# $\mathcal{C}$ri$\mathcal{SP}$ - a Tool for Bump Hunting

Jaroslaw Harezlak

*Indiana University School of Medicine, Indianapolis, IN, USA*

Nancy E. Heckman

*University of British Columbia, Vancouver, BC, Canada*

October 8, 1999

University of British Columbia

Department of Statistics

Technical Report 187

## Abstract

We propose a test of multimodality of regression functions and their derivatives. The test statistic is a critical smoothing parameter ($\mathcal{C}$ri$\mathcal{SP}$), giving the minimum amount of smoothing necessary to force the regression function to satisfy the null hypothesis. The $p$-values are computed via bootstrapping. Our idea is motivated by Silverman's test concerning the number of modes in the density function. Simulation studies indicate that the test works well, even when testing for bumps in the derivative. We apply $\mathcal{C}$ri$\mathcal{SP}$ to children's growth data, to study the number of spurts of growth.

**Key Words:** Bootstrap; Critical smoothing parameter; Growth data; Multimodality testing; Smoothing

Jarek Harezlak is a Biostatistician in the Division of Biostatistics, Indiana University School of Medicine, Indianapolis, IN, 46202-2872 USA (e-mail jharezla@iupui.edu). Nancy Heckman is a Professor in the Department of Statistics, University of British Columbia, Vancouver BC V6T 1Z3 Canada (e-mail nancy@stat.ubc.ca).

# 1 Introduction

Statisticians often want to find an underlying true regression function when given a set of noisy data. We assume, very often, that this function is of a parametric form. Unfortunately once we adopt a specific parametric form of a function, we cannot detect features of the data beyond those found in the specified model. For instance, if we adopt the exponential parametric model, the regression function is restricted to be monotone. However, we might want to allow more flexibility in modeling. We can achieve that by not assuming any specific model, i.e. by taking advantage of nonparametric regression. This approach permits a detection of fine features in the function, e.g. bumps and valleys.

In Section 6 we consider growth data collected on children in Berkeley, California. This data set consists of height measurements taken at least once per year from age one till age eighteen. Our goal is to study the number of growth spurts, that is the number of local maxima in the rate of growth. We allow ourselves flexibility in estimating the derivative of the growth curve by using smoothing techniques. We use a parameter controlling smoothness of the estimate as our test statistic. Our testing method is inspired by Silverman's testing for modes in density functions (1981, 1983). He used the kernel density estimation method, a local approach producing a smooth histogram, with the smoothing parameter (also called a bandwidth), $h$, playing the role of bin-width. If the data are strongly bimodal, we would need a large value of $h$ to obtain a unimodal estimate. Silverman used this idea to test the null hypothesis that the density $f$ has $k$ modes versus the alternative that $f$ has more than $k$ modes.

Early parametric tests for multimodality were proposed by Cox (1966), Engelman and Hartigan (1969), and Wolfe (1971). In the 1980's nonparametric methods in bump-hunting started gaining popularity due to greater availability of computing resources. Besides Silverman's method, Good and Gaskins (1980), Hartigan and Hartigan (1985), Hartigan and Mohanty (1992), Müller and Sawitzki (1991), Minnotte and Scott (1993) and Minnotte (1997) proposed and/or extended tests for multimodality of a density.

Tests for bumps are much more prevalent in the density estimation context than in the regression setting. Schlee (1982) and Heckman (1992) discussed the issues involved with bump hunting in regression but without giving a practical solution. Chaudhuri and Marron (1999) extend their density estimation work to finding bumps in regression functions. Tests for monotonicity of the regression function have been proposed by Bowman, Jones and Gijbels (1998), Hall and Heckman (1998) and Gijbels, Hall, Jones and Koch (1999). Bowman et al's test is an extension of Silverman's idea. We extend Bowman's monotonicity test to $\mathcal{C}\text{ri}\mathcal{SP}$ (Critical Smoothing Parameter), a test of

$$H_0 : \#\text{ of bumps in } m^{(j)}(\cdot) \ \leq k \quad \text{vs.} \quad H_1 : \#\text{ of bumps in } m^{(j)}(\cdot) \ > k \tag{1}$$

where $m^{(j)}(\cdot)$ is the $j^{th}$ derivative of the true regression function ($j = 0, 1, 2, \ldots$). Our testing method is based on the critical value of a smoothing parameter which is defined as the smallest smoothing parameter yielding a regression curve estimate with $k$ bumps. The $p$-value of the test is calculated via bootstrap. Our simulation studies indicate that the method of smoothing isn't crucial.

In Section 2, we give a short outline of the regression estimation methods. Section 3 deals with the definition of a bump and estimation of the number of bumps. Our test is introduced in Section 4, simulation studies are described in Section 5, and the test is applied to the growth data in Section 6.

## 2 Methods of Estimation

Suppose that we observe independent data $(t_i, Y_i)$, $i = 1, \cdots, n$, with $E(Y|t) = m(t)$. A smoothing spline estimate, $\hat{m}_\lambda(\cdot)$, is the minimizer of

$$\sum_{i=1}^{n}(Y_i - m(t_i))^2 + \lambda \int_0^1 [Lm(t)]^2 dt, \tag{2}$$

over an appropriate space of functions. Here $\lambda \geq 0$ is called a smoothing parameter, and $L$ is a linear differential operator of order $k \geq 1$ defined as

$$(Lm)(t) = m^{(k)}(t) + \sum_{j=0}^{k-1} w_j m^{(j)}(t) \quad \text{where } \ w_j \in \mathbb{R} \text{ for } j = 0, \ldots, k-1.$$

We can see that the first part of (2) penalizes for the lack of fit, and the second part penalizes for $m$ when $Lm$ is far from 0. We say that $m_{def}$ is a default model if and only if $Lm_{def}(t) = 0$ almost everywhere $t$. When $Lm = m''$, the resulting estimate of $m$ is the popular cubic smoothing spline. See Heckman and Ramsay (in press) for a discussion of minimizers of (2). These minimizers can also be found via the model-based Kalman filter approach (Ansley, Kohn and Wong 1993; de Jong 1997).

In the growth data analysis (described in Section 6), the default model is commonly assumed to be of the form $m_{def}(t) = \alpha_0 + \alpha_1 \exp(-\gamma t)$ which has corresponding $L$ $Lm = m'' + \gamma m$. The minimizer of (2) with this $L$ is easily calculated using J.O. Ramsay's S-PLUS module Lspline, available on the web at http://lib.stat.cmu.edu. In order to use this method, we need to estimate the parameter $\gamma$ of the default model. We find $\hat{\gamma}$ for each dataset by a nonlinear least squares fit of the parametric model $m(t) = \alpha_0 + \alpha_1 \exp\{-\gamma t\}$. See Heckman and Ramsay (in press) for a discussion of estimation of unknown parameters in the default model.

For comparison purposes we also estimate $m$ via local polynomial smoothing, a form of kernel smoothing. See, for instance, Wand and Jones (1995). We use M.P. Wand's S-PLUS module locpoly, available at http://lib.stat.cmu.edu. Both Lspline and locpoly also produce estimates of the derivatives of $m$.

In one of the steps of our bootstrap procedure, described in Section 4, we obtain an unrestricted estimate of the regression function. The estimate uses a data-dependent smoothing parameter chosen to obtain a good fit. We choose this smoothing parameter by generalized cross validation ($GCV$) for splines and by a plug-in procedure for local polynomial regression estimators. Both procedures have many favorable properties (see Green and Silverman 1994) and are readily available in the above-mentioned S-PLUS functions.

# 3 Counting Bumps

Following the idea of Heckman (1992), we can decide that a bump in $m^{(k)}(\cdot)$ has occurred if the estimate of $m^{(k)}(\cdot)$ rises and then falls over a range of the independent variable. If we let $\hat{Y}_i^{(k)} = \hat{m}^{(k)}(t_i)$, $i = 1, ..., n$, be our estimates, then we say $\hat{m}^{(k)}(\cdot)$ has one $l$-bump in $(t_{j-l}, t_{j+l})$, if

$$\left\{\hat{Y}_{j-l}^{(k)} < \hat{Y}_{j-l+1}^{(k)} < \cdots < \hat{Y}_j^{(k)} > \hat{Y}_{j+1}^{(k)} > \cdots > \hat{Y}_{j+l}^{(k)}\right\}. \tag{3}$$

Unfortunately the restriction of the estimates to be strictly rising and then falling over a range of an independent variable does not seem to work well in practice. This is discussed in Sections 4.1 and 5.1. Therefore in our testing procedure, we utilize

$$\hat{Y}_j^{(k)} > \max\left\{\hat{Y}_{j-l}^{(k)}, \cdots, \hat{Y}_{j-1}^{(k)}, \hat{Y}_{j+1}^{(k)}, \cdots, \hat{Y}_{j+l}^{(k)}\right\} \tag{4}$$

as a basis for saying that $\hat{m}^{(k)}(\cdot)$ has one bump in $(t_{j-l}, t_{j+l})$. We call this an $l$-max. Note that when $l = 1$, (3) and (4) are the same.

One can use $t_i$'s equal to the data points or one can use $t_i$'s on a grid. Thus, the important components in defining a bump are the $t_i$'s used and the value of $l$. For instance, the code used in Bowman et al (1998) takes $l = 1$ and, as default, a grid of 50 $t_i$'s evenly spaced between the minimum and maximum data value. The impact of choice of $l$ and the grid points does not seem to be widely discussed in the bump-hunting literature. We give some guidance on these choices in our simulation studies and data analysis.

# 4 The $\mathcal{C}$ri$\mathcal{SP}$ Test

## 4.1 $\mathcal{C}$ri$\mathcal{SP}$ in Regression

We test (1) with $j = 0$ using the test statistic $\lambda_{crit} = \inf\{\lambda : \hat{m}_\lambda \text{ has } k \text{ bumps}\}$. When $k = 0$ our test is essentially that of Bowman et al (1998) except that we take the bump definition (4). We reject $H_0$ if $\lambda_{crit}$ is too big. The $p$-value of the test is then

$$p = P_{H_0}\{\lambda_{crit} > \lambda_{crit}^{obs}\}$$

where $\lambda_{crit}^{obs}$ is the observed value of $\lambda_{crit}$ and $P_{H_0}$ denotes the supremum (over $m$ in $H_0$) of the probability of the event. If the number of bumps in $\hat{m}_\lambda$ is monotone in $\lambda$, that is, the greater the $\lambda$ the fewer bumps of $\hat{m}_\lambda(\cdot)$, then $p$ is equal to the more simply defined $p^\star$:

$$p^\star = P_{H_0}\{\hat{m}_{\lambda_{crit}^{obs}} \text{ has more than } k \text{ bumps}\}. \tag{5}$$

The monotonicity of the number of modes as a function of the smoothing parameter or the bandwidth is a desirable property. It implies that $p = p^*$. But furthermore, the testing rule is not really sensible if the monotonicity property does not hold. The monotonicity property was established in a few cases. Silverman (1981) proved it for a kernel density estimator based on a normal kernel. Härdle (1989) established the monotonicity property for Gasser-Müller and Nadaraya-Watson estimators when the kernel is normal. However, this feature does not hold for the local linear regression estimator or for the spline regression estimator. Fortunately, simulation studies indicate that, with a bump defined as in (4), data sets where the number of bumps in $\hat{m}_\lambda$ is non-monotone are very unusual. Unfortunately, this is not true for the bump definition (3). Therefore, we use definition (4) and then use $p^\star$ instead of $p$, estimating $p^\star$ via bootstrapping.

The probability in the right hand side of (5) should be calculated as a supremum over the null hypothesis. However, we estimate it using a regression function in the null hypothesis which is close to the data.

The bootstrap procedure using a generic smoothing method SM involves the following steps:

1. Calculate $\lambda_{crit}^{obs}$ and $\hat{m}_{\lambda_{crit}^{obs}}$ using SM.
2. Find $\hat{m}_{unr}$, an unrestricted estimate of $m(\cdot)$ with smoothing parameter for SM chosen by $GCV$ (splines) or the plug-in method (kernel).
3. Find residuals $\varepsilon_i^* = Y_i - \hat{m}_{unr}(t_i)$, $i = 1, ..., n$.
4. Create a bootstrap sample $Y_i^* = \hat{m}_{\lambda_{crit}^{obs}}(t_i) + \varepsilon_i^*$, $i = 1, ..., n$ where the $\varepsilon_i^*$'s are drawn at random with replacement from the $\varepsilon_j^*$'s, $j = 1, 2, \ldots, n$.
5. Find the new estimate $\hat{m}^*_{\lambda_{crit}^{obs}}$ for the bootstrap sample $\{Y_i^*\}_1^n$ using SM.
6. Find the number of bumps in $\hat{m}^*_{\lambda_{crit}^{obs}}$.
7. Repeat 4-6 a large number of times, and calculate the estimate of the $p$-value by counting the number of estimates $\hat{m}^*_{\lambda_{crit}^{obs}}$ having more than $k$ bumps, i.e.

$$\hat{p} = \frac{\#\text{ of estimates } \hat{m}^*_{\lambda_{crit}^{obs}} \text{ with more than } k \text{ bumps}}{\#\text{ of bootstrap samples}}.$$

The smoothing method in step 2 need not be the same as the smoothing method in steps 1 and 5. However, in keeping with the spirit of bootstrapping, steps 1 and 5 should use the same smoothing method. If we use local polynomial smoothing in steps 1 and 5, the definition of the critical smoothing parameter is analogous to $\lambda_{crit}$. We define $h_{crit} = \inf\{h : \hat{m}_h \text{ has } k \text{ bumps}\}$ and we use the same bootstrap steps 4-7 with $\hat{m}_{\lambda_{crit}^{obs}}$ replaced by $\hat{m}_{h_{crit}^{obs}}$.

Since we want to calculate a probability under $H_0$, in step 3 we could calculate the residuals using the restricted estimates $\hat{Y}_i^{crit} = \hat{m}_{\lambda_{crit}^{obs}}(t_i)$. However, the results of our simulation study indicate that using the unrestricted estimates to obtain the residuals gives the right size of the test.

## 4.2 $\mathcal{C}$ri$\mathcal{SP}$ for Derivatives

We can extend our procedure to testing for bumps in the derivatives of $m$, i.e. testing (1) with $j \geq 1$. The bootstrap procedure is the same as that in Section 4.1 except that in step 1 we calculate $\lambda_{crit}^{obs}$ for $\widehat{m^{(j)}}$ and we use $\widehat{m^{(j)}}$ in steps 5, 6, and 7. In Section 5.2 we study the derivative test with $j = 1$.

# 5 Simulation Studies

## 5.1 Regression Function Simulations

We generated the data $Y_i = m(t_i) + \varepsilon_i$ with $t_i = i/101$ and $\varepsilon_i \sim N(0, (0.05)^2)$ for $i = 1, 2, \ldots, 101$. Functions used in the simulation study have none, one or two bumps and are described below. For each function $m$, we produced 500 data sets, and to each data set we applied the bootstrap procedure with bootstrap sample size equal to 500.

Our functions $m$ are of the form

$$m(t|a,b) = 1 + \exp(-4t) + a * B(t|0.25) + b * B(t|0.75)$$

where

$$B(t|c) = \exp\left\{-\frac{(t-c)^2}{2\,(0.1)^2}\right\}.$$

Depending on the values of $a$ and $b$, the function $B$ will produce a bump in $m$ at $t = 0.25$ and/or $t = 0.75$. When $a = 0$ the resulting function has no bump at $t$=0.25. When $a = 0.32$, $m$ has no bump at $t = 0.25$ but $m$ has a shoulder there. The values $a = 0.48$ and $a = 0.64$ yield a medium and large bump respectively. When $b = 0$, there is no bump in the regression function at $t$=0.75. However, values of $b = 0.10, 0.15, 0.20$ give rise to a small, medium and large bump respectively. Plots of $m$ for our different combinations of $a$ and $b$ are given in Figure 1. The detectibility of a bump depends not only on the multiplier of the function $B$, but also on the location of the bump. For instance, when $a = 0.48$, the bump might be rather difficult to detect, since the slope of the underlying function obscures part of the dip and rise in the bump (see Figure 1).

Insert Figure 1 about here.

As mentioned previously, we used bump definition (4). We came to that recommendation by simulation studies using $m(\cdot|a = 0.48, b = 0.15)$. We conducted the test using definitions of a bump as $l$-bump and $l$-max for $l = 2, 3, 4, 5$ as in (3) and (4). The monotonicity of the number of bumps as a function of the smoothing parameter $\lambda$ did not hold when the definition $l$-bump with $l \geq 3$ was used. Also, for all values of $l$ the calculations using the $l$-bump definition were very demanding computationally. With 4-max and 5-max we observed another problem, the non-existence of $\lambda_{crit}$ in some cases. We found that 3-max was an optimum definition of a bump for testing purposes.

We used three smoothing methods in steps 1 and 5 of the bootstrap procedure, the spline method with $Lm = m'' + \gamma m$ (denoted Lspline), local linear smoothing (denoted Kernel), and the spline method with $Lm = m''$. The results from the last method aren't presented here since they're similar to the results using the Lspline method. In step 2 we always used the Lspline procedure.

Tables 1, 2 and 3 show the proportions of the simulations with $p$-values falling below $5\%, 10\%$ and $25\%$. $\mathcal{C}\mathrm{ri}\mathcal{SP}$ performs well, although the test is usually conservative.

Table 2 ($k = 0$) and Table 3 ($k = 0, 1$) show that both smoothing methods have high power when testing at $\alpha = 0.05, 0.10$ and $0.25$. At $\alpha = 0.05$, the kernel method tends to have higher power than the the Lspline method. The curves in Figure 2, showing the proportion of times that the simulated test yields p-value less than or equal to $\alpha$, clearly show that the kernel method is more powerful for $\alpha$ less than around 0.15. Plots for other functions are similar to those in Figure 2.

From the tables, we see that the size is usually smaller than but not too far from the prespecified $\alpha$. Figure 2 shows that this is true for $\alpha$ less than around 0.20. Given these power and size properties, we recommend testing at levels of $\alpha$ between 0.10 and 0.20. If we conduct tests of (1) with prespecified significance levels in this range, we overwhelmingly reject the null hypothesis when we should and rarely reject it when we shouldn't. This is in keeping with comments of Minnotte and Scott (1993). Thus we can use the testing procedure to estimate the number of bumps in a curve by sequential hypothesis testing, that is by conducting tests of (1) for a sequence of $j$ values.

Insert Tables 1, 2, 3 and Figure 2 here.

$\mathcal{C}$ri$\mathcal{SP}$ appears to sometimes perceive a shoulder as a bump. For instance, in Table 1, the rejection rates for $a = 0.32$ are higher than for $a = 0$. We also see that $\mathcal{C}$ri$\mathcal{SP}$ has some trouble when $\alpha = 0.05$ and there is a medium bump at $t = 0.25$ (Table 3, $a = 0.48$ and $b = 0.10$ or $0.15$). A medium bump on a steep part of the curve can be easily smoothed out to yield a monotone estimate.

## 5.2 Derivative Simulations

We now study how $\mathcal{C}$ri$\mathcal{SP}$ performs for testing (1) with $j = 1$. The regression functions are antiderivatives of the functions introduced in the previous section,

$$\begin{aligned} m(t) &= \int_0^t 1 + \exp\{-4s\} + a * B(s|0.25) + b * B(s|0.75)\ ds \\ &= 0.25 - 0.25\exp\{-4t\} \\ &\quad +(0.1)\sqrt{(2\pi)}\,\{a * \Phi((t-0.25)/0.1) + b * \Phi((t-0.75)/0.1\} \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable. We generated the data $Y_i = m(t_i)+\varepsilon_i$ with $t_i = i/101$ and $\varepsilon_i \sim N(0,(0.005)^2)$ for $i = 1, 2, \ldots, 101$. For each model, we produced 500 data sets and for each data set we applied a bootstrap procedure with a sample size equal to 500.

We have chosen six combinations of parameters $a$ and $b$ defining $m$. We have chosen them in such a way as to have two curves where $m'$ has 0 bumps ($a = 0, b = 0$; $a = 0.32, b = 0$), two with one bump ($a = 0, b = 0.20$; $a = 0.32, b = 0.20$), and two with two bumps ($a = 0.64, b = 0.20$; $a = 0.48, b = 0.15$). Recall that a value of $a = 0.32$ produces a shoulder in $m'$ at $t = 0.25$. The sample functions are presented in Figure 3 together with their derivatives.

Insert Figure 3 about here.

Throughout the study, the Lspline method of estimation was used with the penalty $L = m'' + \gamma m'$. For comparison purposes we also used local quadratic smoothing in steps 1 and 5 of the bootstrap procedure. Since the results were comparable, we only present the results of the Lspline method.

Derivative estimation is much more difficult than function estimation. However, our testing procedure did very well. As in the simulations presented in the previous section, the tests are conservative but have good power. Sequential hypothesis tests can be carried out to estimate the number of bumps in the derivative. Again, we recommend testing at levels of $\alpha$ between 0.10 and 0.20. In the analysis of the growth data, presented in the next section, we use $\alpha = 0.15$.

As before, we see increased rejection rates when there is a shoulder at $t = 0.25$. The case with $a = 0.48$ and $b = 0.15$ continues to pose a problem for $\mathcal{C}\mathrm{ri}\mathcal{SP}$.

Insert Table 4 about here.

# 6 Analysis of the Growth Data

We consider growth data from a longitudinal study conducted in Berkeley, California. These data are usually referred to in the literature as the Berkeley growth data. The complete description of the study can be found in Tuddenham and Snyder (1954). Here, we analyze the data from the records of 43 white males and 50 white females born in 1928 or 1929 in Berkeley, California. The data set includes, in most cases, measurements of recumbent length at three months and then at three-month intervals to 21 months. Standing heights were measured annually from age two until eight and thereafter semiannually until age eighteen. The question of interest concerns the existence of growth spurts, that is of local maxima in the derivative of the growth curve.

Following other authors (Gasser, Müller, Köhler, Molinari and Prader 1984a; Gasser et al 1984b; Gasser et al 1985; Ramsay, Bock and Gasser 1994) we have used only the measurements at ages 1, 1.5, 2 and annually thereafter until 9.5 in females and 10 in males, followed by semiannual measurements until age 18, for a total of 27 ages for the males and 28 for the females.

The growth curves were analyzed for each child. We decided on the number of bumps in a child's curve by performing four tests of (1) with $j = 1$ for $k = 0$, 1, 2, and 3. As mentioned in Section 5.2, we chose the significance level of our test to be $\alpha = 0.15$. We used the $p$-values to determine the number of bumps, using either a strict or soft rule.

By the strict decision rule, we reach a conclusion that the speed of growth has exactly $b$ bumps if, for the hypothesis tests $H_0$ : # bumps $\leq k$ the $p$-values are less than or equal to 0.15 for all $k < b$ and the $p$-values are greater than 0.15 for all $b \leq k \leq 3$. When we reject $H_0$ for all $k \leq 3$ we conclude that the number of bumps $b$ is greater than three. In some cases, no classification can be made by the strict decision rule. For instance, we may reject the hypothesis $H_0$ : # bumps $\leq k$ for $k$=0 and 2, but we do not reject it for $k$=1 and 3 at the level $\alpha = 0.15$.

Therefore, we introduce the soft decision rule, which always gives us a classification. For $b$=0, 1, 2, 3, we say that the curve has $b$ bumps if, for the hypothesis tests $H_0$ : # bumps $\leq k$, for all $k < b$ the $p$-values are less than or equal to 0.15, and the $p$-value is greater than 0.15 for $k = b$. As with the strict decision rule, we conclude that a curve has more than 3 bumps if, for all $k \leq 3$, the $p$-values are less than or equal to 0.15.

We have used the bootstrap procedure for derivatives outlined in Section 4.2. In steps 1 and 5 of the bootstrap we employed the Lspline smoothing method with penalty $Lm = m'' + \gamma m'$. In step 2, we found the unrestricted estimate $\hat{m}_{unr}(\cdot)$ of the regression function $m(\cdot)$ via local quadratic kernel smoothing. Lspline was not used in step 2, because the level of noise in the growth data is very small and, in such cases, spline methods almost interpolate the data, giving unrealistically small residuals. For each child, we generated 500 bootstrap samples from the original residuals.

Our data set has infrequent height measurements recorded at unequally spaced times. Therefore, for testing purposes, we evaluated the estimate of the speed of growth at 101 equally spaced values of time between the ages one and eighteen. We used 3-max on this grid of points as the definition of a bump.

By the strict decision rule, we could classify 34 of the 43 males and 39 of the 50 females. By definition, we could classify everyone via the soft decision rule. Figure 4 contains examples of plots of curves of males classified by the strict definition as having one, two, and three bumps, respectively. For each male, we plot the estimates of the growth curves having $b$=1, 2, and 3 growth spurts, and we indicate the one we decided on by a solid curve. Table 5 contains all of our classification results, with percents calculated based on the number classified.

Insert Figure 4 and Table 5 about here.

Comparison of the female and male growth curves suggests that multi-bump curves are more prevalent for boys than for girls. Also, boys and girls are equally easy to classify by the strict definition, with only about 20% of each group being unclassified.

We compare our results with the work described in Ramsay et al (1994). Ramsay et al estimate the ages at which growth spurts occur, the dominant one (pubertal) and less dominant (mid-spurt). In some cases, they find that some children had more than one mid-spurt. Unfortunately we cannot use their presented analysis to provide classifications as in Table 5. But we can infer that they would classify 19 males (44.2%) as having one bump (only the pubertal spurt) and 24 males (55.8%) as having two or more bumps (the pubertal and at least one mid-spurt). They would classify one girl out of 50 as having no growth spurts, 40 girls as having one spurt and 9 girls as having two or more spurts of growth. Our soft rule agrees with Ramsay et al's results for the males, but not for the females.

# 7 Conclusions

In this article, a smoothing parameter based test $\mathcal{C}\mathrm{ri}\mathcal{SP}$ for the number of bumps in regression functions and their derivatives has been proposed. Bootstrap methods have been used to provide p-values. Simulation studies indicated that the $\mathcal{C}\mathrm{ri}\mathcal{SP}$ test was successful in all considered setups. The behaviour of the test was better when the bump was of a bigger size and not situated on a steep slope of a regression function. As mentioned in Section 5, we should test at levels between 10% and 20%. At low levels (e.g. $\leq 0.05$) the power of the proposed test drops substantially, especially for the more challenging regression functions and their derivatives.

$\mathcal{C}\mathrm{ri}\mathcal{SP}$ was applied to the Berkeley growth data to detect the number of spurts of children's growth. Most of the results obtained from the completely data-driven test agree with what one might decide by eye, by looking at an appropriately smoothed derivative estimate. When we compared our test with the results of Ramsay et al (1994), we arrived at very similar conclusions for the male data. For the female data, we detected spurts of growth on more occasions than did Ramsay et al.

It should be noted that the procedure introduced in this article is complicated, and many issues regarding it are still unresolved. Throughout the simulation studies, we obtained a few surprising results. For instance, in the simulations involving the first derivative of the regression function, for some models we got better results (not presented here) with a higher level of noise added to the true regression function than with a lower level of noise. Also, we were able to detect a small bump on a flat portion of the curve more often than a large bump on a steep slope.

In the introduced procedure, the residuals were calculated using the $GCV$ criterion for the smoothing parameter choice in spline smoothing and with a plug-in procedure for local polynomial smoothing. Both methods choose the optimal smoothing parameter for the estimation of the function $m$. However, when we test for the number of bumps in the derivative of $m$, we could probably improve the performance of $\mathcal{C}\mathrm{ri}\mathcal{SP}$ if we used a procedure tailored to choosing the best smoothing parameter for $m'$, rather than for $m$.

Another issue of interest is the definition of a bump. We defined it in terms of the values of the original function $m$. However, we could define it as a zero down-crossing of the derivative of $m$.

When we define a bump as (3), we encounter the unpleasant property of the number of bumps being a non-monotone function of a smoothing parameter. This causes some philosophical problems in the testing procedure. This property is not necessary for $\mathcal{C}\mathrm{ri}\mathcal{SP}$ to work, but without the monotonicity property we cannot provide the intuitive reasons behind our test.

A modification of $\mathcal{C}\mathrm{ri}\mathcal{SP}$ to deal with correlated or heteroscedastic errors is worth pursuing. This adjustment would most likely involve a new estimate of $\hat{m}_{unr}$ and a different method of bootstrapping the residuals. A version of $\mathcal{C}\mathrm{ri}\mathcal{SP}$ for higher dimensional data would also be useful.

**Acknowledgements:** The research of both authors was supported by the National Science and Engineering Research Council of Canada.

# References

Ansley, C. , Kohn R. and Wong, C. (1993), "Nonparametric spline regression with prior information," *Biometrika*, 80, 75–88.

Bowman, A.W., Jones, M.C. and Gijbels, I. (1998), "Testing Monotonicity of Regression," *Journal of Computational and Graphical Statistics*, 7, 489-500.

Chaudhuri, P. and Marron, J.S. (1999), "SiZer for Exploration of Structures in Curves," *Journal of the American Statistical Association*, 94, 807–203 .

Cox, D.R. (1966), "Notes on the Analysis of Mixed Frequency Distributions," *British Journal of Mathematical and Statistical Psychology,* 19, 39-47.

de Jong, P. (1997), "Fixed interval smoothing," manuscript.

Engelman L. and Hartigan, J.A. (1969), "Percentage Points of a Test for Clusters," *Journal of the American Statistical Association,* 64, 1647-1648.

Gasser, T., Müller, H.-G., Köhler, W., Molinari L. and Prader A. (1984a), "Nonparametric Regression Analysis of Growth Curves," *The Annals of Statistics,* 12, 210-229.

Gasser, T., Köhler, W., Müller, H.-G., Kneip, A., Largo, R., Molinari L. and Prader A. (1984b), "Velocity and Acceleration of Height Growth Using Kernel Estimation," *The Annals of Human Biology,* 11, 397-411.

Gasser, T., Müller, H.-G., Köhler, W., Prader A., Largo, R. and Molinari L. (1985), "An Analysis of the Mid-growth and Adolescent Spurts of Height Based on Acceleration," *The Annals of Human Biology,* 12, 129-148.

Gijbels, I., Hall, P., Jones, M.C. and Koch, I. (1999), "Tests for Monotonicity of a Regression Mean with Guaranteed level," manuscript.

Good, I.J. and Gaskins, R.A. (1980), "Density Estimation and Bump-Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data," *Journal of the American Statistical Association,* 75, 42-56.

Green P.J. and Silverman, B.W. (1994), *Nonparametric Regression and Generalized Linear Models - A Roughness Penalty Approach*, Chapman and Hall.

Hall, P. and Heckman, N.E. (1998), "Testing for Monotonicity of a Regression Mean without Selecting a Bandwidth," Department of Statistics Technical Report 178, University of British Columbia.

Hartigan, J.A. and Hartigan, P.M. (1985), "The Dip Test of Unimodality," *The Annals of Statistics*, 13, 70-84.

Hartigan, J.A. and Mohanty, S. (1992), "The RUNT Test for Multimodality," *Journal of Classification*, 9, 63-70.

Heckman, N.E. (1992), "Bump Hunting in Regression Analysis," *Statistics and Probability Letters*, 14, 141-152.

Heckman, N.E. and Ramsay, J.O. (in press), "Penalized Regression with Model-Based Penalties," *Canadian Journal of Statistics.*

Minnotte, M.C. (1997), "Nonparametric Testing of the Existence of Modes," *The Annals of Statistics*, 25, 1646-1660.

Minnotte, M.C. and Scott, D.W. (1993), "The Mode Tree: A Tool for Visualization of Nonparametric Density Features," *Journal of Computational and Graphical Statistics*, 2, 51-68.

Müller, D.W. and Sawitzki, G. (1991), "Excess Mass Estimates and Tests for Multimodality," *Journal of the American Statistical Association,* 86, 738-746.

Ramsay, J.O., Bock, R.D. and Gasser, T. (1994), "Comparison of Height Acceleration Curves in the Fels, Zurich, and Berkeley Growth Data," *Annals of Human Biology*, 22, 413-426.

Schlee, W. (1982), "Nonparametric Tests of the Monotony and Convexity of Regression," in *Nonparametric Statistical Inference, vol. II* (eds: B.V. Gnedenko, M.L. Puri & I. Vincze), Amsterdam: North-Holland, 823-836.

Silverman, B.W. (1981), "Using Kernel Density Estimates to investigate Multimodality," *Journal of Royal Statistical Society, Series B*, 43, 97-99.

– – – – – (1983), "Some Properties of a Test for Multimodality Based on Kernel Density Estimates," in London Mathematical Society Lecture Note Series # 79

*Probability, Statistics and Analysis* (eds: J.F.C. Kingman & G.E.H. Reuter), 248-259.

Tuddenham, R.D., and Snyder, M.M. (1954), "Physical Growth of California Boys and Girls form Birth to Eighteen Years," *University of California Publications in Child Development*, 1, 183-364.

Wand, M.P. and Jones, M.C. (1995), *Kernel Smoothing*, Chapman and Hall.

Wolfe, S.J. (1971), "On the Unimodality of $L$ Functions," *The Annals of Mathematical Statistics*, 42, 912-918.

Table 1: Zero Bumps: Proportion of times $H_0$ is rejected at level $\alpha$ based on 500 simulations of $Y_i = m(t_i|a, b) + \varepsilon_i$, $i = 1, \ldots, 101$, $\varepsilon_i \sim N(0, (0.05)^2)$.

| | | | $H_0$ : # of bumps $\leq k$ | | |
|---|---|---|---|---|---|
| **$a, b$** | **Level** | Method | **$k = 0$** | $k = 1$ | $k = 2$ |
| **$a = 0, b = 0$** | **$\alpha = 0.05$** | Lspline | 0.008 | 0.020 | 0.034 |
| | | Kernel | 0.024 | 0.028 | 0.038 |
| | **$\alpha = 0.10$** | Lspline | 0.036 | 0.050 | 0.076 |
| | | Kernel | 0.066 | 0.062 | 0.070 |
| | **$\alpha = 0.25$** | Lspline | 0.168 | 0.184 | 0.222 |
| | | Kernel | 0.224 | 0.222 | 0.228 |
| **$a = 0.32, b = 0$** shoulder at $t$=0.25 | **$\alpha = 0.05$** | Lspline | 0.030 | 0.034 | 0.064 |
| | | Kernel | 0.028 | 0.042 | 0.084 |
| | **$\alpha = 0.10$** | Lspline | 0.062 | 0.098 | 0.134 |
| | | Kernel | 0.078 | 0.092 | 0.178 |
| | **$\alpha = 0.25$** | Lspline | 0.176 | 0.294 | 0.318 |
| | | Kernel | 0.254 | 0.310 | 0.380 |

Table 2: One Bump: Proportion of times $H_0$ is rejected at level $\alpha$ based on 500 simulations of $Y_i = m(t_i|a,b) + \varepsilon_i$, $i = 1, \ldots, 101$, $\varepsilon_i \sim N(0, (0.05)^2)$.

| | | | $\boldsymbol{H_0:}$ **# of bumps** $\boldsymbol{\leq k}$ | | |
|---|---|---|---|---|---|
| $\boldsymbol{a, b}$ | **Level** | Method | $k=0$ | $\boldsymbol{k=1}$ | $k=2$ |
| $\boldsymbol{a=0, b=0.15}$ small bump at $t=0.75$ | $\boldsymbol{\alpha = 0.05}$ | Lspline | 0.986 | 0.014 | 0.022 |
| | | Kernel | 1.000 | 0.012 | 0.020 |
| | $\boldsymbol{\alpha = 0.10}$ | Lspline | 0.998 | 0.036 | 0.052 |
| | | Kernel | 1.000 | 0.022 | 0.048 |
| | $\boldsymbol{\alpha = 0.25}$ | Lspline | 1.000 | 0.134 | 0.168 |
| | | Kernel | 1.000 | 0.126 | 0.166 |
| $\boldsymbol{a=0.32, b=0.20}$ shoulder at $t=0.25$ large bump at $t=0.75$ | $\boldsymbol{\alpha = 0.05}$ | Lspline | 1.000 | 0.048 | 0.032 |
| | | Kernel | 1.000 | 0.038 | 0.042 |
| | $\boldsymbol{\alpha = 0.10}$ | Lspline | 1.000 | 0.088 | 0.056 |
| | | Kernel | 1.000 | 0.076 | 0.068 |
| | $\boldsymbol{\alpha = 0.25}$ | Lspline | 1.000 | 0.244 | 0.202 |
| | | Kernel | 1.000 | 0.236 | 0.212 |
| $\boldsymbol{a=0.32, b=0.10}$ shoulder at $t=0.25$ small bump at $t=0.75$ | $\boldsymbol{\alpha = 0.05}$ | Lspline | 0.678 | 0.042 | 0.048 |
| | | Kernel | 0.698 | 0.056 | 0.048 |
| | $\boldsymbol{\alpha = 0.10}$ | Lspline | 0.802 | 0.088 | 0.078 |
| | | Kernel | 0.782 | 0.106 | 0.110 |
| | $\boldsymbol{\alpha = 0.25}$ | Lspline | 0.930 | 0.264 | 0.206 |
| | | Kernel | 0.878 | 0.274 | 0.310 |
| $\boldsymbol{a=0.64, b=0}$ large bump at $t=0.25$ | $\boldsymbol{\alpha = 0.05}$ | Lspline | 0.546 | 0.078 | 0.032 |
| | | Kernel | 0.838 | 0.066 | 0.044 |
| | $\boldsymbol{\alpha = 0.10}$ | Lspline | 0.812 | 0.088 | 0.078 |
| | | Kernel | 0.870 | 0.120 | 0.112 |
| | $\boldsymbol{\alpha = 0.25}$ | Lspline | 0.968 | 0.152 | 0.254 |
| | | Kernel | 0.992 | 0.292 | 0.302 |

Table 3: Two Bumps: Proportion of times $H_0$ is rejected at level $\alpha$ based on 500 simulations of $Y_i = m(t_i|a,b) + \varepsilon_i$, $i = 1, \ldots, 101$, $\varepsilon_i \sim N(0, (0.05)^2)$.

| | | | $\boldsymbol{H_0}$ **: # of bumps** $\leq \boldsymbol{k}$ | | | |
|---|---|---|---|---|---|---|
| $\boldsymbol{a, b}$ | **Level** | Method | $k = 0$ | $k = 1$ | $\boldsymbol{k = 2}$ | $k = 3$ |
| $\boldsymbol{a = 0.64, b = 0.20}$ large bumps at $t = 0.25, 0.75$ | $\boldsymbol{\alpha = 0.05}$ | Lspline | 1.000 | 0.958 | 0.020 | 0.024 |
| | | Kernel | 0.922 | 0.996 | 0.016 | 0.016 |
| | $\boldsymbol{\alpha = 0.10}$ | Lspline | 1.000 | 0.978 | 0.044 | 0.050 |
| | | Kernel | 0.972 | 1.000 | 0.032 | 0.050 |
| | $\boldsymbol{\alpha = 0.25}$ | Lspline | 1.000 | 0.998 | 0.158 | 0.172 |
| | | Kernel | 1.000 | 1.000 | 0.146 | 0.172 |
| $\boldsymbol{a = 0.48, b = 0.15}$ medium bumps at $t = 0.25, 0.75$ | $\boldsymbol{\alpha = 0.05}$ | Lspline | 1.000 | 0.368 | 0.034 | 0.020 |
| | | Kernel | 0.976 | 0.708 | 0.078 | 0.022 |
| | $\boldsymbol{\alpha = 0.10}$ | Lspline | 1.000 | 0.586 | 0.074 | 0.060 |
| | | Kernel | 1.000 | 0.814 | 0.120 | 0.064 |
| | $\boldsymbol{\alpha = 0.25}$ | Lspline | 1.000 | 0.832 | 0.206 | 0.172 |
| | | Kernel | 1.000 | 0.918 | 0.242 | 0.226 |
| $\boldsymbol{a = 0.48, b = 0.10}$ medium bump at $t = 0.25$ small bump at $t = 0.75$ | $\boldsymbol{\alpha = 0.05}$ | Lspline | 0.650 | 0.420 | 0.036 | 0.036 |
| | | Kernel | 0.640 | 0.738 | 0.100 | 0.044 |
| | $\boldsymbol{\alpha = 0.10}$ | Lspline | 0.816 | 0.614 | 0.076 | 0.070 |
| | | Kernel | 0.748 | 0.828 | 0.162 | 0.092 |
| | $\boldsymbol{\alpha = 0.25}$ | Lspline | 0.934 | 0.864 | 0.224 | 0.234 |
| | | Kernel | 0.914 | 0.910 | 0.316 | 0.244 |

Table 4: Zero, One and Two Bumps: Proportion of times $H_0$ is rejected at level $\alpha$ based on 500 simulations of $Y_i = m(t_i|a,b) + \varepsilon_i$, $i = 1, \ldots, 101$, $\varepsilon_i \sim N(0, (0.005)^2)$.

| | | $\boldsymbol{H_0}$ **: # of bumps** $\boldsymbol{\leq k}$ | | | |
|---|---|---|---|---|---|
| **Zero bumps** | **Level** | $\boldsymbol{k=0}$ | $k=1$ | $k=2$ | |
| | $\alpha = 0.05$ | 0.034 | 0.014 | 0.046 | |
| $\boldsymbol{a=0, b=0}$ | $\alpha = 0.10$ | 0.072 | 0.050 | 0.104 | |
| | $\alpha = 0.25$ | 0.188 | 0.208 | 0.290 | |
| $\boldsymbol{a=0.32, b=0}$ | $\alpha = 0.05$ | 0.054 | 0.018 | 0.088 | |
| shoulder at $t = 0.25$ | $\alpha = 0.10$ | 0.084 | 0.074 | 0.180 | |
| | $\alpha = 0.25$ | 0.204 | 0.258 | 0.400 | |
| **One bump** | **Level** | $k=0$ | $\boldsymbol{k=1}$ | $k=2$ | |
| $\boldsymbol{a=0, b=0.20}$ | $\alpha = 0.05$ | 1.000 | 0.008 | 0.014 | |
| large bump at $t = 0.75$ | $\alpha = 0.10$ | 1.000 | 0.032 | 0.038 | |
| | $\alpha = 0.25$ | 1.000 | 0.140 | 0.204 | |
| $\boldsymbol{a=0.32, b=0.20}$ | $\alpha = 0.05$ | 1.000 | 0.020 | 0.042 | |
| shoulder at $t = 0.25$ | $\alpha = 0.10$ | 1.000 | 0.054 | 0.076 | |
| small bump at $t = 0.75$ | $\alpha = 0.25$ | 1.000 | 0.220 | 0.244 | |
| **Two bumps** | **Level** | $k=0$ | $k=1$ | $\boldsymbol{k=2}$ | $k=3$ |
| $\boldsymbol{a=0.64, b=0.20}$ | $\alpha = 0.05$ | 0.986 | 0.898 | 0.026 | 0.026 |
| large bumps at | $\alpha = 0.10$ | 0.996 | 0.982 | 0.064 | 0.076 |
| $t = 0.25, 0.75$ | $\alpha = 0.25$ | 0.998 | 0.998 | 0.186 | 0.242 |
| $\boldsymbol{a=0.48, b=0.15}$ | $\alpha = 0.05$ | 0.832 | 0.090 | 0.104 | 0.044 |
| medium bumps at | $\alpha = 0.10$ | 0.904 | 0.290 | 0.164 | 0.116 |
| $t = 0.25, 0.75$ | $\alpha = 0.25$ | 0.968 | 0.662 | 0.306 | 0.292 |

Table 5: Classifications of male and female growth curves by the number of growth spurts, using the strict and soft decision rules.

| **Number of bumps** | | 0 | 1 | 2 | 3 | $\geq 3$ |
|---|---|---|---|---|---|---|
| Males | strict | 0 | 10 (29.4%) | 17 (50.0%) | 6 (17.6%) | 1 (2.9%) |
| | soft | 0 | 17 (50.0%) | 19 (44.2%) | 6 (14.0%) | 1 (2.3%) |
| Females | strict | 0 | 17 (43.6%) | 15 (38.5%) | 5 (12.8%) | 2 (5.1%) |
| | soft | 1 (2.0%) | 24 (48.0%) | 18 (36.0%) | 5 (10.0%) | 2 (4.0%) |

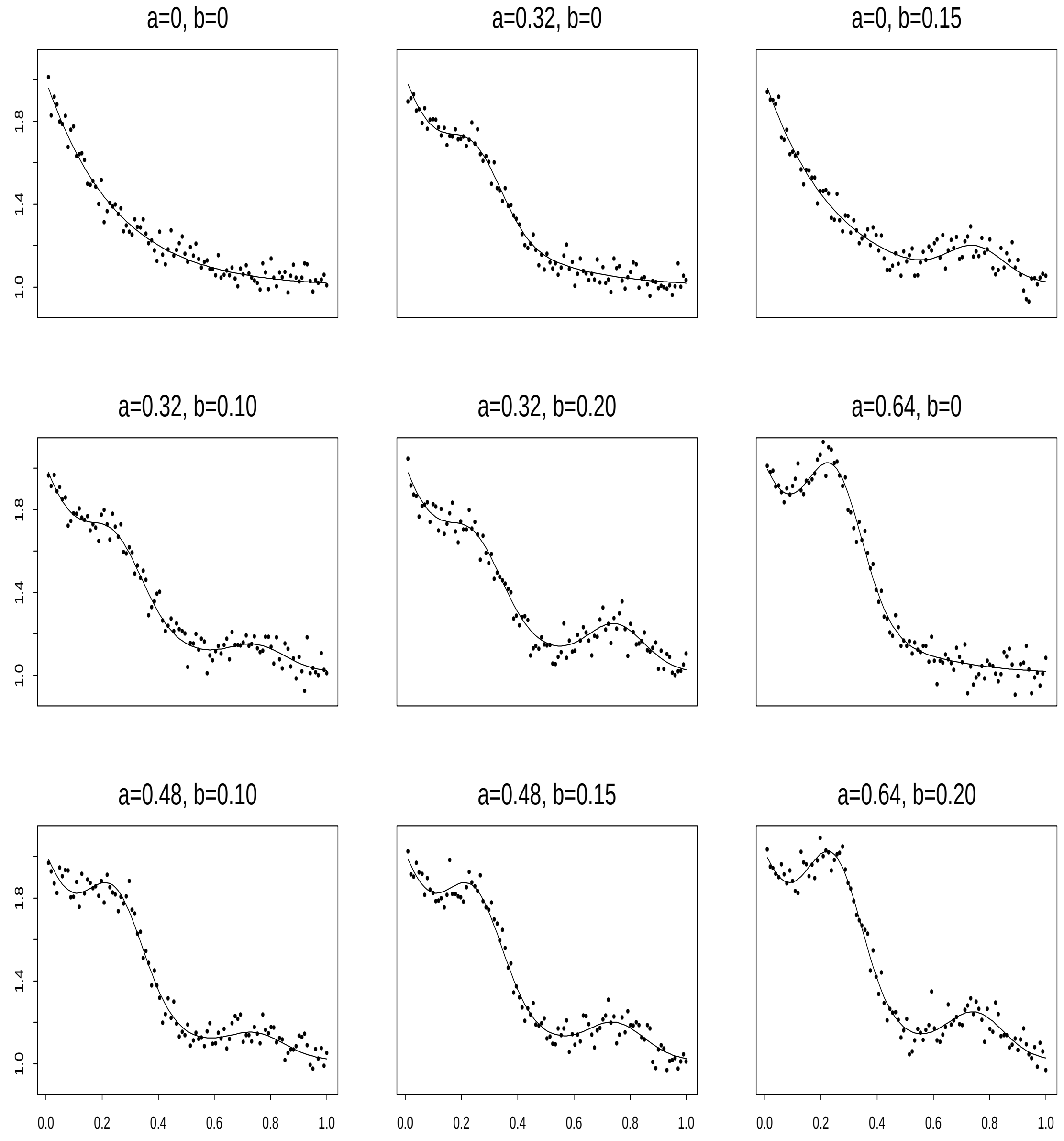


Figure 1: Functions $m(t|a,b) = 1 + \exp(-4t) + a * B(t|0.25) + b * B(t|0.75)$ and simulated data with the level of noise $\sigma = 0.05$.

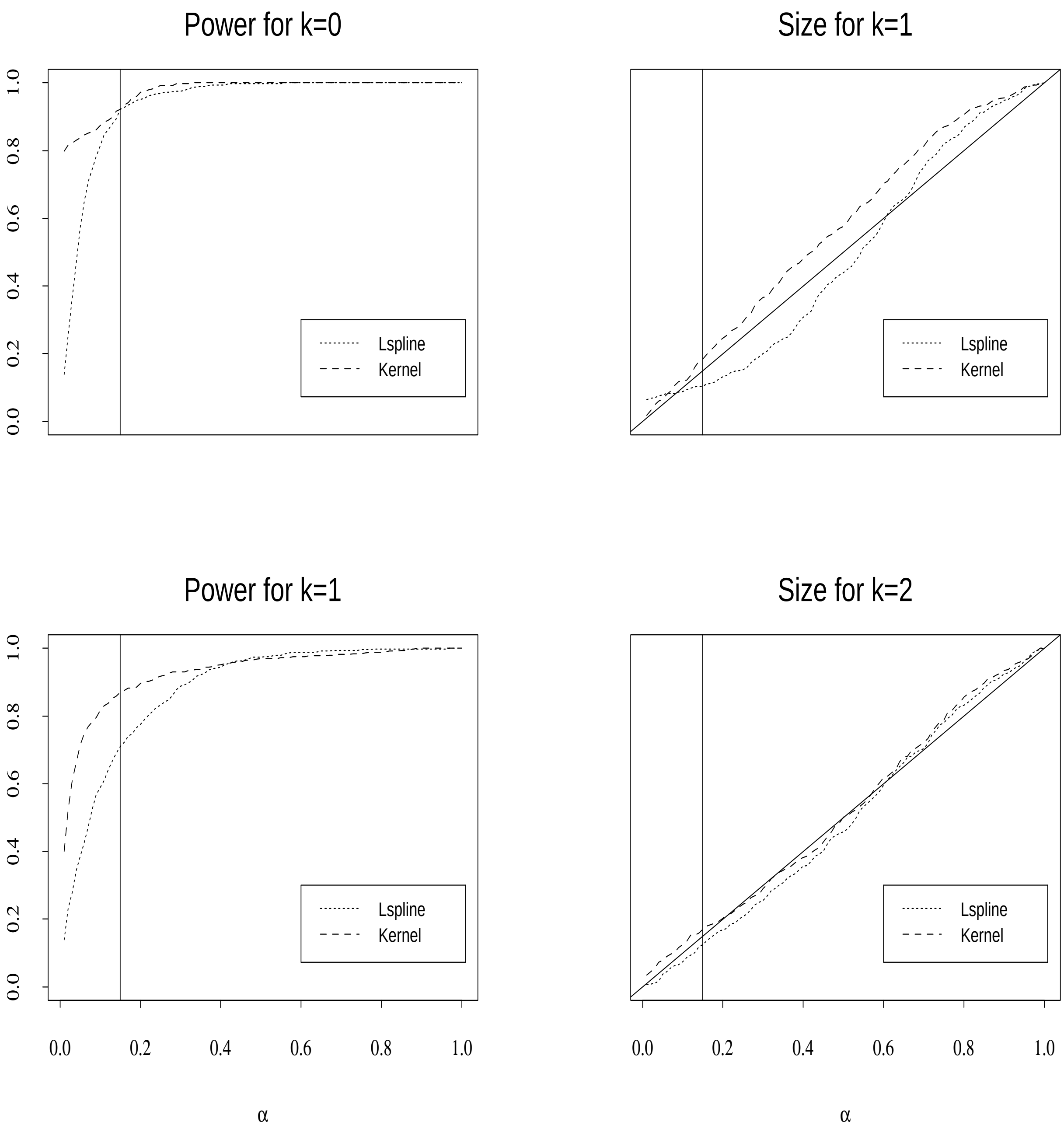


Figure 2: Power and size estimates for the functions $m(t|a = 0.64, b = 0)$ (top) and $m(t|a = 0.48, b = 0.15)$ (bottom). The vertical line is drawn at $\alpha = 0.15$ and a 45 degree line is drawn on the size plots.

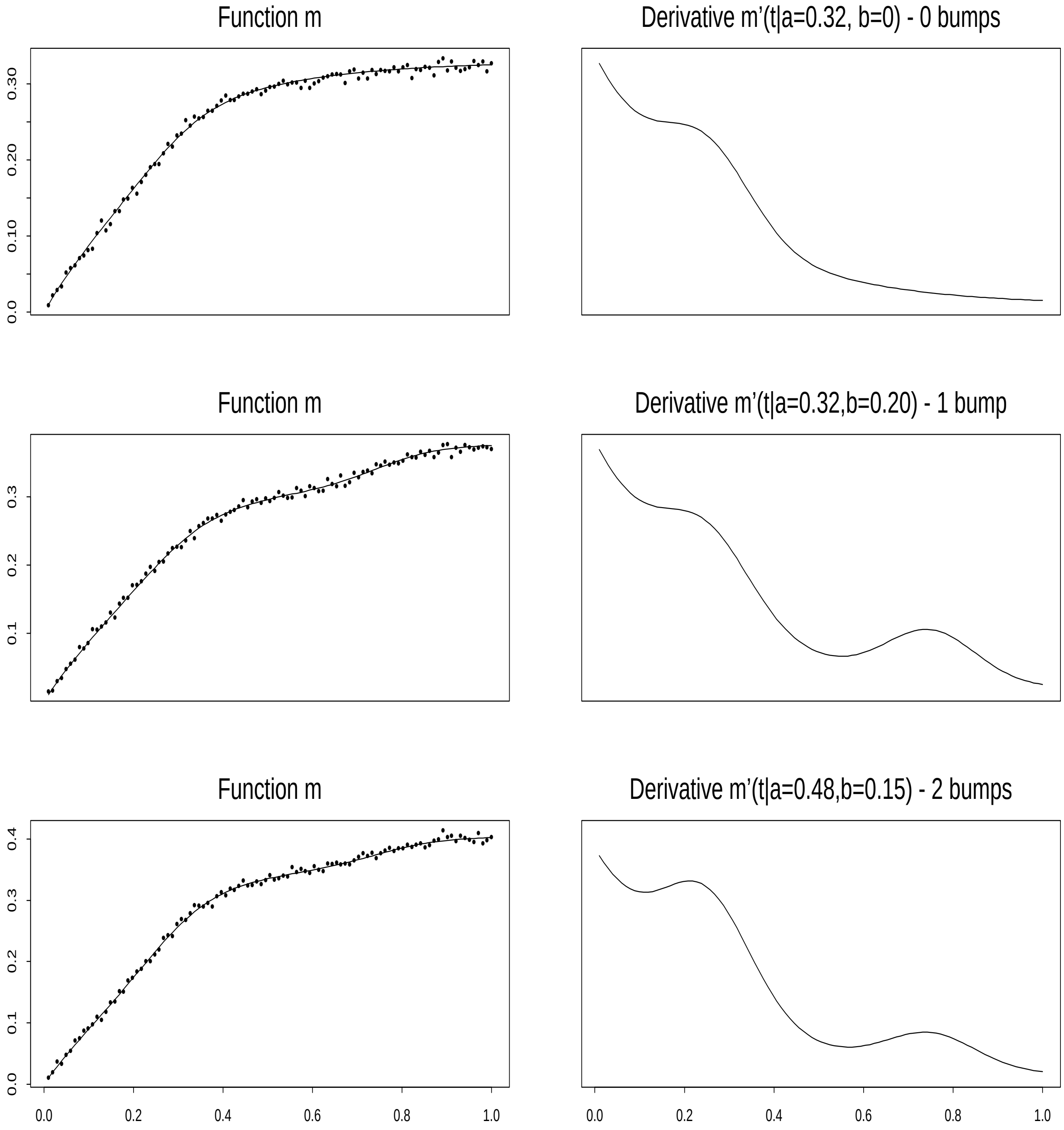


Figure 3: Functions $m(t|a, b)$ (left) and their derivatives $m'$ (right) with 0, 1 and 2 bumps. Also included are data sets (left) with noise level $\sigma = 0.005$.

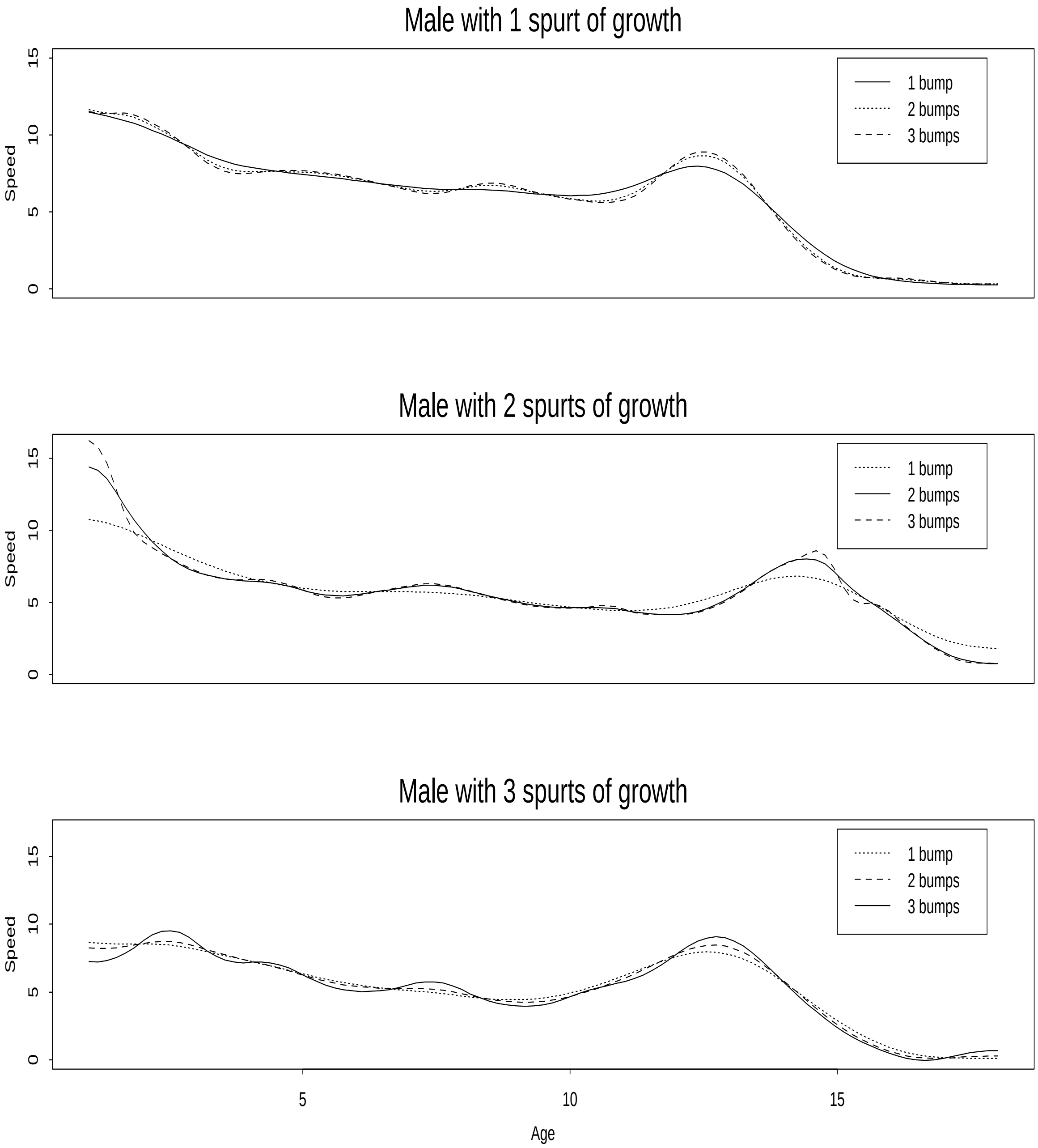


Figure 4: Estimates of the speed of growth of males with 1, 2, and 3 bumps respectively.