Published in final edited form as:

Comput Stat Data Anal. 2010 April 1; 54(4): 891–905. doi:10.1016/j.csda.2009.09.036.

Penalized spline estimation for functional coefficient regression models

Yanrong Cao^a, Haiqun Lin^b, Tracy Z. Wu^c, and Yan Yu^d

Yanrong Cao: Yanrong_Cao@manulife.com; Haiqun Lin: haiqun.lin@yale.edu; Tracy Z. Wu: tracy.z.wu@jpmchase.com; Yan Yu: Yan.Yu@uc.edu

^aManulife Financial, 200 Bloor St. E., Toronto, ON M4W1E5, Canada

^bYale University, United States

^cJP Morgan Chase Bank, 1111 Polaris Parkway Columbus, OH 43240, United States

^dThe University of Cincinnati, PO BOX 210130, Cincinnati, OH 45221, United States

Abstract

The functional coefficient regression models assume that the regression coefficients vary with some “threshold” variable, providing appreciable flexibility in capturing the underlying dynamics in data and avoiding the so-called “curse of dimensionality” in multivariate nonparametric estimation. We first investigate the estimation, inference, and forecasting for the functional coefficient regression models with dependent observations via penalized splines. The P-spline approach, as a direct ridge regression shrinkage type global smoothing method, is computationally efficient and stable. With established fixed-knot asymptotics, inference is readily available. Exact inference can be obtained for fixed smoothing parameter λ , which is most appealing for finite samples. Our penalized spline approach gives an explicit model expression, which also enables multi-step-ahead forecasting via simulations. Furthermore, we examine different methods of choosing the important smoothing parameter λ : modified multi-fold cross-validation (MCV), generalized cross-validation (GCV), and an extension of empirical bias bandwidth selection (EBBS) to P-splines. In addition, we implement smoothing parameter selection using mixed model framework through restricted maximum likelihood (REML) for P-spline functional coefficient regression models with independent observations. The P-spline approach also easily allows different smoothness for different functional coefficients, which is enabled by assigning different penalty λ accordingly. We demonstrate the proposed approach by both simulation examples and a real data application.

1. Introduction

Functional (varying) coefficient regression (FCR) model (Hastie and Tibshirani, 1993; Chen and Tsay, 1993; Cai et al., 2000b) is an important tool in multivariate nonparametric analysis. The regression function of FCR has the form

$$E(y|\mathbf{X}=\mathbf{x}, U=u) = \sum_{j=1}^d a_j(u)x_j, \quad (1)$$

where y is a response variable, $\mathbf{x}$ is d -dimensional covariate, u is a lower dimensional (often univariate) threshold variable, and $a_j(\cdot)$ is an unknown function of threshold variable u and often called the functional (varying) coefficient. By allowing the varying feature of the regression coefficients, FCR is able to capture some nonlinearity.

Some familiar nonlinear time series models are within the framework of the functional coefficient regression model, for example, the exponential autoregressive model (EXPAR, Haggan and Ozaki, 1981), threshold autoregressive model (TAR, Tong, 1990), and functional autoregressive model (FAR, Chen and Tsay, 1993). The FCR models have found important applications in many areas, such as ecological studies and time series modeling. Recently, Huang and Shen (2004) used functional coefficient models to estimate and to forecast US Gross National Product (GNP), the empirical behavior of which usually displays severe nonlinearities.

One common approach to estimating FCR models is the local method. Chen and Tsay (1993) used local constant approach and developed a nice iterative algorithm for estimation of $a_j(\cdot)$. Xia and Li (1999) and Cai et al. (2000a) used local linear technique. While local approaches are very nice, it is often evidenced that penalized spline (P-spline) approach, as a global estimation approach, may offer computational advantage with efficiency and stableness (Ruppert et al., 2003; Yu and Ruppert, 2002; Jarrow et al., 2004). In addition, as a global estimation with an explicit model expression, spline approach enables easy-to-implement multi-step-ahead forecasting (Huang and Shen, 2004). Different smoothness for different functional coefficients, which is enabled by assigning different penalty λ , can also be easily implemented using P-spline approach.

Huang and Shen (2004) introduce polynomial (regression) spline approach to estimating the coefficient functions globally. Generally there are three approaches to spline estimation: smoothing splines, regression splines, and penalized splines. Smoothing splines require as many parameters as the number of observations. Regression splines use a small number of knots placed judiciously and elaborate algorithms such as the TURBO (Friedman and Silverman, 1989) and MARS (Friedman, 1991) are needed for knots selection. Penalized splines (O'Sullivan, 1986; Eilers and Marx, 1996; Ruppert and Carroll, 2000; Ruppert, 2002) combine the features of smoothing splines and regression splines. A roughness penalty is incorporated with a relative large number of knots. Smoothing parameter is used to balance the goodness-of-fit and smoothness. The number and location of knots are no longer crucial as long as the minimum number of knots is reached. Equally spaced knots are used in Huang and Shen (2004), where the number of knots is selected by AIC criterion. Usually, the optimal number of knots is chosen in the range from 2 to 10. They are applying the regression spline technique in essence. Elaborate algorithm of knot placement, such as TURBO or MARS, is not utilized in order to reduce computation burden. However, the small number of knots used might result in under-fitting. Moreover, equally spaced knots may not capture the true pattern in data.

This paper is mainly concerned with the application of the penalized spline approach (Ruppert et al., 2003) to the functional coefficient regression models under dependence. We investigate P-spline estimation, inference, and forecasting for the functional coefficient regression models. In the nonlinear time series framework, let $\{Y_t, \mathbf{X}_t, U_t\}$ be jointly strictly stationary process with $\mathbf{X}_t$ endogenous or exogenous variables. We model $a_j(\cdot)$ by a penalized spline. Different smoothness is allowed for different coefficient function $a_j(\cdot)$ by assigning different smoothing parameter λ_j . Exact inference is established, which is most appealing for finite samples and is also demonstrated through a Monte Carlo variance study. We also obtain fixed-knot asymptotics, similar as usual parametric modeling. As a global

smoothing method, our proposed P-spline approach provides a parsimonious explicit expression of the estimated model which facilitates the multi-step-ahead forecasting.

Selection of the penalty parameter λ to control the trade-off between the goodness of fit and smoothness is crucial. Generalized Cross-Validation (GCV) has been widely used in spline literature. However, some research shows GCV could under-smooth or over-smooth the regression curves, especially the derivatives, when possible correlation structure exists (Jarow et al., 2004). In local estimation, for example, a modified multi-fold cross-validation (MCV) criterion was used for selection of bandwidth in FCR (Cai et al., 2000a). Ruppert (1997) proposed an interesting empirical bias bandwidth selection (EBBS) in univariate local smoothing. In this paper, we further implement and examine different methods of choosing the important penalty parameter λ : generalized cross-validation (GCV); modified multi-fold cross-validation (MCV); and an extension of Ruppert's empirical bias bandwidth selection (EBBS) to P-splines in FCR. Another popular candidate of smoothing parameter selection in P-spline approach uses mixed model framework through restricted maximum likelihood (REML). See Ruppert et al. (2003) for discussion on REML for univariate P-spline regression. We also implement REML for P-spline FCR with independent observations.

The rest of the paper is organized as follows. Section 2 presents the model, the proposed penalized spline approach, asymptotics and finite sample properties, and multiple-step-ahead forecasting using the fitted model. Section 3 describes the implementation. Section 4 discusses the MCV, GCV, EBBS, and REML criteria for selecting the penalty parameter. Section 5 applies the proposed approach to simulation examples and the series of quarterly US real gross national product (GNP) data. Section 6 concludes the paper.

2. Model and estimation methods

2.1. Functional coefficient models

We consider the model:

$$y_t = f(\mathbf{x}_t, u_t) + \varepsilon_t.$$

The conditional mean function is

$$f(\mathbf{x}_t, u_t) = \sum_{j=1}^d a_j(u_t) x_{tj} = a_1(u_t) x_{t1} + a_2(u_t) x_{t2} + \cdots + a_d(u_t) x_{td}, \quad (2)$$

where

1. $y_t \in \mathbf{R}$, $\mathbf{x}_t \in \mathbf{R}^d$, $u_t \in \mathbf{R}$, $t = 1, \dots, n$;
2. $a_j, j = 1, \dots, d$, are the unknown univariate coefficient functions;
3. ε_t are independent of $(\mathbf{x}_s, u_s)$ for $s \leq t$ and $\varepsilon_t, t' \neq t$; with $E(\varepsilon_t) = 0$ and constant variance σ^2 .

The above setup is suitable for both regression and time series models. In a pure time series context, both $\mathbf{x}$ and threshold variable u consist of some lagged values of y_t . Throughout this paper, we consider univariate time series model framework and ignore conditioning on exogenous variables. Thus the conditional mean $E[y_t | \mathbf{x}_t, u_t] = E[y_t | y_{t-s}, y_{t-s+1}, \dots, y_{t-k}]$ for some $0 < k \leq s < t$. Extension to models with both endogenous and exogenous variables is conceptually straightforward.

2.2. Penalized spline estimation

The unknown univariate function $a_j(\cdot)$ can be estimated by a penalized spline (Ruppert and Carroll, 2000; Ruppert, 2002). Assume that

$$a_j(u) = \delta_j + \delta_{j,1}u + \delta_{j,2}u^2 + \cdots + \delta_{j,p}u^p + \sum_{k=1}^K \delta_{j,p+k}(u - \mathcal{K}_k)_+^p, \quad (3)$$

where p is spline degree and $\{\mathcal{K}_k\}_{k=1}^K$ are spline knots. We could use different degree and knots for different function a_j . However, one of the appealing features of penalized splines is that the number and location of knots are no longer crucial and the smoothness can easily be controlled by a single smoothing parameter λ_j . We will further discuss the tuning parameter selection in Section 4. Model (3) uses the so-called truncated power function basis. This basis is convenient for illustration purpose. Other bases, for example, B-splines, can be used. Similar fits are obtained when B-spline basis is used.

Define the spline coefficient vector $\delta_j = (\delta_j, \delta_{j,1}, \dots, \delta_{j,p+K})^T$ and spline basis

$$\mathbf{B}_j(u) = \left(1 \ u \ \cdots \ u^p (u - \mathcal{K}_1)_+^p \ \cdots \ (u - \mathcal{K}_K)_+^p \right).$$

Then, our spline model for functional coefficient a_j is

$$a_j(u) = \mathbf{B}_j(u) \delta_j.$$

Also denote the row vectors

$$\begin{aligned} \mathbf{Z}_j(\mathbf{x}_s, u_s) &= \left(x_{sj} \ x_{sj}u_s \ \cdots \ x_{sj}u_s^p \ x_{sj}(u_s - \mathcal{K}_1)_+^p \ \cdots \ x_{sj}(u_s - \mathcal{K}_K)_+^p \right), \\ \mathbf{Z}_s &= \left(\mathbf{Z}_1(\mathbf{x}_s, u_s) \ \cdots \ \mathbf{Z}_j(\mathbf{x}_s, u_s) \ \cdots \ \mathbf{Z}_d(\mathbf{x}_s, u_s) \right), \end{aligned}$$

and column spline coefficient vector

$$\delta = \left(\delta_1^T, \dots, \delta_j^T, \dots, \delta_d^T \right)^T.$$

Define the “design” matrix $\mathbf{Z}$ with its s th row $\mathbf{Z}_s$. The mean function can then be represented as:

$$E[y_s | \mathbf{x}_s, u_s] := f(\mathbf{x}_s, u_s) = \mathbf{Z}_s \delta. \quad (4)$$

Let $\mathbf{D}_j$ be an appropriate semi-definite symmetric matrix. For example, $\mathbf{D}_j$ can be diagonal with its last K diagonal elements equal to one and the rest equal to zero, which penalizes on jumps in the p th derivatives of $a_j(\cdot)$ (see Ruppert and Carroll, 2000). Alternatively, $\mathbf{D}_j$ can be a positive semi-definite symmetric matrix such that

$$\delta^T \mathbf{D}_j \delta = \int \left[a_j''(s) \right]^2 ds,$$

which yields the usual quadratic integral penalty. Let $\lambda_j \geq 0$ be the penalty parameter for a_j . Separate penalty parameters are used to allow different smoothness for different coefficient functions. Denote in matrix notation

$$\lambda = \text{diag}(\lambda_1, \dots, \lambda_1, \dots, \lambda_d, \dots, \lambda_d)$$

and the penalty matrix $\mathbf{D}$ with its j th diagonal block $\mathbf{D}_j$. Let $\mathbf{y}$ be the observed time series vector.

We obtain that the penalized least squares estimator of δ is the shrinkage estimate of linear ridge regression

$$\widehat{\delta}(\lambda) = (\mathbf{Z}^T \mathbf{Z} + n\lambda \mathbf{D})^{-1} \mathbf{Z}^T \mathbf{y}, \quad (5)$$

minimizing the penalized average sum of squared error

$$Q_{n,\lambda}(\delta) = Q_n(\delta) + \lambda \delta^T \mathbf{D} \delta = \frac{1}{n} (\mathbf{y} - \mathbf{Z}\delta)^T (\mathbf{y} - \mathbf{Z}\delta) + \lambda \delta^T \mathbf{D} \delta. \quad (6)$$

The smoothing or “hat” matrix is

$$\mathbf{S}(\lambda) = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z} + n\lambda \mathbf{D})^{-1} \mathbf{Z}^T. \quad (7)$$

The model mean can be fitted by

$$f(\mathbf{x}_t, u_t) = \mathbf{S}(\lambda) \mathbf{y} = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z} + n\lambda \mathbf{D})^{-1} \mathbf{Z}^T \mathbf{y}. \quad (8)$$

2.3. Finite sample properties and inference with λ fixed

We have shown that the spline coefficients are ridge regression type linear estimator as in Eq. (5). Since both $\mathbf{x}$ and threshold variable u consist of some lagged values of y_t , the design matrix $\mathbf{Z}$ only involves lagged values of y_t . When conditional mean estimation is considered with smoothing parameter λ fixed, this is very similar to linear ridge regression with fixed design. Hence, we can get exact forms of variance estimates (or so-called Sandwich formula) as in usual ridge regression. The covariance matrix of the spline coefficients $\widehat{\delta}(\lambda)$ is

$$\text{cov}(\widehat{\delta}(\lambda)) = \sigma^2 (\mathbf{Z}^T \mathbf{Z} + n\lambda \mathbf{D})^{-1} (\mathbf{Z}^T \mathbf{Z}) (\mathbf{Z}^T \mathbf{Z} + n\lambda \mathbf{D})^{-1}. \quad (9)$$

In nonparametric regression, the trace of smoothing matrix $\mathbf{S}(\lambda)$ is often called the *degrees of freedom* of the fit (Hastie and Tibshirani, 1993), namely, $\text{df}_{\text{fit}} = \text{tr}(\mathbf{S}(\lambda))$. It has the rough interpretation as the equivalent number of parameters. A nearly unbiased estimator of σ^2 is

$$\widehat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{Z}\widehat{\delta}(\lambda))^T (\mathbf{y} - \mathbf{Z}\widehat{\delta}(\lambda))}{n - \text{tr}(\mathbf{S}(\lambda))}. \quad (10)$$

The covariance matrix of the fitted vector $\widehat{\mathbf{f}} = \mathbf{S}(\lambda) \mathbf{y}$ is

$$\widehat{\text{cov}}(\mathbf{f}) = \sigma^2 \mathbf{S}(\lambda) \mathbf{S}(\lambda)^\top. \quad (11)$$

This can be used to form pointwise confidence intervals (Hastie and Tibshirani, 1993).

2.4. Properties and inferences when $\lambda_n \rightarrow 0$

The penalized spline approach lies in a gray area between parametric and nonparametric estimation. The analysis of fixed-knot P-spline asymptotics we use is more like standard parametric modeling that the parameters can be estimated at root-n rate (Gray, 1994; Carroll et al., 1999).

We denote $\lambda = \max(\lambda_1, \dots, \lambda_d)$ and denote λ by λ_n to imply that the value of the smoothing parameter depends on the sample size. The assumptions of the following theorems are given in the Appendix. The proofs are a combination of work similar as in Yu and Ruppert (2002) and Huang and Shen (2004). To save the space, proofs are omitted but an extended version of the paper with proofs is downloadable from <http://statqa.cba.uc.edu/~yuy/CLWY-extended.pdf>.

Theorem 1. *Under the assumptions given in the Appendix, if the smoothing parameter $\lambda_n = o(1)$, then there exists a sequence of penalized least square estimators $(\hat{\delta}_{n,\lambda_n})$ minimizing Q_{n,λ_n} of Eq. (6), and is a strong consistent estimator of δ_0 .*

Theorem 2. *Under the assumptions given in the Appendix, if the smoothing parameter $\lambda_n = o(n^{-1/2})$, then a sequence of penalized least squares estimators exists and is asymptotically normally distributed*

$$n^{1/2} (\hat{\delta}_{n,\lambda_n} - \delta_0) \rightarrow N(\mathbf{0}, \sigma^2 \mathbf{\Omega}^{-1}), \quad (12)$$

where $\mathbf{\Omega}$ is the almost sure limit of $\frac{\mathbf{Z}^\top \mathbf{Z}}{n}$, which is shown to exist and positive definite in the Appendix.

Our results of (12) enable us to make inferences of the coefficient functions. A joint confidence region for a set of spline coefficients can also be constructed, which is sometimes called “global confidence sets” (Hastie and Tibshirani, 1993). However, for moderate sample size, we recommend inference based on the finite sample results in Section 2.3.

2.5. Multi-step-ahead forecast

Forecasting is always one of the important issues in time series analysis. For nonlinear time series data, it is not easy to make multi-step-ahead forecast. For example, Chen et al. (2004) recently developed a nice multi-stage local polynomial smoother, which requires optimal bandwidth selections for each of the two stage predictions. Since our penalized spline model (8) provides an explicit expression, we can follow Huang and Shen (2004) and conduct forecasting based on simulations using the explicit model. A nice feature of this simulation based method is that not only the point forecasts but the interval forecasts and density forecasts can be also obtained.

The idea is very simple. Suppose we have a time series $\{y_1, \dots, y_n\}$ that follows a functional coefficient model. From the estimation of proposed model, an explicit model expression can

be produced and thereby be used for forecasting. The k -step-ahead simulated series can be obtained by recursively using the fitted model to do 1-step-ahead prediction. In each step, the error term can be generated by random sampling with replacement (bootstrap) from the fitted model residuals. One could also consider repeated sampling of the error terms for obtaining several point predictions. We can then use the sample mean of the simulated series as the k -step-ahead point forecast. In addition, this simulated series can be used to construct interval forecasts and density forecasts. For details, we refer readers to Huang and Shen (2004). This prediction method is implemented in real data application in Section 5.

3. Implementation

We provide some guidance regarding selections of knots and threshold variables in this section. We will investigate the selection of smoothing parameters in the next.

3.1. Choosing the knots

Unlike regression splines, the number of knots as well as the knot locations of penalized splines are no longer crucial, as long as the minimum number of knots is reached. Given a fixed value of the number of knots K , we recommend that the knots be placed at equally spaced sample quantiles of the predictor variable, which in this context is the index u . For example, if there are 9 knots, then they would be at the 10th percentile, 20th percentile, etc. Ruppert (2002) has a detailed study of the choice of K . He observes that the value of K is not too important, provided it is large enough. We simply choose approximately $\min(n/4, 40)$ knots. Our experiments show that this works well.

3.2. Selection of the threshold variable

The threshold variable u may be set a priori. When there is no prior information, we can adaptively choose the threshold variable similarly as variable selection in regression models. We can select the candidate threshold by stepwise addition followed by stepwise deletion using criterion such as GCV or AIC etc. Selection of lags of significant variables $\mathbf{x}_t$ can be similarly implemented. Of course, a limitation is that u is assumed to be some univariate variable. In a more general setup, we may adopt single-index functional coefficient models where the threshold variable can be obtained as an index of the form $\boldsymbol{\alpha}^T \mathbf{x}_t$ by direct penalized (nonlinear) least squares estimation. This may be considered in future research.

4. Selection of smoothing parameters

The selection of appropriate smoothing parameters is crucial to good curve fitting. GCV that approximates CV is a computationally expedient criterion, which is popular in spline literature. In this section, we examine different methods of choosing the important penalty parameter λ : modified multi-fold cross-validation (MCV), generalized cross-validation (GCV), and an extension of empirical bias bandwidth selection (EBBS) to P-splines FCR. Empirical selection of smoothing parameters (EBBS, Ruppert, 1997) provides an alternative and efficient new method in this field. We extend current literature by (i) implementing a revised EBBS for both local method and P-spline method to estimate functional coefficient regression models; and (ii) providing a comparison of GCV, MCV and EBBS performances when used for smoothing parameter selection for both spline smoothing and local smoothing. In addition, we implement mixed model framework through restricted maximum likelihood (REML) for P-spline FCR with independent observations. These are further explored in the simulation study in Section 5.1.

4.1. MCV

In Cai et al. (2000a), a modified multi-fold cross-validation (MCV) criterion is used for selection of bandwidth. They divided the whole sample into several subseries and used Q subseries of length $n - qm$ ($q = 1, \dots, Q$) to estimate the model and to compute “1-step” forecasting error for the $n - qm + 1$ th to the $n - (q - 1)m$ th observations. Optimal

bandwidth (properly sample size scaled, i.e., $h\left(\frac{n}{n - qm}\right)^{1/5}$) minimizes forecasting errors sum over Q sub-samples, i.e.,

$$\begin{aligned} AMS(h) &= \sum_{q=1}^Q AMS_q(h), \\ AMS_q(h) &= \frac{1}{m} \sum_{t=n-qm+1}^{n-qm+m} \left[y_t - \sum_{j=1}^d \hat{a}_j(u_t) x_{tj} \right]^2. \end{aligned}$$

In practice, the authors recommend to use $m = [0.1n]$ and $Q = 4$ instead of $m = 1$, though optimal choices and resulting computational costs are not clear. We initially implemented MCV for local approach and found its performance is not very satisfactory based on our limited study (e.g. Fig. 6, Section 5.1): (i) The deviances between curve fits based on MCV bandwidth and true curves are non-ignorable; (ii) Estimation using multiple subseries is not efficient in terms of computation time.

4.2. GCV

Generalized cross-validation score (GCV) has been commonly used in spline literature. Consider P-spline smoothing when quadratic penalty is used. We choose the smoothing parameter λ by minimizing

$$GCV(\lambda) := \frac{n^{-1}(\mathbf{y} - \mathbf{Z}\hat{\delta}(\lambda))^T (\mathbf{y} - \mathbf{Z}\hat{\delta}(\lambda))}{(1 - n^{-1}\text{tr}\mathbf{S}(\lambda))^2}, \quad (13)$$

where $\mathbf{S}(\lambda)$ is the “hat” matrix such that $\hat{\mathbf{y}} = \mathbf{S}(\lambda)\mathbf{y}$. The trace of the smoothing matrix $\mathbf{S}(\lambda)$ of Eq. (7) can be calculated as follows $\text{tr}(\mathbf{S}(\lambda)) = \text{tr}\{(\mathbf{Z}^T\mathbf{Z} + n\lambda\mathbf{D})^{-1}\mathbf{Z}^T\mathbf{Z}\}$, so that only the traces of matrices of dimensions $\mathbf{Z}^T\mathbf{Z}$ need to be evaluated. GCV defined in (13) is only valid if the hat matrix $\mathbf{S}(\lambda)$ is independent of $\mathbf{y}$. When this is not the case, approximations are necessary by evaluating (13) at observed values, and its validity is not clear for time series case where lags of $\mathbf{y}$ appear in $\mathbf{S}(\lambda)$. This is also one concern that evokes our consideration for another criterion, namely EBBS, described next. For comparison purpose, we also implement GCV criterion for bandwidth selection in local linear estimation for functional coefficient regression models.

4.3. EBBS

EBBS (Empirical Bias Bandwidth Selection) developed by Ruppert (1997) for choosing the bandwidth for local regression can be extended to other smoothing parameters. We modify EBBS for use with P-spline FCR models. Jarrow et al. (2004) observed that an EBBS based method could improve the selection of smoothing parameters in an interest rate term structure case study when derivative estimation is of interest with presence of autocorrelation. EBBS minimizes an estimate of the mean squared error (MSE) of function of interest over λ through grid search. The bias of the fitted values is modeled as a function

of the smoothing parameter. The variance of the fitted values can be obtained based on asymptotic formula or an exact result if available. To estimate the MSE, the estimated bias is squared and added to the estimated variance.

For our P-spline functional coefficient regression models, EBBS minimizes an estimate of MSE of function of interest $f(\mathbf{x}_t, u_t)$ averaged across observed time grid t . To estimate

$$\text{MSE}(\widehat{f}(\mathbf{x}_t, u_t); t, \lambda) := \text{MSE}(\widehat{f}_t; t, \lambda),$$

we need to estimate both bias and variance of $\widehat{f}_t(t, \lambda)$, which is the t th fitted values of y given λ . Then $\text{MSE}(\widehat{f}_t; t, \lambda)$'s are averaged over $t = 1, \dots, n$ and minimized over λ through grid search. The variance of $\widehat{f}_t(t, \lambda)$ is estimated by Eqs. (10) and (11) in Section 2.3.

EBBS estimates bias at any fixed t by computing the fit at t for a range of values of the smoothing parameter and then fitting a curve to model bias. One implementation of EBBS for P-splines uses the fact that, to the first order, the bias of a P-spline at λ is $\gamma(t)\lambda$ for some $\gamma(t)$ (Wand, 1999). Let $\widehat{f}_t(t, \lambda)$ be depending on maturity and λ . Compute $\{\lambda_\ell, \widehat{f}_t(t, \lambda_\ell)\}$, $\ell = 1, \dots, L$, where $\lambda_1 < \dots < \lambda_L$ is the grid of values of λ used for selecting λ by GCV. For this grid, in our example we used $L = 50$ values of λ such that their logarithms to base 10 were equally spaced between -7 and 6 . For any fixed t , fit a straight line to the data $\{\lambda_\ell, \widehat{f}_t(t, \lambda_\ell) : \ell = 1, \dots, L\}$. Let the slope of the line be $\widehat{\gamma}(t)$. Then the estimate of squared bias at t and λ_ℓ is $(\widehat{\gamma}(t) \lambda_\ell)^2$. Alternatively, we may set $\lambda \rightarrow 0$ to obtain an approximate unbiased estimate as the benchmark. Its difference with estimate $\widehat{f}_t(t, \lambda_\ell)$ may be considered as the “empirical” bias estimate.

For local smoothing, bias is estimated “empirically” similar as Ruppert (1997). we calculate $\widehat{f}_t(t, h)$ at a grid of bandwidths h_ℓ . Then we fit a polynomial to $(\widehat{f}_t(t, h_\ell), h_\ell)_{\ell=1}^L$. The intercept is taken as unbiased estimate of f_t and the bias is calculated accordingly.

4.4. REML

In the usual regression context with independent data, a nice feature of P-spline estimation is that it has linear mixed model representation (see Ruppert et al., 2003). Smoothing parameters can be automatically obtained through standard mixed model framework using the restricted maximum likelihood (REML) approach. Computation implementation is straightforward with standard mixed model softwares (see Wand, 2003).

In particular, rewrite the row vectors

$$\mathbf{Z}_j(\mathbf{x}_s, u_s) = (x_{sj} \ x_{sj} u_s \ \cdots \ x_{sj} u_s^p \ x_{sj}(u_s - \mathcal{K}_1)_+^p \ \cdots \ x_{sj}(u_s - \mathcal{K}_K)_+^p) = (\mathbf{A}_j(\mathbf{x}_s, u_s) \mathbf{C}_j(\mathbf{x}_s, u_s)),$$

where $\mathbf{A}_j(\mathbf{x}_s, u_s)$ and $\mathbf{C}_j(\mathbf{x}_s, u_s)$ correspond to the polynomial part and truncated power basis part respectively. Denote

$$\mathbf{Z}_s = (\mathbf{A}_1(\mathbf{x}_s, u_s) \ \cdots \ \mathbf{A}_d(\mathbf{x}_s, u_s) \ \mathbf{C}_1(\mathbf{x}_s, u_s) \ \cdots \ \mathbf{C}_d(\mathbf{x}_s, u_s)) = (\mathbf{A}_s \ \mathbf{C}_s),$$

The column spline coefficient vector can be expressed by coefficients corresponding to polynomial part and truncated power basis part respectively. $\boldsymbol{\delta} = (\boldsymbol{\beta}^\top \boldsymbol{\gamma}^\top)^\top$. Define the “design” matrix $\mathbf{A}$ and $\mathbf{C}$ with its s th row $\mathbf{A}_s$ and $\mathbf{C}_s$.

The mixed model can then be represented as:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (14)$$

where $\boldsymbol{\gamma} \sim N(0, \sigma_\gamma^2 \mathbf{I})$, $\boldsymbol{\varepsilon} \sim N(0, \sigma_\varepsilon^2 \mathbf{I})$ for independent observations. Then we have

$$\lambda = \sigma_\varepsilon^2 / \sigma_\gamma^2.$$

It is also fairly straightforward to implement REML smoothing parameter estimates to accommodate separate smoothing parameters for different coefficient functions, by simply introducing block variance matrix for $\sigma_\gamma^2 \mathbf{I}$.

However, REML implementation with dependent observations is far from trivial. For a simple univariate P-spline smoothing with correlated errors, Krivobokova and Kauermann (2007) recently examined REML smoothing parameter selection when $\boldsymbol{\varepsilon} \sim N(0, \sigma_\varepsilon^2 \mathbf{V})$, where $\mathbf{V}$ is an unknown error correlation matrix to be estimated. The implementation is not trivial and extension to our functional coefficient regression model with dependent observations is not clear. This may be investigated in the future research.

Remark 1. When different λ 's are used for different a_j 's, minimization of penalized sum of squares over multi (d)-dimension is sometimes computationally expensive. An approximate solution can be used, e.g. Ruppert and Carroll (2000). To minimize GCV over grids of

$\{\lambda_j\}_{j=1}^d$, a 2-step algorithm is used. In the first step, GCV is minimized with a common smoothing parameter, that is, $\lambda_1 = \dots = \lambda_d = \lambda$, which is chosen from a trial sequence of grid

values. In the second step, starting with this common smoothing parameter, $\{\lambda_j\}_{j=1}^d$ are selected one at a time by minimizing the GCV criterion. More specifically, λ_i is set equal to its minimum GCV value with the other λ_j , $j \neq i$ fixed. For details of the algorithm, we refer to Ruppert and Carroll (2000). Similar follows 2-step algorithm for EBBS.

Remark 2. Like in our P-spline approach, separate smoothing parameters for coefficient functions can be adopted in local smoothing method as well. An implementation of such idea can follow the backfitting algorithm by Opsomer and Ruppert (1999). However, the use of different bandwidths should be careful since it involves heavy computing when using local smoothing in our model context and should only be used when the curvatures of coefficient functions are very much different.

5. Examples

5.1. A simulation study

We consider an EXPAR model

$$y_t = a_1(y_{t-1})y_{t-1} + a_2(y_{t-1})y_{t-2} + \varepsilon_t,$$

where $a_1(u) = 0.138 + (0.316 + 0.928u)e^{-3.89u^2}$, $a_2(u) = -0.437 - (0.659 + 1.260u)e^{-3.89u^2}$, and $\{\varepsilon_t\}$ are i.i.d. $N(0, \sigma^2)$. This model has been widely studied in the literature (Haggan and Ozaki, 1981; Cai et al., 2000a; Huang and Shen, 2004).

We replicate simulation 100 times and each time we generate a time series with length n and error noise level σ . We try different sample sizes and noise levels, namely, $n = 200, 400$,

1000, and $\sigma = .1, .2, .4$ to gain better understanding of the model performance. For penalized spline approach, we use quadratic splines (degree = 2) with approximately $\min(n/4, 40)$ number of knots placed in the quantiles (see Section 3.1). Quantile knots are nice if the true data have non-balanced change-points. The smoothing parameter λ is selected by GCV, EBBS, and MCV (see detailed discussion in Section 4). We also implement Huang and Shen (2004) in Matlab, where the tuning parameters (optimal number of knots: $nknots$) are selected by exhaustive searching in the range from 2 to 10 using the suggested AIC criterion $AIC = \log(ASE) + 2df/n$, where df is simply the total number of parameters. In this particular example, $df = 2(1 + \text{degree} + nknots)$, and the average sum squared residuals $ASE = n^{-1}(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\delta}})^T(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\delta}})$. Then equal spaced knots are placed. In principle, HS polynomial spline approach can be considered as a special case of penalized spline without penalty and with smaller number of equal spaced knots. In the study, we also implement both B-spline basis and truncated power basis. The results are very similar, which is as expected.

For ease of comparison as in Huang and Shen (2004), the Root Average Squared Error (RASE) of a_j over some fine grid points $\{u_i\}_{i=1}^{n_{grid}}$ are used. Define RASE as

$$RASE = \left\{ n_{grid}^{-1} \sum_{i=1}^{n_{grid}} [\hat{a}(u_i) - a(u_i)]^2 \right\}^{1/2}.$$

We use 240 equally spaced grid points within the data range as in HS. Huang and Shen (2004) found from simulation studies that the local linear method of Cai et al. (2000a) usually gave estimates with larger bias than those given by polynomial splines, especially around the modes. It is not surprising that they also found that the spline method, allowing different smoothness for different coefficient functions, performs better than the local method with single bandwidth. Thus, in this paper we will focus comparison on the polynomial spline approach of Huang and Shen (2004).

Table 1 gives the mean and standard error of RASEs from 100 replications for estimation of $a_1(\cdot)$ and $a_2(\cdot)$ by Huang and Shen's polynomial spline (HS) and penalized spline using GCV, MCV and EBBS. When the sample size is relatively small ($n = 200, 400$), penalized spline with GCV approach seems to consistently generate considerable gains in estimation of the functional coefficients. P-spline with GCV gives the smallest RASEs across different sample sizes and noise levels. When sample size grows as big as $n = 1000$, all the listed approaches give very similar performance in terms of RASEs. EBBS seems to give slightly smaller RASEs than HS and MCV.

We further examine the forecast performance of different approaches. Table 2 gives the Average Absolute Prediction Error (AAPE) of 1- to 12-step-ahead forecasts by Huang and Shen's polynomial spline (HS) and penalized spline using GCV, MCV and EBBS. AR(2) forecast results are also shown in the table as a reference. Interestingly, MCV seems to yield the best AAPE over different sample size and for noise levels $\sigma = .1, .2$. Mixed results are observed when noise level $\sigma = .4$. AR(2) model also yields comparable forecast error, though its functional coefficient estimates are clearly off from the simulated true function. Overall, Table 2 shows that in terms of AAPE, P-spline approaches seem to yield somewhat smaller AAPE than HS.

Fig. 1 shows the average P-spline fits of functional coefficients with smoothing parameters chosen by EBBS, GCV, MCV along with their 95% Monte Carlo confidence intervals when $n = 400$, $\sigma = .2$. Visually, penalized spline approach gives close estimation to the true function. For even smaller noise level, e.g. $\sigma = .05$, the fitted coefficient functions and true

functions virtually overlay each other. We observe from Fig. 1 that average EBBS fits are closest to the true function. The RASEs of the average fits of $a_1(\cdot)$ and $a_2(\cdot)$ by EBBS, GCV, MCV, and HS are respectively 0.004, 0.018, 0.029, 0.028; 0.009, 0.071, 0.035, 0.039. That is, the bias of EBBS functional coefficients fits seem to be the smallest. However, EBBS fits gives bigger variance, which is consistent with Table 1 findings. Overall, from our limited simulation study, if accurate estimate of the functional coefficients is one of the main objectives for FCR modeling, we would recommend GCV as the choice for P-spline smoothing parameter selection for both accuracy and computational expedience.

We then further conduct a Monte Carlo variance study. We compare the performance of the estimated covariance matrix by ridge type regression with fixed λ (or Sandwich formula, SW) discussed in Section 2.3 Eq. (7) to (11) and asymptotic form with $\lambda = 0$ of Eq. (12) in Section 2.4 with the true covariance matrix estimated by Monte Carlo simulation.

Fig. 2 gives the average penalized spline fit of functional coefficients $a_1(\cdot)$ and $a_2(\cdot)$ using GCV over 100 replications when $n = 400$, $\sigma = .2$, the corresponding 95% Monte Carlo (MC) confidence intervals, and the estimated 95% sandwich (SW) confidence intervals using Eqs. (10) and (11) in Section 2.3. We observe that the MC and Sandwich (SW) variance estimator are very close. The distance (Euclidean) between the Sandwich variance estimator and MC is also small. However, the estimated asymptotic 95% confidence intervals (not shown in the figure) assuming trivial smoothing parameter ($\lambda = 0$) almost “blow up” for coefficient $a_1(\cdot)$ and are much wider than the MC and Sandwich variance estimates for coefficient $a_2(\cdot)$. The deviation is huge. This is not too surprising since we expect that the Sandwich formula gives an exact variance, which would outperform the heavily asymptotic variance estimator overestimating the variance. This is especially true for finite samples. Carroll et al. (1999) (Section 4.4) observe similar features of “blow up” variance when the smoothing parameter converges to 0 in classical measurement error problems using P-splines and derive theoretical calculations. It appears that this feature exists similarly for our dependent time series data, which further confirms that it is important to have a roughness penalty (Carroll et al., 2004). We also experiment Monte Carlo variance study with independent regressors. We observe that Sandwich variance estimator matches MC variance estimator almost perfectly with moderate sample size and the asymptotic variance with $\lambda = 0$ slightly overestimates, which is not surprising.

In addition, Figs. 3–5 show the average P-spline fits of functional coefficients with smoothing parameters chosen by GCV, MCV, EBBS, REML along with their 95% Monte Carlo confidence intervals for a similar design model but with independent regression design $n = 200$, $\sigma = .2$. Three figures are presented for better visual display. For independent data, EBBS and GCV seem to outperform MCV and REML. Average EBBS fits are closest to the true curves. Whereas, MCV and REML fits deviate to the true curve more. For even smaller noise level, e.g. $\sigma = .05$, or bigger sample size, say, $n = 1000$, the fitted coefficient functions and true functions virtually overlay each other and all the above smoothing parameter selection criteria virtually give the same fits and confidence intervals. From our limited study, we also observe that with independent predictors in regression model setting, the fit is usually better than with dependent data in time series setting, which is not surprising.

We also implemented GCV, EBBS, and MCV criteria for local linear approach for functional coefficient regression models. We implemented MCV according to what is literally described in Cai et al. (2000a) and found its performance is not as satisfactory compared with GCV and EBBS, confirming what we have discussed in Section 4. Fig. 6 illustrates the average local linear fits with different criteria of smoothing parameter selection along with their 95% Monte Carlo confidence intervals with independent

regression design $n = 200$, $\sigma = .2$. Both EBBS and GCV give close estimates of the functional coefficients. It is interesting that in this case the extended EBBS for local linear FCR models seems to be the best.

5.2. Forecasting US real GNP

The US real GNP data is of its own right of interest and has been studied in the previous literature (see Huang and Shen, 2004). For ease of comparison, we use the same quarterly US real GNP $\{z_t\}$ from the first quarter of 1947 to the first quarter of 1991 with total 177 observations, where the present measure of real GNP is valued in 1982 prices. The data are obtained from the Citibase database and are seasonally adjusted so that data of different quarters are comparable (Tsay, 2002). We apply our approach to the transformed data

$y_t = 100 \log \left(\frac{z_t}{z_{t-1}} \right)$, the annual US Real GNP Growth Rate, with 176 data points.

For comparison of forecast performance, we split the series into two subseries. The “training” subseries consist of the first 174 data points $\{y_1, \dots, y_{174}\}$ to model the dynamic behavior of US real GNP by functional coefficient models. The remaining 12 data points $\{y_{175}, \dots, y_{176}\}$ are used to check the performance of forecasting based on the fitted model.

Following the model selection in Huang and Shen (2004), we use

$$y_t = a_1(y_{t-2})y_{t-1} + a_2(y_{t-2})y_{t-2} + \epsilon_t.$$

In implementing the proposed penalized spline approach, 40 knots are used for both $a_1(\cdot)$ and $a_2(\cdot)$. Smoothing parameter is chosen in the additive penalized spline algorithm. Penalties λ are selected using GCV. We then make forecasts for 1- to 12-step ahead (Section 2.5) based on the fitted model. HS polynomial spline method is also implemented for comparison. The optimal pair of number of equally spaced knots for $a_1(\cdot)$ and $a_2(\cdot)$ are selected by AIC from 2 to 10. We obtain 3 as the optimal numbers of knots by AIC.

Estimated GNP and 95% confidence intervals of the penalized spline fit are given in Fig. 7. Table 3 gives the Absolute Prediction Error (APE) of each of 1- to 12-step-ahead forecasts by HS polynomial spline and penalized spline approaches. AR(2) and AR(3) forecast results are also shown in the table as a reference. The penalized spline approach yields the smallest Average Absolute Prediction Error (AAPE) of 0.3344 for these 12 forecasts compared with 0.3983 given by HS polynomial spline method, 0.5460 by AR(2), and 0.5269 by AR(3). Moreover, Table 3 shows that among all 12-step forecasts, penalized spline approach generates better forecasts in all but steps 3 and 5. The estimates and forecasts from both penalized and polynomial spline approaches are illustrated in Fig. 8.

6. Conclusion

We propose penalized spline estimation for functional coefficient regression model under dependence. Different smoothness is allowed for different varying coefficients. This penalized spline approach can be fit directly by shrinkage ridge type regression and is computationally expedient. For moderate size samples, we recommend exact forms of inference for fixed λ . Different methods of choosing tuning parameters, that is, MCV, GCV, and EBBS, are considered and adapted to both local linear and P-spline approaches to functional coefficient models. From our limited study, we would recommend GCV as the choice for P-spline smoothing parameter selection for both computational efficiency and accuracy of the functional coefficient estimates. Monte Carlo variance studies are also conducted. As a global approach, our proposed P-spline approach provides an explicit model

expression which also facilitates multi-step-ahead forecasting, as that is shown in GNP data application. It is our hope that this proposed method generates a nice alternative tool in terms of estimation, inference, and forecasting for functional coefficient regression models.

Acknowledgments

We would like to thank the editors and two referees for their extremely helpful comments, which led to significant improvements to the paper.

Appendix

Assumption 1. The marginal density of U_t is bounded away from 0 and infinity uniformly on $\mathcal{C}$

Assumption 2. The eigenvalues of $E(\mathbf{X}_t \mathbf{X}_t^\top | u_t)$ are uniformly bounded away from 0 and infinity for all $u \in \mathcal{C}$

Assumption 3. $\{(Y_t, \mathbf{X}_t, U_t)\}$ is jointly strictly stationary and ergodic.

Assumption 4. $E(|x_{tj}|^m) < \infty, j = 1, \dots, d$ for reasonably large m . $E(|\varepsilon_t|^{4+r}) < \infty$ for some $r > 0$.

Assumption 2 is necessary for identification of the coefficient functions in the functional coefficient model. Assumptions 1, 3 and 4 are commonly used in the literature.

References

- Cai Z, Fan J, Li R. Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*. 2000a; 95:888–902.
- Cai Z, Fan J, Yao Q. Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*. 2000b; 95:941–956.
- Carroll RJ, Maca JD, Ruppert D. Nonparametric regression with errors in covariate. *Biometrika*. 1999; 86:541–554.
- Carroll RJ, Ruppert D, Crainiceanu C, Tosteson TD, Karagas MR. Nonlinear and nonparametric regression and instrumental variables. *Journal of the American Statistical Association*. 2004; 99:736–750.
- Chen R, Tsay RS. Functional-coefficient autoregressive models. *Journal of the American Statistical Association*. 1993; 88:298–308.
- Chen R, Yang L, Hafner C. Nonparametric multi-step ahead prediction in time series analysis. *Journal of the Royal Statistical Society: Series B*. 2004; 66:669–686.
- Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science*. 1996; 11:89–102.
- Friedman JH, Silverman B. Flexible parsimonious smoothing and additive modelling (with discussion). *Technometrics*. 1989; 31:3–39.
- Friedman JH. Multivariate adaptive regression splines (with discussion). *The Annals of Statistics*. 1991; 19:1–141.
- Gray RJ. Spline-based tests in survival analysis. *Biometrics*. 1994; 50:640–652. [PubMed: 7981391]
- Haggan V, Ozaki T. Modeling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*. 1981; 68:189–196.
- Hastie TJ, Tibshirani R. Varying-coefficient models. *Journal of the Royal Statistical Society: Series B*. 1993; 55:757–796.
- Huang JZ, Shen H. Functional coefficient regression models for nonlinear time series: A polynomial spline approach. *Scandinavian Journal of Statistics*. 2004; 31:515–534.

- Jarrow R, Ruppert D, Yu Y. Estimating the interest rate term structure of corporate debt with a semiparametric penalized spline model. *Journal of the American Statistical Association*. 2004; 99:57–65.
- Krivobokova T, Kauermann G. A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association*. 2007; 102:1328–1337.
- Opsomer J, Ruppert D. A root-n consistent backfitting estimator for semiparametric additive modelling. *Journal of Computational and Graphical Statistics*. 1999; 8:715–732.
- O’Sullivan F. A Statistical perspective on ill-posed inverse problems. *Statistical Science*. 1986; 1:502–518.
- Ruppert D. Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Journal of the American Statistical Association*. 1997; 92:1049–1062.
- Ruppert D. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*. 2002; 11:735–757.
- Ruppert D, Carroll RJ. Spatially-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics*. 2000; 42:205–223.
- Ruppert, D.; Wand, MP.; Carroll, RJ. *Semiparametric Regression*. Cambridge University Press; 2003.
- Tong, H. *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press; 1990.
- Tsay, RS. *Analysis of Financial Time Series*. Wiley; 2002.
- Wand MP. On the optimal amount of smoothing in penalised spline regression. *Biometrika*. 1999; 86:936–940.
- Wand MP. Smoothing and mixed models. *Computational Statistics*. 2003; 18:223–249.
- Xia Y, Li WK. On the estimation and testing of functional-coefficient linear models. *Statistica Sinica*. 1999; 9:735–758.
- Yu Y, Ruppert D. Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*. 2002; 97:1042–1054.

Further reading

- Fan J, Zhang W. Statistical estimation in varying coefficient models. *The Annals of Statistics*. 1999; 27:1491–1518.
- Fan J, Yao Q, Cai Z. Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society: Series B*. 2003; 65:57–80.

Fig. 1. EXPAR. The upper plot: $a_1(\cdot)$. The lower plot: $a_2(\cdot)$. The average P-spline fits of functional coefficients with smoothing parameters chosen by EBBS, GCV, MCV along with their 95% Monte Carlo confidence intervals are shown respectively with the true function for dependent observations with $n = 400$, $\sigma = .2$.

Fig. 2. EXPAR. The upper plot: $a_1(\cdot)$. The lower plot: $a_2(\cdot)$. True functions and average penalized spline (GCV) estimates. The solid curves are the corresponding 95% Monte Carlo (MC) confidence intervals. The dashed curves are the estimated 95% Sandwich (SW) confidence intervals.

Fig. 3.

The upper plot: $a_1(\cdot)$. The lower plot: $a_2(\cdot)$. True functions and the average P-spline estimates based on GCV, EBBS, REML, and MCV are presented for independent observations with $n = 200$, $\sigma = .2$.

Fig. 4.

The average P-spline fits of functional coefficients $a_1(\cdot)$ with smoothing parameters chosen by GCV, MCV, EBBS, REML along with their 95% Monte Carlo confidence intervals are shown respectively with the true function for independent observations with $n = 200$, $\sigma = .2$.

Fig. 5.

The average P-spline fits of functional coefficients $a_2(\cdot)$ with smoothing parameters chosen by GCV, MCV, EBBS, REML along with their 95% Monte Carlo confidence intervals are shown respectively with the true function for independent observations with $n = 200$, $\sigma = .2$.

Fig. 6.

The upper plot: $a_1(\cdot)$. The lower plot: $a_2(\cdot)$. The average local linear fits by GCV, EBBS, and MCV along with their 95% Monte Carlo confidence intervals are shown respectively with the true function for independent observations with $n = 200$, $\sigma = .2$.

Fig. 7.
Estimation and 95% confidence intervals with penalized spline method (GCV).

Fig. 8. Estimates and forecasts of US real GNP growth rate with penalized spline (GCV) and HS polynomial spline.

Table 1

EXPAR simulation example. Mean (Lines 3, 7, 11) and standard error (Lines 4, 8, 12) of RASEs (10^{-2}) for estimation of $a_1(\cdot)$; and mean (Lines 5, 9, 13) and standard error (Lines 6, 10, 14) of RASEs (10^{-2}) for estimation of $a_2(\cdot)$; by Huang and Shen's polynomial spline (HS) and penalized spline using GCV, MCV and EBBS.

σ	$n = 200$					$n = 400$					$n = 1000$				
	HS	GCV	MCV	EBBS		HS	GCV	MCV	EBBS		HS	GCV	MCV	EBBS	
.1	6.69	5.77	6.79	6.69		4.77	4.27	4.77	4.76		2.73	2.73	2.73	2.73	
	0.31	0.29	0.33	0.31		0.24	0.22	0.24	0.24		0.12	0.11	0.12	0.12	
	4.06	3.70	4.07	4.06		3.42	3.36	3.47	3.47		2.48	2.47	2.52	2.50	
	0.14	0.14	0.14	0.14		0.12	0.12	0.12	0.12		0.09	0.09	0.10	0.09	
.2	10.62	8.45	10.57	10.57		7.26	6.00	7.14	7.12		4.33	4.07	4.31	4.31	
	0.46	0.34	0.46	0.46		0.33	0.24	0.31	0.31		0.19	0.17	0.19	0.19	
	7.05	6.56	6.97	6.97		6.08	5.64	6.04	6.00		3.99	3.84	3.90	3.97	
	0.29	0.27	0.25	0.25		0.25	0.22	0.25	0.23		0.16	0.14	0.14	0.15	
.4	19.48	13.52	19.55	19.45		12.01	9.13	11.91	11.88		2.73	2.73	2.73	2.73	
	0.83	0.57	0.87	0.82		0.47	0.32	0.46	0.45		0.12	0.11	0.12	0.12	
	13.06	11.30	13.12	13.05		9.14	8.07	9.22	9.17		2.48	2.47	2.52	2.50	
	0.40	0.38	0.42	0.40		0.27	0.25	0.29	0.27		0.09	0.09	0.10	0.09	

Table 2

Forecast prediction of EXPAR simulation example. Average Absolute Prediction Error (AAPE) by Huang and Shen's polynomial spline (HS), penalized spline using GCV, MCV, EBBS, and AR(2).

σ	Method	$n = 200$	$n = 400$	$n = 1000$
$\sigma = .1$	HS	0.3842	0.3842	0.6325
	GCV	0.3532	0.3532	0.6056
	MCV	0.3488	0.3488	0.5377
	EBBS	0.3531	0.3531	0.6890
	AR(2)	0.3556	0.3556	0.7277
$\sigma = .2$	HS	0.1443	0.3021	0.5082
	GCV	0.1468	0.2733	0.4606
	MCV	0.1390	0.2621	0.4573
	EBBS	0.1488	0.2845	0.4839
	AR(2)	0.1859	0.3045	0.4906
$\sigma = .4$	HS	0.1625	0.3126	0.5104
	GCV	0.1614	0.3094	0.5016
	MCV	0.1653	0.3090	0.4994
	EBBS	0.1627	0.3082	0.5021
	AR(2)	0.1776	0.3252	0.5003

Table 3

US real GNP growth data example: Absolute Prediction Error (APE) of 1- to 12-step-ahead forecasts.

	Penalized spline	Polynomial spline	AR(2)	AR(3)
1-step-ahead	0.1454	0.1455	0.1649	0.1772
2-step-ahead	0.2576	0.2577	0.2973	0.2078
3-step-ahead	0.1389	0.1302	0.2185	0.1344
4-step-ahead	0.1608	0.1742	0.0336	0.1130
5-step-ahead	0.2631	0.2497	0.4380	0.3884
6-step-ahead	0.1837	0.2103	0.4047	0.3723
7-step-ahead	0.4747	0.5126	0.7391	0.7209
8-step-ahead	0.0963	0.1818	0.4009	0.3884
9-step-ahead	0.3694	0.4891	0.7111	0.7017
10-step-ahead	0.0841	0.2302	0.4602	0.4513
11-step-ahead	0.8089	0.9677	1.2166	1.2081
12-step-ahead	1.0302	1.2311	1.4673	1.4587
Average APE	0.3344	0.3983	0.5460	0.5269