Estimating Curves and Derivatives with Parametric Penalized Spline Smoothing

Jiguo Cao*, Jing Cai

Department of Statistics & Actuarial Science, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

Liangliang Wang

Department of Statistics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2

Abstract

Accurate estimation of an underlying function and its derivatives is one of the central problems in statistics. Parametric forms are often proposed based on the expert opinion or prior knowledge of the underlying function. However, these strict parametric assumptions may result in biased estimates when they are not completely accurate. Meanwhile, nonparametric smoothing methods, which do not impose any parametric form, are quite flexible. We propose a parametric penalized spline smoothing method, which has the same flexibility as the nonparametric smoothing methods. It also uses the prior knowledge of the underlying function by defining an additional penalty term using the distance of the fitted function to the assumed parametric function. Our simulation studies show that the parametric penalized spline smoothing method can obtain more accurate estimates of the function and its derivatives than the penalized spline smoothing method. The parametric penalized spline smoothing method is also demonstrated by estimating the human height function and its derivatives from the real data.

Keywords: growth curve, nonlinear regression, parameter cascading

*Corresponding author, Email: jca76@sfu.ca

1. Introduction

Intensive research has been conducted to model human growth over time, known as the human height function. Several parametric models have been widely adopted in research and thought to be sensible. For example, Preece and Baines proposed PB-1 model in 1978 (Preece and Baines, 1978); Bock and Thissen suggested a triple logistic model in 1980 (Bock and Thissen, 1980); Kanefuji and Shohoji derived two models in 1990 (Kanefuji and Shohoji, 1990); and Jolicoeur and his co-workers proposed three different models from 1988 to 1992 (Jolicoeur et al., 1988, 1992). However, these parametric models may not be completely valid for the height function, because they cannot properly model the entire growth curve or the rate of change (Jolicoeur et al., 1992).

To relax the parametric assumption, it is also popular to apply nonparametric smoothing methods to estimate the human growth curve and its derivatives (Gasser et al., 1984, 1985; Ramsay et al., 1994; Ramsay and Silverman, 2005). These nonparametric smoothing methods estimate the growth curve completely from the data without making any parametric assumptions on the growth curve. A shortcoming of these methods is that they completely ignore the expert opinion or prior knowledge of the growth curve as reflected in the parametric models.

Motivated by the above dilemma, a general parametric penalized spline smoothing method is proposed in order to combine both information from the data and the prior knowledge. The parametric penalized spline smoothing method still estimates the underlying function nonparametrically. Different from the traditional nonparametric smoothing methods, the nonparametric function is evaluated with three terms: the first term measures the fit of the nonparametric function to the data, the second term measures the distance of the nonparametric function to the parametric model proposed based on expert knowledge, and the third term controls the roughness of the nonparametric function. These three terms are added together to evaluate the overall performance of the nonparametric function. The smoothing parameters control the trade-off between fitting to data, fidelity to the parametric model, and the roughness of the nonparametric function.

The rest of the paper is organized as follows. The parametric penalized spline smoothing method is introduced in Section 2. The finite sample performance of this method is evaluated in Section 3. It is also compared with the parametric nonlinear regression method and the penalized spline smoothing method. The parametric penalized spline smoothing method is demonstrated by estimating the human height function and its derivatives from the real data in Section 4. Finally, conclusions are given in Section 5. The matlab code for implementing the parametric penalized spline smoothing method is available in the website http://www.stat.sfu.ca/∼cao/Research/PPSS.html

2. Method

Suppose a function $x(t)$ is measured at some discrete points t_i, $i = 1, \cdots, n$. These measurements are denoted as y_i, and are assumed to have a mean $x(t_i)$. Our objective is to estimate the function $x(t)$ from the noisy measurements y_i.

If based on some expert (prior) knowledge on the function $x(t)$, a parametric form may be assumed for the underlying function $x(t)$, which is denoted as $h(t|\boldsymbol{\theta})$. The parameter $\boldsymbol{\theta}$ can be estimated by the regression method (Bates and Watts, 1988). However, the parametric function may not be completely valid for capturing the true underlying true function. Therefore, several nonparametric methods are proposed to estimate $x(t)$ without making any parametric assumption on $x(t)$.

One popular nonparametric method is the spline smoothing method (Ramsay and Silverman, 2005). Let $\phi_g(t)$, $g = 1, \dots, G$, be some basis functions, then any smooth function, $x(t)$, can be approximated with a linear combination of these basis functions

$$x(t) = \sum_{g=1}^{G} c_g \phi_g(t) = \boldsymbol{\phi}(t)^T \mathbf{c}\,, \tag{1}$$

where $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_G(t))^T$ is a vector of basis functions, and $\mathbf{c} = (c_1, \dots, c_G)^T$ is a vector of basis coefficients. The basis coefficients can be estimated by minimizing the sum of squared errors

$$\texttt{SSE}(\mathbf{c}) = \sum_{i=1}^{n} (y_i - x(t_i))^2\,.$$

The basis functions are often chosen as B-spline basis functions (de Boor, 2001), because they are only non-zero in a local intervals, and zero elsewhere. This property is called *compact support*, which is essential for efficient computation. The B-spline basis functions are determined by the number of knots and their locations. The number of basis functions, G, equals sum of the degree of the B-spline basis functions and the number of interior knots plus one. It is not easy to choose the optimal number of knots and their locations, which is an infinite dimensional optimization problem. Methods for knot selection are discussed in Eubank (1988), Friedman and Silverman (1989), Wahba (1990), Friedman (1991) and Stone et al. (1997). These methods select knots from a set of candidate knots using a technique similar to stepwise regression.

An alternative method is the penalized spline smoothing method. It uses a saturated number of basis functions. For example, one knot is put at each location with an observation. To prevent overfitting the data, a roughness penalty term is added to control the smoothness of the fitted curve. The roughness penalty term is often defined by the derivative of the fitted curve. The basis coefficients are then estimated by minimizing the penalized sum of squared errors

$$\texttt{PENSSE}(\mathbf{c}) = \sum_{i=1}^{n}(y_i - x(t_i))^2 + \lambda \int_{t_1}^{t_n} \left[\frac{d^m x(t)}{dt^m}\right]^2 dt\,.$$

If the nonparametric function, $x(t)$, is of the main interest to estimate, m is usually set to 2; otherwise, m is chosen as $m = j + 2$ to estimate the derivative $d^j x(t)/dt^j$, $j = 1, 2, \ldots$.

The penalized spline smoothing method completely relies on the data, and ignores any expert opinion of the underlying function. In order to make up for this shortcoming, the parametric penalized spline smoothing method is proposed to combine both information from the data and the expert opinion. Suppose some parametric function, denoted as $h(t|\boldsymbol{\theta})$, is proposed to model the underlying function, $x(t)$, based on some expert opinion, which may not be completely valid for the underlying function. The parametric penalized spline smoothing method estimates $x(t)$ as a linear combination of basis functions as defined in (1). A saturated number of basis functions are chosen here. The basis coefficients are estimated by minimizing

$$J(\mathbf{c}|\boldsymbol{\theta}) = \sum_{i=1}^{n}(y_i - x(t_i))^2 + \lambda_1 \int_{t_1}^{t_n} [x(t) - h(t|\boldsymbol{\theta})]^2 dt + \lambda_2 \int_{t_1}^{t_n} \left[\frac{d^m x(t)}{dt^m}\right]^2 dt\,, \quad (2)$$

where the first term measures the fit to the data, the second term measures the fidelity to the parametric model, and the last term controls the roughness of the fitted function. The smoothing parameters, λ_1 and λ_2, control the trade-off among these three terms.

Given any value of $\boldsymbol{\theta}$, the estimate for the basis coefficients can be derived by minimizing $J(\mathbf{c}|\boldsymbol{\theta})$:

$$\hat{\mathbf{c}}(\boldsymbol{\theta}) = [\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \mathbf{Q} + \mathbf{R}]^{-1}[\boldsymbol{\Phi}^T\mathbf{y} + \boldsymbol{\eta}(\boldsymbol{\theta})],$$

where $\boldsymbol{\Phi}$ is a $n \times G$ basis matrix with the (i, g)-entry as $\phi_g(t_i)$, the data vector $\mathbf{y} = (y_1, \ldots, y_n)^T$, and

$$\begin{aligned}
\mathbf{Q} &= \lambda_1 \int_{t_1}^{t_n} \boldsymbol{\phi}(t)\boldsymbol{\phi}(t)^T dt, \\
\mathbf{R} &= \lambda_2 \int_{t_1}^{t_n} D^m\boldsymbol{\phi}(t) D^m\boldsymbol{\phi}(t)^T dt, \\
\boldsymbol{\eta}(\boldsymbol{\theta}) &= \lambda_1 \int_{t_1}^{t_n} \boldsymbol{\phi}(t) h(\boldsymbol{\theta}, t) dt.
\end{aligned}$$

The conditional estimate $\hat{\mathbf{c}}(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$. The parameter $\boldsymbol{\theta}$ can be estimated by minimizing the sum of squared errors

$$H(\boldsymbol{\theta}) = \sum_{i=1}^{n} [y_i - \hat{x}(t_i)]^2 = \sum_{i=1}^{n} [y_i - \hat{\mathbf{c}}(\boldsymbol{\theta})^T \boldsymbol{\phi}(t_i)]^2.$$

To summarize, the basis coefficients $\mathbf{c}$ and $\boldsymbol{\theta}$ are estimated at two nested levels of optimization. In the inner optimization level, $\mathbf{c}$ is estimated conditional on $\boldsymbol{\theta}$. In the outer optimization level, $\mathbf{c}$ is removed from the parameter space as a function of $\boldsymbol{\theta}$, and $\boldsymbol{\theta}$ is then estimated. This estimation procedure is called *parameter cascading*, which has been applied to estimate differential equations (Ramsay et al., 2007) and to estimate linear mixed-effects models (Cao and Ramsay, 2010).

One could tune the trade-off between the smoothness of the fitted function and its distance to the parametric function by varying the values of λ_1 and λ_2 in (2). This flexibility is one advantage of the parametric penalized spline smoothing method. The values for λ_1 and λ_2 may be chosen subjectively based on the expert opinion of the underlying function. We use the K-fold cross validation (Efron and Tibshirani, 1993) as the objective criterion to choose the optimal values for λ_1 and λ_2.

3. Simulations

Some simulation studies are implemented to evaluate the performance of the parametric penalized spline smoothing method, and to compare it with the parametric nonlinear regression method and the penalized spline smoothing method.

3.1. Simulation 1

Many reasonable parametric height functions have been proposed. Among them, the JPA function (Jolicoeur et al., 1992), denoted as $h(t|\boldsymbol{\theta})$, and the PB function (Preece and Baines, 1978), denoted as $f(t|\boldsymbol{\theta})$, are chosen as the true functions in our simulation studies. These two parametric functions are expressed as

$$\text{JPA} \quad : \quad h(t|\boldsymbol{\theta}_1) = A\frac{[B_1(t+E)]^{C_1} + [B_2(t+E)]^{C_2} + [B_3(t+E)]^{C_3}}{1 + [B_1(t+E)]^{C_1} + [B_2(t+E)]^{C_2} + [B_3(t+E)]^{C_3}} \tag{3}$$

$$\text{PB} \quad : \quad f(t|\boldsymbol{\theta}_2) = A - \frac{2(A-B)}{\exp[(t-E)/D_1] + \exp[(t-E)/D_2]}, \tag{4}$$

where $\boldsymbol{\theta}_1 = (A, B_1, B_2, B_3, C_1, C_2, C_3, E)^T$ is a vector of parameters in the JPA function $h(t|\boldsymbol{\theta}_1)$, and $\boldsymbol{\theta}_2 = (A, B, D_1, D_2, E)^T$ is a vector of parameters in the PB function $f(t|\boldsymbol{\theta}_2)$. The true parameter values for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are set as the nonlinear regression estimates from real height measurements of one girl in a Berkeley growth study (Tuddenham and Snyder, 1954), which are $\boldsymbol{\theta}_1 = (166.91, 0.56, 0.13, 0.08, 0.53, 3.41, 23.84, 0.0001)^T$ and $\boldsymbol{\theta}_2 = (167.27, 158.27, 7.69, 0.80, 14.01)^T$. Figure 1 displays the JPA and PB functions, and their first and second derivatives using the true parameter values. It shows that these two functions are almost identical to each other using the true parameter values, but their first and second derivatives are very different. Each true function is evaluated at 31 points from age 1 to age 18, with spacing quarterly while the child is one year old, annually from two to eight years, and biannually thereafter. These time points are also consistent with the Berkeley growth data. The simulated data are then generated by adding white noise with a variance σ^2 to the true values. We use the value of 3, 7, 10 millimeters as our values of σ, and evaluate four estimation methods under different scales of measurement errors. The simulation is implemented with 500 replicates.

Four methods are compared based on how accurately they estimate the height function and the first and second derivatives from the noisy data. The

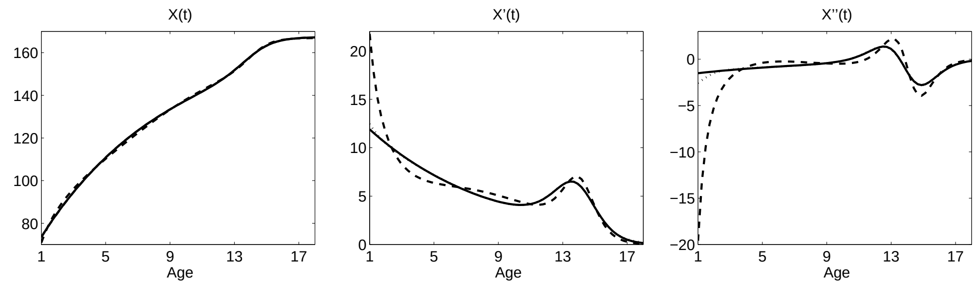

Figure 1: The JPA function (3) and the PB function (4), and their first and second derivatives using the true parameter values. The dotted lines are the estimates for the curve, $x(t)$, the first derivative, $x'(t)$, and the second derivative, $x''(t)$, using the parametric penalized spline smoothing method (PPSS) when the PB function is the true function used to simulate data. The PPSS estimates are almost identical to the true ones.

first method is the parametric nonlinear regression (PNR), where the JPA function $h(t|\boldsymbol{\theta}_1)$ is chosen as the parametric function. The second method is the penalized spline smoothing (PSS) method, in which order 6 B-spline basis functions are chosen with one knot put in each measurement location. The roughness penalty term is defined with the second or fourth derivatives of the nonparametric function, which are called PSS2 and PSS4, respectively. The third method is the parametric penalized spline smoothing (PPSS) method. It uses the same basis system as the penalized spline smoothing method, and the roughness penalty term is defined with the fourth derivatives of the nonparametric function; The parametric function $h(t|\boldsymbol{\theta}_1)$ in (2) is chosen as the JPA function. The fourth method is Wahba's standard smoothing splines method, which is implemented using Dr. Chong Gu's `gss` package in R.

These four methods are evaluated by comparing their average point-wise root mean squared errors (RMSEs) for the estimates of the height function and the first and second derivatives, which are defined as

$$\texttt{RMSE}(\hat{x}(t)) \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{500}\sum_{j=1}^{500}[\hat{x}_j(t_i) - x(t_i)]^2}\,,$$

$$\texttt{RMSE}(\hat{x}'(t)) \;=\; \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{500}\sum_{j=1}^{500}\left[\hat{x}'_j(t_i)-x'(t_i)\right]^2},$$

$$\texttt{RMSE}(\hat{x}''(t)) \;=\; \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{500}\sum_{j=1}^{500}\left[\hat{x}''_j(t_i)-x''(t_i)\right]^2},$$

where $\hat{x}_j(t_i)$, $\hat{x}'_j(t_i)$, $\hat{x}''_j(t_i)$, $i = 1, \ldots, n$, $j = 1, \ldots, 500$, are the estimated height function and the first and second derivatives in the j-th simulation at t_i, and $x(t_i)$, $x'(t_i)$, and $x''(t_i)$, are the true height function and the first and second derivatives at t_i.

When the PB function is the true function used to simulate data, Table 1 displays the average point-wise RMSEs for the estimates of the height function and the first and second derivatives using the four methods. The PPSS method is always better than the other three methods. The PNR method is sensitive to the variance of the noise, performing much better with a small variance of the noise. Both the PPSS method and PSS method are relatively robust to the variance of the noise. Because the `gss` package is only able to estimate $x(t)$, the first and second derivatives, $x'(t_i)$ and $x''(t_i)$, are estimated using the finite-difference method from the estimated $\hat{x}(t)$. The smoothing spline method has smaller RMSEs for $\hat{x}(t)$ and $\hat{x}'(t)$ than the PSS method, but the RMSE for $\hat{x}''(t)$ is very large, which may be caused by the error from the finite-difference method.

In terms of estimating the growth function $x(t)$, the PPSS method is slightly better than the PNR method, and reduced the RMSE by about 15% than the PSS4 method. When estimating the first derivative, $x'(t)$, the PPSS method only reduces RMSE by 1% than the PNR method when $\sigma = 3$, but this improvement increases to 10% when $\sigma = 10$. The RMSEs of the estimated $\hat{x}'(t)$ with the PPSS method are reduced more than 32% than those with the PSS4 method for all values of the standard deviations of the noise. Similar results are found when estimating the second derivative $x''(t)$. The RMSE of the estimated second derivatives using the PPSS method are only 55% of that using the PNR method when $\sigma = 10$. The PPSS method also reduces the RMSE of the estimated second derivatives by more than 40% than the PSS4 method for any scales of noise.

Figure 2 displays RMSEs for the estimates of the curve, $x'(t)$, the first derivative, $x'(t)$, and the second derivative, $x''(t)$, using the PPSS and PSS4 methods when the PB function is the true function used to simulate data.

Table 1: The average point-wise root mean squared errors (RMSEs) for the estimates of the height function and the first and second derivatives using three methods, when the data are simulated based on the PB function as the true function. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4), and Wahba's standard smoothing spline method (SS). The PB function is defined in (4).

Noise	**Method**	$\mathtt{RMSE}(\hat{x}(t))$	$\mathtt{RMSE}(\hat{x}'(t))$	$\mathtt{RMSE}(\hat{x}''(t))$
$\sigma = 3$	PPSS	0.101	0.093	0.114
	PNR	0.101	0.094	0.116
	PSS2	0.127	0.199	0.449
	PSS4	0.123	0.158	0.258
	SS	0.127	0.173	5.199
$\sigma = 7$	PPSS	0.231	0.204	0.269
	PNR	0.232	0.207	0.308
	PSS2	0.290	0.435	0.960
	PSS4	0.260	0.299	0.449
	SS	0.265	0.288	4.834
$\sigma = 10$	PPSS	0.325	0.273	0.338
	PNR	0.326	0.305	0.615
	PSS2	0.424	0.632	1.372
	PSS4	0.375	0.426	0.623
	SS	0.374	0.379	5.702

The PPSS method has smaller pointwise RMSEs over almost the entire interval than the PSS4 method, although the PSS4 method has slightly smaller RMSEs in [13,16] where the growth curve has some sharp changes.

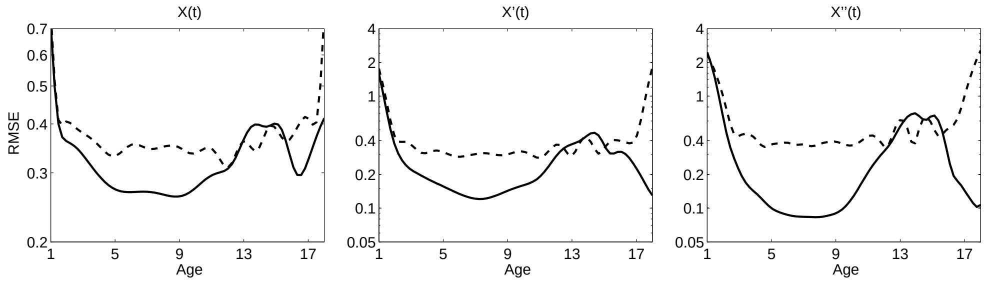

Figure 2: The pointwise root mean squared errors (RMSE) for the estimates of the curve, $x'(t)$, the first derivative, $x'(t)$, and the second derivative, $x''(t)$, using the parametric penalized spline smoothing (PPSS) method (solid lines) and the penalized spline smoothing (PSS4) method (dashed lines) when the PB function is the true function used to simulate data.

When the JPA function is the true function used to simulate data, the PNR method will be favoured since it has the correct parametric model for the data. Table 2 shows the average point-wise RMSEs for the estimated $\hat{x}(t)$, $\hat{x}'(t)$, $\hat{x}''(t)$ using the four methods in this scenario. When the noise standard deviation is set to 3, 7, or 10, the PPSS method has the same $\texttt{RMSE}(\hat{x}(t))$ with the PNR method, but $\texttt{RMSE}(\hat{x}'(t))$ and $\texttt{RMSE}(\hat{x}''(t))$ with the PPSS method is larger than the PNR method. The PPSS method always has smaller values for $\texttt{RMSE}(\hat{x}(t))$, $\texttt{RMSE}(\hat{x}'(t))$, and $\texttt{RMSE}(\hat{x}''(t))$ than the PSS method. The smoothing spline method has comparable RMSEs for $\hat{x}(t)$ and $\hat{x}'(t)$ with the PSS2 method, but the RMSE for $\hat{x}''(t)$ is very large, which may be caused by the error from the finite-difference method.

Figure 3 shows the pointwise RMSEs for the estimates of the curve, $x'(t)$, the first derivative, $x'(t)$, and the second derivative, $x''(t)$, using the PPSS and PSS4 methods when the JPA function is the true function used to simulate data. The PPSS method has smaller pointwise RMSEs over most of the

Table 2: The average point-wise root mean squared errors (RMSEs) for the estimates of the height function and the first and second derivatives using three methods, when the data are simulated based on the JPA function as the true function. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4), and Wahba's standard smoothing spline method (SS). The JPA function is defined in (3).

Noise	**Method**	$\texttt{RMSE}(\hat{x}(t))$	$\texttt{RMSE}(\hat{x}'(t))$	$\texttt{RMSE}(\hat{x}''(t))$
$\sigma = 3$	PPSS	0.093	0.087	0.192
	PNR	0.093	0.080	0.120
	PSS2	0.154	0.332	0.958
	PSS4	0.161	0.284	0.687
	SS	0.156	0.353	23.400
$\sigma = 7$	PPSS	0.220	0.186	0.318
	PNR	0.220	0.181	0.265
	PSS2	0.305	0.541	1.425
	PSS4	0.283	0.391	0.804
	SS	0.309	0.526	14.917
$\sigma = 10$	PPSS	0.315	0.268	0.425
	PNR	0.315	0.265	0.384
	PSS2	0.430	0.722	1.824
	PSS4	0.388	0.499	0.941
	SS	0.422	0.648	15.633

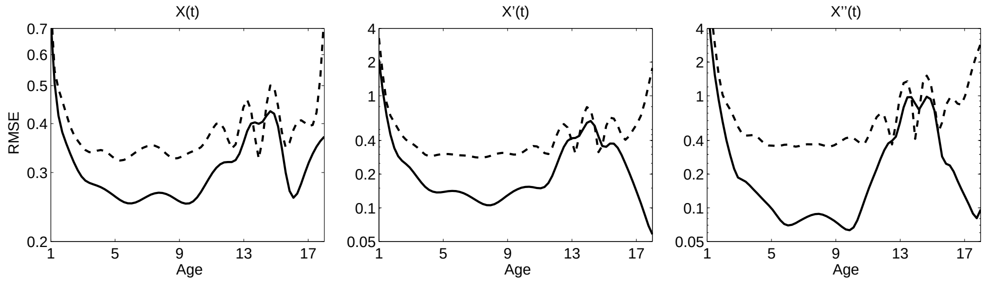

Figure 3: The pointwise root mean squared errors (RMSE) for the estimates of the curve, $x'(t)$, the first derivative, $x'(t)$, and the second derivative, $x''(t)$, using the parametric penalized spline smoothing (PPSS) method (solid lines) and the penalized spline smoothing (PSS4) method (dashed lines) when the JPA function is the true function used to simulate data.

interval than the PSS4 method.

3.2. Simulation 2

We choose two parametric functions:

$$h(t|\boldsymbol{\theta}_1) = a + b \cdot t + c \cdot t^2, \tag{5}$$

$$f(t|\boldsymbol{\theta}_2) = \frac{\exp(\alpha t)}{\cos(\beta t)}, \tag{6}$$

where $\boldsymbol{\theta}_1 = (a, b, c)^T$ is a vector of parameters in the parametric function $h(t|\boldsymbol{\theta}_1)$, and $\boldsymbol{\theta}_2 = (\alpha, \beta)^T$ is a vector of parameters in the parametric function $f(t|\boldsymbol{\theta}_2)$. The function $f(t|\boldsymbol{\theta}_2)$ is treated as the true parametric function with the true parameter value $\boldsymbol{\theta}_2 = (2, 3)^T$. The true function is evaluated at 21 equally-spaced points in [-0.1,0.1]. The white noises with the variance, σ^2, are added to the true values to generate the simulated data. The value of σ is chosen as 0.005, 0.01, and 0.05 in three separate simulation studies. Each simulation study is implemented with 500 replicates. The above two parametric functions are chosen because they have almost the same values, but their first and second derivatives are different, as shown in Figure 4.

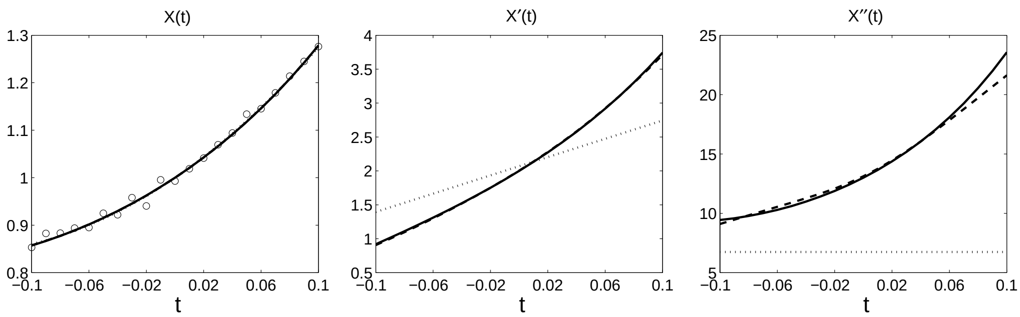

Figure 4: The estimates for the curve, $x(t)$, the first derivative, $x'(t)$, and the second derivative, $x''(t)$, using the parametric penalized spline smoothing method (PPSS). The solid lines are the true curve, $x(t) = \exp(2t)/\cos(3t)$, and the true derivatives. The dash lines are the estimates using PPSS model. The dotted lines are the estimate using parametric nonlinear regression (PNR) method. The circles are the simulation data, which have white noises with the variance $\sigma^2 = 0.01^2$. The PPSS and PNR estimates for $x(t)$ are too close to the true curve $x(t)$ to be distinguishable; The PPSS estimates for $x'(t)$ is too close to the true derivative $x'(t)$ to be distinguishable.

The curve and the first and second derivatives are estimated with the three methods. The first method is the parametric nonlinear regression (PNR) method which uses $h(t|\boldsymbol{\theta}_1)$ as the parametric function. The second method is the penalized spline smoothing (PSS) method, in which order 6 B-spline basis functions are chosen with one knot put in each measurement location. The roughness penalty term is defined with the second or fourth derivatives of the nonparametric function, which is called PSS2 and PSS4, respectively. The third method is the parametric penalized spline smoothing (PPSS) method. It uses the same basis system as the penalized spline smoothing method, and the roughness penalty term is defined by the fourth derivative of the nonparametric function. The parametric function $h(t|\boldsymbol{\theta})$ in (2) is chosen as the polynomial function $h(t|\boldsymbol{\theta}_1)$.

The summary of the estimates from the data simulated with the large scale of noises ($\sigma = 0.05$) is displayed in Table 3. The PPSS method has the smallest `RMSE` of the estimates for the curve and the first and second derivatives. The PNR method suffers with the large bias of the estimates, while standard deviations of the estimates using the PSS2 and PSS4 methods are too large.

When the data are simulated with the median scale of noises ($\sigma = 0.01$), the summary for the estimates are displayed in Table 4. The estimates for the first and second derivatives with the PPSS method have the smallest `RMSE`. The PNR method has slightly smaller `RMSE` of the curve estimates than the PPSS method, but the bias of curve estimates with the PNR method is tenfold of that obtained using the PPSS method. The PNR method also obtains very large biased estimates of the first and second derivatives.

When the data are simulated with the small scale of noises ($\sigma = 0.005$), the summary for the estimates are displayed in Table 5. The PPSS method is slightly better than PSS4 method, while the PNR and PSS2 method perform the worst overall. The PNR method obtains very large biased estimates, because wrong parametric form is assumed. The estimates with the PSS2 method have the extremely large standard deviations.

4. Applications

Figure 5 displays some real height measurements of one girl in a Berkeley growth study (Tuddenham and Snyder, 1954). The girl is measured at 31 non-equally-spaced time points, with four measurements when she is one

Table 3: The average point-wise bias(Bias), standard deviations (SDs), and root mean squared errors (RMSEs) for the estimates of the curve and the first and second derivatives using three methods, when the data are simulated based on $f(t) = \exp(2t)/\cos(3t)$ as the true function with the noise standard deviation 0.05. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4).

Estimator	**Method**	$\lvert\texttt{Bias}\rvert * 10^3$	$\texttt{SD} * 10^3$	$\texttt{RMSE} * 10^3$
$\hat{x}(t)$	PPSS	1.85	17.75	17.88
	PNR	1.86	17.76	17.89
	PSS2	1.41	21.30	21.36
	PSS4	0.81	20.47	20.50

Estimator	**Method**	$\lvert\texttt{Bias}\rvert * 10^2$	$\texttt{SD} * 10^2$	$\texttt{RMSE} * 10^2$
$\hat{x}'(t)$	PPSS	8.92	40.08	41.47
	PNR	37.54	25.87	46.80
	PSS2	8.25	77.69	78.48
	PSS4	3.25	67.17	67.25

Estimator	**Method**	$\lvert\texttt{Bias}\rvert$	$\texttt{SD}$	$\texttt{RMSE}$
$\hat{x}''(t)$	PPSS	3.45	6.50	7.68
	PNR	7.70	3.25	8.52
	PSS2	5.25	32.91	34.79
	PSS4	1.14	20.65	20.69

Table 4: The average point-wise bias(Bias), standard deviations (SDs), and root mean squared errors (RMSEs) for the estimates of the curve and the first and second derivatives using three methods, when the data are simulated based on $f(t) = \exp(2t)/\cos(3t)$ as the true function with noise $\sigma = 0.01$. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4).

Estimator	**Method**	$\lvert\texttt{Bias}\rvert * 10^3$	$\texttt{SD} * 10^3$	$\texttt{RMSE} * 10^3$
$\hat{x}(t)$	PPSS	0.16	4.48	4.49
	PNR	1.60	3.70	4.11
	PSS2	0.24	5.57	5.57
	PSS4	0.15	4.56	4.56

Estimator	**Method**	$\lvert\texttt{Bias}\rvert * 10^2$	$\texttt{SD} * 10^2$	$\texttt{RMSE} * 10^2$
$\hat{x}'(t)$	PPSS	0.80	17.30	17.33
	PNR	36.35	5.26	36.96
	PSS2	2.95	34.24	34.53
	PSS4	0.51	18.05	18.06

Estimator	**Method**	$\lvert\texttt{Bias}\rvert$	$\texttt{SD}$	$\texttt{RMSE}$
$\hat{x}''(t)$	PPSS	0.35	7.26	7.28
	PNR	7.51	0.66	7.55
	PSS2	3.47	28.45	30.01
	PSS4	0.29	7.61	7.62

Table 5: The average point-wise absolute value of bias(Bias), standard deviations (SDs), and root mean squared errors (RMSEs) for the estimates of the curve and the first and second derivatives using three methods, when the data are simulated based on $f(t) = \exp(2t)/\cos(3t)$ as the true function with noise $\sigma = 0.005$. The methods include the parametric penalized spline smoothing (PPSS), the parametric nonlinear regression (PNR), and the penalized spline smoothing with the roughness penalty term defined with the second derivative (PSS2) or the fourth derivative (PSS4).

Estimator	**Method**	$\lvert\texttt{Bias}\rvert * 10^3$	$\texttt{SD} * 10^3$	$\texttt{RMSE} * 10^3$
$\hat{x}(t)$	PPSS	0.09	2.21	2.21
	PNR	1.60	1.79	2.51
	PSS2	0.13	3.53	3.53
	PSS4	0.09	2.21	2.22

Estimator	**Method**	$\lvert\texttt{Bias}\rvert * 10^2$	$\texttt{SD} * 10^2$	$\texttt{RMSE} * 10^2$
$\hat{x}'(t)$	PPSS	0.76	8.70	8.75
	PNR	36.73	2.61	36.91
	PSS2	1.55	36.05	36.12
	PSS4	0.77	8.73	8.78

Estimator	**Method**	$\lvert\texttt{Bias}\rvert$	$\texttt{SD}$	$\texttt{RMSE}$
$\hat{x}''(t)$	PPSS	0.46	3.65	3.69
	PNR	7.58	0.33	7.59
	PSS2	3.70	60.08	61.26
	PSS4	0.46	3.66	3.70

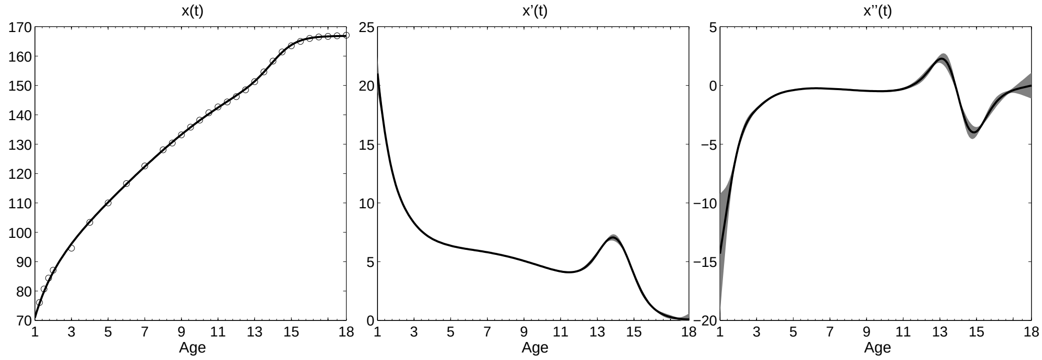

Figure 5: The estimate for the height function $x(t)$, the first derivative $x'(t)$, and the second derivative $x''(t)$ with the parametric penalized spline smoothing method. The circles are the real height measurements of one girl in one Berkeley growth study. The shaded areas are the corresponding 95% point-wise confidence intervals.

year old, annual measurements from two to eight years and biannual measurements from eight to eighteen years. The growth dynamic process of this girl can be studied by estimating the height function $x(t)$, and the first and second derivatives $x'(t)$, $x''(t)$ using the parametric penalized spline smoothing method. The JPA function (3) is used as the parametric function $h(t)$ defined in (2). Ten-fold cross-validation is employed to choose the optimal values for the smoothing parameters. It is displayed in Figure 6, which is minimized when $\log_{10}(\lambda_1) = -0.9$ and $\log_{10}(\lambda_2) = -3.9$ using the grid search method. The parametric penalized spline smoothing method estimates the parameters in the JPA function as $A = 166.90, B_1 = 0.56, B_2 = 0.13, B_3 = 0.08, C_1 = 0.53, C_2 = 3.42, C_3 = 24.00, E = 0.0001$.

The estimates for the height function, $x(t)$, and the first and second derivatives, $x'(t)$ and $x''(t)$, are displayed in Figure 5. The height function is strictly increasing, and becomes flat around 16. The grow rate function, $x'(t)$, decreases sharply until 3, and decreases slowly until 13. It then bumps up at 13, peaks at 14, and decreases quickly to zero at 18. The acceleration function, $x''(t)$, quickly increases until 3, becomes flat from 3 to 12, and then has a peak around 13 and a valley around 15.

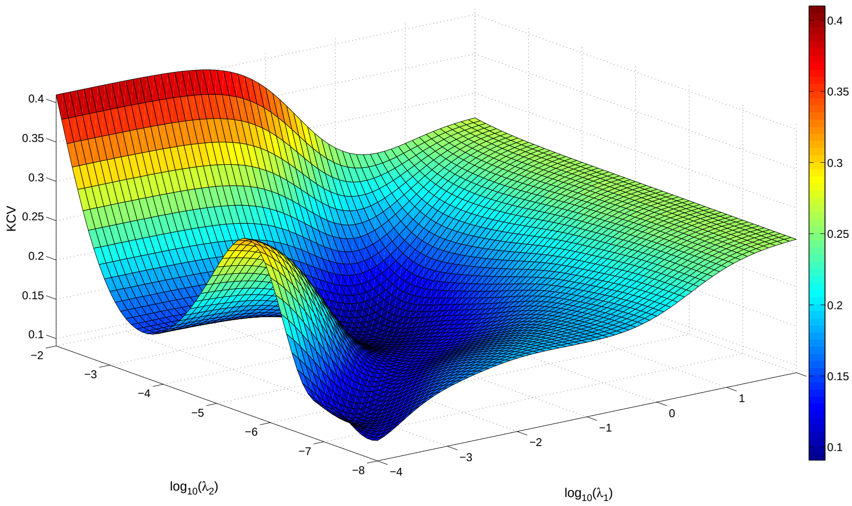

Figure 6: K-fold cross-validation (KCV) calculated based on (**??**) by setting $K = 10$. It is minimized when $\log_{10}(\lambda_1) = -0.9$ and $\log_{10}(\lambda_2) = -3.9$ using the grid search method.

The 95% point-wise confidence intervals for the height function, $x(t)$, and the first and second derivatives, $x'(t)$ and $x''(t)$, are obtained with the parametric bootstrap method (Efron and Tibshirani, 1993). The parametric bootstrap method is implemented as follows. The simulated data are generated by adding white noise with the estimated variance $\hat{\sigma}^2$ to the estimated height function, $\hat{x}(t)$. The height function, $x(t)$, and the first and second derivatives, $x'(t)$ and $x''(t)$, are then estimated from the simulated data with the parametric penalized spline smoothing method. The above process are done using 500 replications of the simulation. The 95% point-wise confidence intervals are obtained from the 500 estimates of the height function and the first and second derivatives.

5. Conclusions

It is of great interest to estimate a smooth function and its derivatives accurately. Some parametric models are often proposed based on some expert opinion or prior knowledge of the function, but the parametric assumption may not be accurate. The nonparametric smoothing methods are very flexible, but do not consider any expert (prior) knowledge on the function.

The parametric penalized spline smoothing method combines the advantages of the parametric models and the nonparametric smoothing method. It uses a linear combination of basis functions to estimate the underlying function nonparametrically. It uses the expert knowledge of the function by defining a penalty term as the distance of the fitted function to the parametric function. It uses a saturated number of basis functions, and a roughness penalty term is defined using the second derivative of the fitted function to prevent overfitting. As suggested by one reviewer, the parametric penalized spline method may be considered as an extension of the partial spline models introduced in Wahba (1990).

The introduction of the parametric model $h(t|\theta)$ can be thought of as a "nuisance parameter (model)", and the goal of the simulations would be to test whether this nuisance model is significant or not. The simulation studies show that the parametric penalized spline smoothing method can obtain more accurate estimates of the function and its derivatives than the parametric regression method and penalized spline smoothing method when the parametric model is not correct. On the other hand, even when the parametric model is correct, the parametric penalized spline smoothing method is still comparable to the parametric regression method.

Acknowledgements

The authors would like to thank Prof. Gilles Celeux, the associate editor, and two reviewers for their very constructive suggestions and comments, which are very helpful for improving this paper. The authors would also like to thank Ms. Oksana Chkrebtii and Ms. Alisha Albert-Green for carefully proofreading this manuscript. This research was supported by a discovery grant from Natural Sciences and Engineering Research Council of Canada (NSERC) to J. Cao.

References

Bates, D. M., Watts, D. B., 1988. Nonlinear Regression Analysis and Its Applications. Wiley, New York.

Bock, R., Thissen, D., 1980. Statistical problems of fitting individual growth curves. In F.E. Johnston, A.F. Roche and C. Susanne (eds.) *Human Physical Groth and Maturation: Methodologies and Factors.* Plenum, New York.

Cao, J., Ramsay, J. O., 2010. Linear mixed effects modeling by parameter cascading. Journal of the American Statistical Association 105, 365–374.

de Boor, C., 2001. A Practical Guide to Splines. Springer, New York.

Efron, B., Tibshirani, R. J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

Eubank, R. L., 1988. Spline smoothing and nonparametric regression. Marcel Dekker, New York and Basil.

Friedman, J. H., 1991. Multivariate adaptive regression splines (with discussion). The Annals of Statistics 19, 1–141.

Friedman, J. H., Silverman, B. W., 1989. Flexible parsimonious smoothing and additive modeling (with discussion). Technometrics 31, 3–39.

Gasser, T., Muller, H., Kohler, W., Largo, R., Molinari, L., Prader, A., 1985. Human height growth: Correlational and multivariate structure of velocity and acceleration. Annals of Human Biology 12, 501–515.

Gasser, T., Muller, H., Kohler, W., Molinari, L., Prader, A., 1984. Nonparametric regression analysis of growth curves. The Annals of Statistics 12, 210–229.

Jolicoeur, P., Pontier, J., Abidi, H., 1992. Asymptotic models for the longitudinal growth of human stature. American Journal of Human Biology 4, 461–468.

Jolicoeur, P., Pontier, J., Pernin, M., Sempe, M., 1988. A lifetime asymptotic growth curve for human height. Biometrics 44, 995–1003.

Kanefuji, K., Shohoji, T., 1990. On a growth model of human height. Growth Development and Aging 54, 155–165.

Preece, M. A., Baines, M. J., 1978. A new family of mathematical models describing the human growth curve. Annals of Human Biology 5, 1–24.

Ramsay, J. O., Altman, N., Bock, R. D., 1994. Variation in height acceleration in the Fels growth data. The Canadian Journal of Statistics 22, 89–102.

Ramsay, J. O., Hooker, G., Campbell, D., Cao, J., 2007. Parameter estimation for differential equations: a generalized smoothing approach (with discussion). Journal of the Royal Statistical Society, Series B 69, 741–796.

Ramsay, J. O., Silverman, B. W., 2005. Functional Data Analysis, 2nd Edition. Springer, New York.

Stone, C. J., Hansen, M., Kooperberg, C., Truong, Y. K., 1997. Polynomial splines and their tensor products in extended linear modeling (with discussion). The Annals of Statistics 25, 1371–1470.

Tuddenham, R., Snyder, M., 1954. Physical growth of california boys and girls from birth to eighteen years. University of California Publications in Child Development 1, 183–364.

Wahba, G., 1990. Spline models for observational data. Society for Industrial and Applied Mathematics, Philadelphia.