing Splines. *Biometrika*, 81(4): 673-680.

Kong, E., Y. Xia. 2012. A Single-Index Quantile Regression Model and Its Estimation. *Econometric Theory*, 28: 730-768.

Kuehn, L.-A., N. Petrosky-Nadeau, L. Zhang. 2012. An Equilibrium Asset Pricing Model with Labor Market Search. Charles A. Dice Center Working Paper No. 2012-1.

Lee, S. 2003. Efficient Semiparametric Estimation of a Partially Linear Quantile Regression Model. *Econometric Theory*, 19(01): 1-31.

Lettau M., S. Ludvigson. 2001. Resurrecting the (c)capm: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, 109(6):1238-1287.

Lewellen J., S. Nagel. 2006. The conditional capm does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82(2):289-314.

Li, K. C. 1991. Sliced Inverse Regression for Dimension Reduction, *Journal of the American Statistical Association*, 86:316-342.

Li, Y., Y. Liu, J. Zhu. 2007. Quantile Regression in Reproducing Kernel Hilbert Spaces. *Journal of American Statistical Association*, 102(477): 255-268.

Li, Y., L. Yang. 2011. Underconditioning and Overconditioning: Testing the Conditional CAPM and the Conditional Fama-French Three-Factor Model, *Journal of Empirical Finance*, 18(5): 972-992.

Mammen, E. 1993. Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics*, 21:255285.

Nychka, D., G. Gray, P. Haaland, D. Martin, M. O'Connell. 1995. A Nonparametric Regression Approach to Syringe Grading for Quality Improvement. *Journal of American Statistical Association*, 90(432): 1171-1178.

Pastor L., R. F. Stambaugh. 2003. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*, 111:642-85.

Petkova R., L. Zhang. 2005. Is value riskier than growth? *Journal of Financial Economics*, 78(1):187-202.

Ruppert, D. 2002. Selecting the Number of Knots for Penalized Splines. *Journal of Computational and Graphical Statistics*, 11: 735-757.

Ruppert, D., M. P. Wand, R. J. Carroll. 2003. Semiparametric Regression. New York: Cambridge University Press.

Shanken, J., 1990. Intertemporal asset pricing : An Empirical Investigation. *Journal of Econometrics*, 45(1-2):99-120.

Stock, J., M. Watson. 2003. Has the business cycle changed? Evidence and explanations. Monetary policy and uncertainty, Federal Reserve Bank of Kansas City, 9-56.

Wang, K. Q. 2003. Asset pricing with conditioning information: A new test. *The Journal of Finance*, 58(1):161-196.

Wang, J., L. Xue, L. Zhu, Y. Chong. 2010. Estimation for a Partial-Linear Single-Index Model. *The Annals of Statistics*, 38: 246-274.

Wang, H., Z. Zhu, J. Zhou. 2009. Quantile Regression in Partially Linear Varying Coefficient Models. *The Annals of Statistics*, 37: 3841-3866.

Welch, I., A. Goyal. 2008. A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, 21(4):1455-1508.

Wu, C.F.J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis (with discussions). *Annals of Statistics*, 14:1261-1350.

Wu, T. Z., H. Lin, Y. Yu. 2011. Single Index Coefficient Models for Nonlinear Time Series. *Journal of Nonparametric Statistics*, 23(1):37-58.

Wu, Z., K. Yu, Y. Yu. 2010. Single-Index Quantile Regression. *Journal of Multivariate Analysis*, 101: 1607-1621.

Xia, Y., W. Härdle. 2006. Semi-Parametric Estimation of Partially Linear Single-index Models. *Journal of Multivariate Analysis*, 97: 1162-1184.

Xia Y., W. K. Li. 1999. On single-index coefficient regression models. *Journal of the American Statistical Association*, 94(448):1275-1285.

Yu, K., M. C. Jones. 1998. Local Linear Quantile Regression. *Journal of American Statistical Association*. 93: 228-238.

Yu, K., Z. Lu. 2004. Local Linear Additive Quantile Regression. *Scandinavian journal of statistics*, 31: 333-346.

Yu, Y., D. Ruppert. 2002. Penalized Spline Estimation for Partially Linear Single-Index Models. *Journal of American Statistical Association*, 97: 1042-1054.

Yu, Y., K. Yu, M. Li, Q. Wang. 2009. Semiparametric Estimation for Time-Inhomogeneous Diffusions. *Statistica Sinica*, 19:843-867.

Yuan, M. 2006. GACV for Quantile Smoothing Splines. *Computational Statistics and Data Analysis*, 5: 813-829.

Zhang, L. 2005. The value premium. *The Journal of Finance*, 60(1):67-103.

Zhu, L., M. Huang, R. Li. 2011. Semiparametric Quantile Regression with High-Dimensional Covariates. *Statistica Sinica*, 22:1379-1401.