The Annals of Statistics
1999, Vol. 27, No. 1, 299–337

DISCONTINUOUS VERSUS SMOOTH REGRESSION[1]

BY HANS-GEORG MÜLLER AND ULRICH STADTMÜLLER

University of California, Davis and Universität Ulm

Given measurements (x_i, y_i), $i = 1, \dots, n$, we discuss methods to assess whether an underlying regression function is smooth (continuous or differentiable) or whether it has discontinuities. The variance of the measurements is assumed to be unknown, and is estimated simultaneously. By regressing squared differences of the data formed with various span sizes on the span size itself, we obtain an asymptotic linear model with dependent errors. The parameters of this asymptotic linear model include the sum of the squared jump sizes as well as the variance of the measurements. Both parameters can be consistently estimated, with mean squared error rates of convergence of $n^{-2/3}$ for the sum of squared jump sizes and n^{-1} for the error variance. We derive the asymptotic constants of the mean squared error (MSE) and discuss the dependence of MSE on the maximum span size L. The test for the existence of jumps is formulated for the null hypothesis that the sum of squared jump sizes is 0. The asymptotic distribution of the test statistic is obtained essentially via a central limit theorem for U-statistics. We motivate and illustrate the methods with data surrounded by a scientific controversy concerning the question whether the growth of children occurs smoothly or rather in jumps.

1. Introduction.

1.1. *Background.* It is customary in applications of nonparametric regression analysis to assume that the function to be estimated is smooth; in fact, this assumption provides the basic motivation as well as technical justification for using smoothing methods. However, as was learned in recent years, in a number of important applications, the underlying function is smooth everywhere except at a critical number of points where jump discontinuities may occur. Examples from various scientific fields are the Nile data [Cobb (1978)], the coal mining disaster data [Jarrett (1979)], single channel patch clamp recordings [Fredkin and Rice (1992)], the segmentation of DNA sequences [Churchill (1992)], stock market data [Wang (1995)] and the crown-heel lengths growth data of Lampl, Veldhuis and Johnson (1992). Some methods that have been proposed for nonparametric regression analysis like wavelet implementations with coefficient shrinkage [see, for instance,

Received March 1997; revised August 1998.
[1]Research supported in part by a NATO grant, NSA Grant MDA 904-96-1-0026 and NSF Grants DMS-94-04906 and DMS-96-25984.
AMS 1991 *subject classifications*. 62G07, 62G10.
Key words and phrases. Asymptotic linear model, variance estimation, goodness of fit, jump detection, model selection, rate of convergence, saltatory growth, test for discontinuity, U-statistics.

Donoho, Johnstone, Kerkyacharian and Picard (1995)] in fact regularly turn out curve estimates with nonsmooth features.

If discontinuities or change-points are indeed present within an otherwise smooth regression function, their locations can be estimated efficiently with methods that have been developed recently; compare the proposals by Müller (1992), Wu and Chu (1993), Eubank and Speckman (1994) and Wang (1995). Once the locations have been obtained, one can then adapt common smoothing techniques to the presence of jump discontinuities by treating the change-point locations as endpoints of the support; this can be achieved with boundary kernels if one uses convolution-type kernel estimators or with modified locally weighted least squares-type kernel estimators [Müller (1993)]. For further work on the intersection between smoothing and change-points as well as on inference for change-points compare, for instance, Hinkley (1970), Bhattacharya and Brockwell (1976), Hall and Titterington (1992), Carlstein, Müller and Siegmund (1994) and Müller and Song (1997).

An important motivation for the work reported here is that if one does assume that the function of interest contains discontinuities, then the resulting curve estimates with discontinuities are not only quantitatively but also qualitatively different from smooth curve estimates, which one desires to obtain if the underlying curve is indeed smooth. The appearance of estimated curves when discontinuities are assumed to be present is strikingly different from estimated curves under global smoothness assumptions, paving the way for substantially different conclusions. A case in point is the application to growth data, to be discussed in more detail in Sections 2 and 5.

Application of common smoothing methods like smoothing splines, kernel and locally weighted least squares estimates (with the possible exception of wavelets) invariably will lead to smooth curve estimates, whether discontinuities are present or not. If discontinuities are present, they will be oversmoothed and will not be visible in resulting curve estimates.

It is therefore of interest to gather as much knowledge as possible regarding the question whether such jumps really exist in the data. For curve estimates based on wavelets, jumps may appear but could be artifacts of the method. If indeed no jumps or sharp cusps exist, smooth curve estimates should be used. However, if jumps exist, modified smoothing methods which allow for the inclusion of nonsmooth parts must be used. Essentially, this is a problem of model selection, where the choice is between a class of smooth regression functions and a larger class of functions which include discontinuities.

It is, then, an important data-analytic problem to develop tools for diagnostic assessment and inference with the aim of aiding the statistician in the decision whether an unknown function, which cannot be parametrically specified, should be modelled as a globally smooth function or as a function which is smooth but contains isolated discontinuities. The statistics proposed in this paper are designed to provide relevant information for this decision. Furthermore, they may also be of general interest as diagnostic tools in applied regression analysis.

1.2. *Proposed model.* We describe now the basic modelling framework within which we discuss diagnostics and tests for discontinuities. Consider the following classes of functions. For given constants $M, \xi > 0$, let

$$S_C(M) = \Big\{ f\colon [0,1] \to \mathbf{R},\ f \text{ is continuously differentiable}, \sup_{0 \le x \le 1} |f'(x)| \le M \Big\}, \tag{1.1}$$

$$S_D(\xi) = \Big\{ f\colon [0,1] \to \mathbf{R},\ f(x) = \sum_{i=0}^{m-1} c_i 1_{I_i}(x), \tag{1.2}$$

where $m \in \{1, 2, 3, \dots\}$ is an arbitrary integer, $c_0, c_1, \dots, c_{m-1}$ is an arbitrary sequence of reals, and

$$I_i = [\tau_i, \tau_{i+1}), \qquad 0 \le i \le m-2, \qquad I_{m-1} = [\tau_{m-1}, 1],$$

for an arbitrary sequence τ_i with

$$\tau_0 = 0 < \tau_1 < \tau_2 < \cdots < \tau_m = 1 \text{ and } \min_{1 \le i \le m} (\tau_i - \tau_{i-1}) \ge \xi \Big\}.$$

Here, $S_C(M)$ denotes a class of continuously differentiable functions, whereas $S_D(\xi)$ is the class of discontinuous step functions with an arbitrary but finite number of steps and a minimum distance ξ between successive jump points. We note that the restriction to functions with support $[0,1]$ is made only for ease of notation.

The class of regression functions to be considered in our statistical model then draws on S_C and S_D as follows: the underlying regression function is assumed to be composed as a sum of a function from $S_C(M)$ and a function from $S_D(\xi)$. In particular, we do not wish to limit the possible number of jumps which occur in the discontinuous part; the development in the following will include the case where the number of jumps is not fixed but may increase as the sample size n increases. This reflects the expectation that one should be able to detect more and more finely grained discontinuities for larger and larger sample sizes. It is therefore natural to assume that the regression function for which we wish to ascertain whether it is continuous or not is allowed to depend on n.

One quantity which needs to remain fixed for varying sample sizes n and indeed emerges as our natural measure of the amount of discontinuity in the data is

$$\gamma = \sum_{i=0}^{m-2} (c_{i+1} - c_i)^2, \tag{1.3}$$

which is defined for any $h \in S_D$. Here and in all of the following we adopt the convention that $\sum_{i=i_1}^{i_2} \alpha_i = 0$ if $i_2 < i_1$, so that $\gamma = 0$ if $m = 1$.

A second quantity of auxiliary nature which also needs to remain fixed throughout and measures the “interaction” between continuous and discontinuous parts is obtained given a continuous function $g \in S_C$ and a step function $h \in S_D$,

$$(1.4)\qquad \begin{aligned} \delta &= \int_0^1 g'(t)^2\,dt + 2\sum_{i=0}^{m-2} g'(\tau_{i+1})(c_{i+1} - c_i) \\ &= \int_0^1 g'(t)^2\,dt + 2\int_0^1 g'(t)\,dh(t). \end{aligned}$$

We now define for $\gamma \geq 0$, $\delta \in \mathbf{R}$ the class of functions

$$(1.5)\qquad \begin{aligned} &S_{\gamma,\delta}(M,\xi) \\ &= \bigg\{ f\colon [0,1] \to \mathbf{R},\ f(x) = g(x) + h(x),\ g \in S_C(M),\ h \in S_D(\xi), \\ &\qquad \sum_{i=0}^{m-2} (c_{i+1} - c_i)^2 = \gamma,\ \int_0^1 g'(t)^2\,dt + 2\int_0^1 g'(t)\,dh(t) = \delta \bigg\}. \end{aligned}$$

It is assumed that the data are recorded according to the fixed design regression model

$$(1.6)\qquad y_{i,n} = f_n(x_{i,n}) + \varepsilon_{i,n}, \qquad i = 1,\dots,n,$$

where $x_{i,n} = i/n$ (equidistant design) and $f_n \in S_{\gamma,\delta}(M,\xi)$, for a fixed large M and a sequence $\xi = \xi_n$ to be specified in (A1) below. The errors $\varepsilon_{i,n}$ are i.i.d. and are assumed to satisfy

$$(1.7)\qquad E\varepsilon_{i,n} = 0, \qquad E\varepsilon_{i,n}^2 = \sigma^2 \quad \text{and} \quad E\varepsilon_{i,n}^4 = \mu_4 < \infty.$$

Note that the regression function f_n is allowed to depend on n. Within the class of functions $S_{\gamma,\delta}(M,\xi)$, $\gamma = 0$ defines the subclass of "smooth" functions, whereas $\gamma > 0$ guarantees that the functions have jump discontinuities. The null hypothesis of a smooth regression function thus corresponds to $\gamma = 0$, while the case of a discontinuous regression function corresponds to $\gamma > 0$. We note that in particular, the number of jumps $m = m(n)$ may depend on and grow with n, and also that the sizes of individual jumps $(c_{i+1,n} - c_{i,n})$ may depend on n. In the following, we omit subscripts n whenever feasible.

The estimation of pure step functions $f(x) = \sum_{i=0}^{m-1} c_i 1_{I_i}(x)$ for a fixed number of jumps was thoroughly investigated in Yao (1984) and Yao and Au (1989). We note that for a given function $f \in S_{\gamma,\delta}(M,\xi)$, the decomposition $f = g + h$ with $h \in S_D(\xi)$, $g \in S_C(M)$, is unique up to a constant which can be shifted between h and g. This does not matter for our purposes.

1.3. *Aims and overview.* Our aims are to estimate the two parameters $\sigma^2 = E\varepsilon_i^2$ and $\gamma = \sum_{i=0}^{m-1}(c_{i+1} - c_i)^2$ and to test the null hypothesis $H_0\colon \gamma = 0$ of a smooth function. We note that there is an extensive literature on the estimation of the error variance σ^2 when the regression function f is "smooth," say, $f \in S_C(M)$. These estimates work by using squared differences of the data of various orders [see, for example, Rice (1984), Gasser, Sroka and Jennen–Steinmetz (1989), Hall, Kay and Titterington (1990)]. A simple example is the estimate

$$\tilde{\sigma}^2 = \frac{1}{2n} \sum_{i=1}^{n-1} (y_{i+1} - y_i)^2,$$

which was proposed in Rice (1984).

If $f \in S_{\gamma, \delta}$ with $\gamma > 0$, these difference scheme estimates are disturbed by the jumps, which inflate the finite variance estimates; however, this inflation effect disappears asymptotically for a fixed finite number of jumps. It will not disappear, however, on sets $S_{\gamma, \delta}$ with $\gamma > 0$, where the number of jumps is potentially unlimited. It is therefore of considerable interest to estimate the error variance σ^2 simultaneously with γ, in the presence of jumps. These two parameters are complementary in the sense that seemingly erratic data y_i could be caused by either a high level of noise variance or by the presence of jump discontinuities. Discrimination between these two opposite causes is a fairly difficult task, and this paper will provide some relevant tools to attack this problem.

The paper is organized as follows: a controversy on the existence of saltatory growth serves as data-analytic motivation for the comparison of discontinuous versus smooth regression. This controversy is introduced in the following Section 2 and its discussion is resumed in Section 6. The proposed statistics and test procedures as well as the asymptotic linear model which forms the backbone for our approach are described in Section 3. The main asymptotic results on consistency, rates of convergence and asymptotic normality are in Section 4. Results of simulations and finite sample aspects are discussed in Section 5. The main proofs are compiled in Section 7. Predominantly technical calculations are deferred to Appendix A.1, while Appendix A.2 provides more details on a preliminary study of fitting multiple change-points in a nonparametric regression setting.

2. A Controversy on saltatory growth.

2.1. *The "saltation and stasis" hypothesis.* In 1992, Lampl, Veldhuis and Johnson published a study on the growth of infants in the journal *Science*. They claimed that their study confirmed the "saltatory growth" or "saltation and stasis" hypothesis: this hypothesis refers to the existence of jump discontinuities in the growth of infants. The authors claimed that the existence of jump discontinuities can be inferred from daily measurements of crown-heel length. This finding was supported by anecdotal accounts of rapid overnight growth of children who were reported to have gained more than one inch in height during a single night. Lampl, Veldhuis and Johnson (1992) also added a "stasis" component, periods between saltations during which very little growth was supposed to occur.

The findings of Lampl, Veldhuis and Johnson (1992) were disputed in a 1995 *Science* article by Heinrichs, Munson, Counts, Cutler and Baron. In their response, Lampl, Cameron, Veldhuis and Johnson (1995) upheld their original findings. Whereas Lampl et al. argued on the basis of their growth measurements, Heinrichs et al. reported their own crown-heel lengths measurements and interpreted their data as not containing any evidence for the saltation and stasis hypothesis. They also pointed out that the biological requirements for saltatory growth not only violate the dogma "Natura non facit saltus" but are genuinely daunting. To initiate a saltatory growth spurt, all cells in the growth zones of the bones, the epiphyses, would have to synchronize their cell cycle in order to achieve a noticeable overall height gain by dividing simultaneously. All this activity would have to be squeezed into a very brief time interval.

In the following, we discuss the data on crown-heel lengths of a single infant as reported in Figure 1 of Heinrichs et al. (1995). These data are 30 daily height measurements from approximately 67 to 97 days of age, and were also used by Lampl et al. (1995); see their Figure 1, in an attempt to refute the interpretation of Heinrichs et al. Lampl et al. claimed to demonstrate that these data do contain evidence for saltatory growth. We are aware of the limitations of inference that can be drawn from data of one infant only, and any conclusions are at best tentative. Nevertheless, this auxologic debate provides additional motivation for the procedures proposed in Section 3 below.

A scatterplot of these data (crown-heel length measurements versus age in days) appears in the panels of Figure 1. The simplest approach to modelling is to fit a straight line, assuming that infant growth over a limited time period of 30 days is approximately linear. Alternatively, one could assume that the data are generated from an underlying smooth (nonparametric) growth curve, allowing for smooth deviations from a linear trajectory. One would then apply any one of a variety of smoothing methods. A smooth fit using the method of local linear fitting by locally weighted least squares [see, e.g., Fan and Gijbels (1996)] with appropriate bandwidth choice is shown in Figure 1 (upper left panel), together with the simple linear least squares regression fit.

2.2. *Preliminary change-point analysis of infant growth data*. An alternative to the smooth fits in the upper left panel of Figure 1 is to apply nonparametric regression with change-points [see, e.g., Hall and Titterington (1992), Müller (1992), Wu and Chu (1993), Loader (1996)]. In such methods, one typically first assumes a fixed number ν of jump discontinuities in an otherwise smooth regression function. Then the ν change-points are located according to a local criterion, searching for maximal jump sizes. In a last step, a smooth fit is obtained on the $(\nu + 1)$ segments defined by the ν estimated change-point locations. This produces smooth curve estimates on the segments, with discontinuities where segments adjoin. An application of this idea with local linear fitting to the infant growth data for $\nu = 0, 1, 2, 3$ is

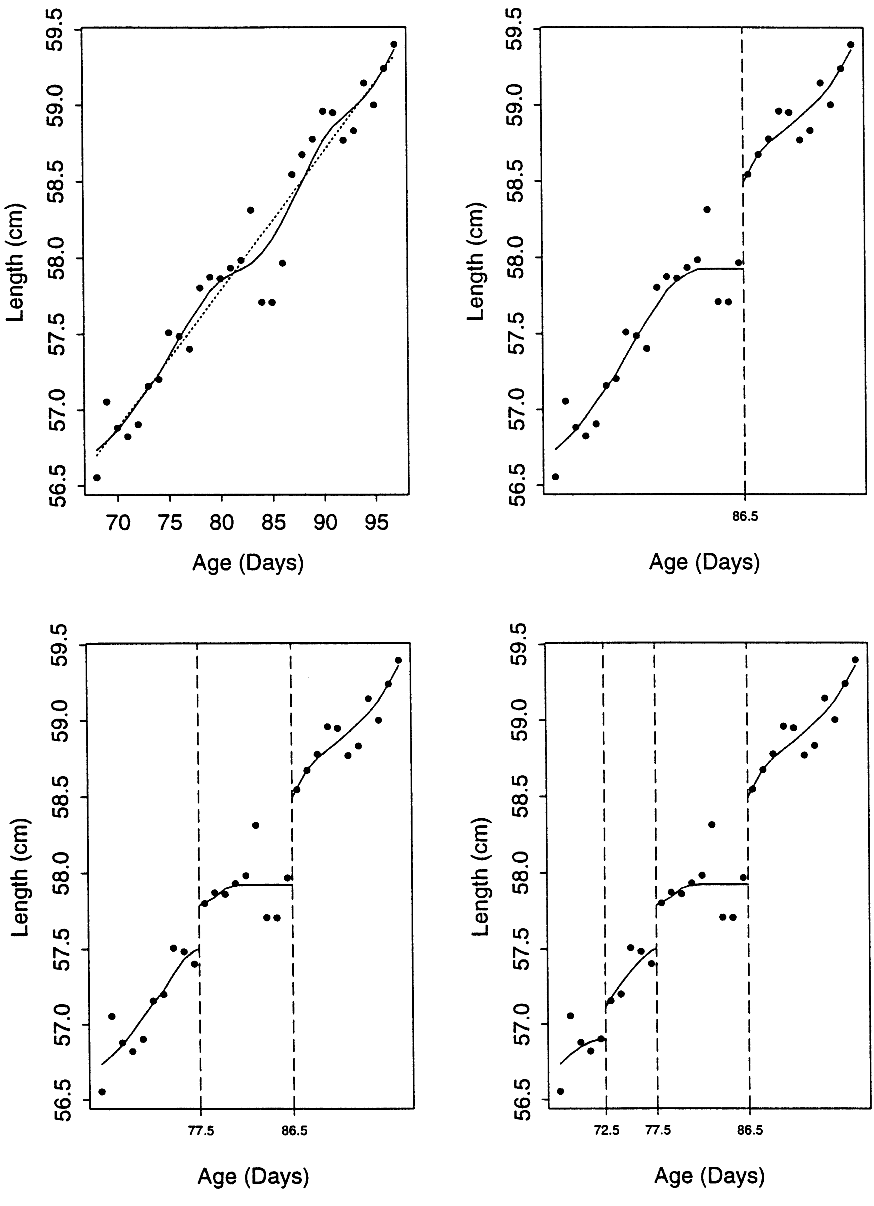

FIG. 1. *Crown-heel lengths data* [*from Heinrichs, Munson, Counts, Cutler and Baron*, 1995] *for one infant boy, length measured daily in cm, from age* 68 *to* 97 *days* ($n = 30$). *Superimposed are several regression fits*: *Upper left*: *simple least squares regression line* (*dashed*) *and smooth fit by weighted least squares fitting of local lines* (*solid*). *Upper right*: *monotone smooth fit with one jump discontinuity, using local line fitting. Lower left*: *monotone smooth fit with two jump discontinuities. Lower right*: *monotone smooth fit with three jump discontinuities.*

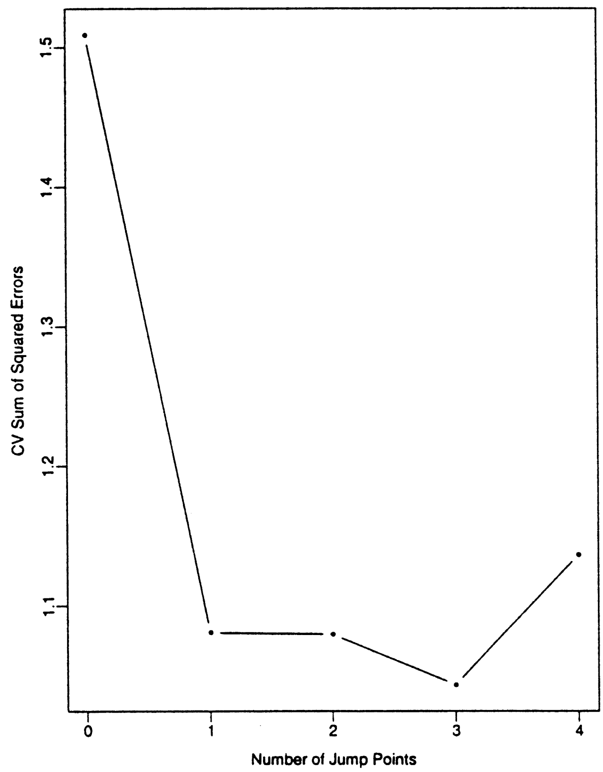

FIG. 2. *Cross-validation sum of squares as function of the number of jump points for the crown-heel lengths data.*

demonstrated in Figure 1. The fits with one (upper right panel) or two (lower left panel) change-points appear quite plausible.

One point of particular interest is a comparison of the quality of the smooth fits with ν change-points, where $\nu \geq 0$. One straightforward method is minimization of the cross-validation sum of squares

$$\mathrm{CVSS}(\nu) = \sum_{i=1}^{n} \left(y_i - \hat{y}_\nu^{(-i)}(x_i)\right)^2$$

for given scatterplot data (x_i, y_i), $i = 1, \dots, n$. Here, $\hat{y}_\nu^{(-i)}(x_i)$ is the fit obtained at x_i, assuming ν change-points and excluding the data point (x_i, y_i) when constructing the fit. The resulting plot for CVSS(ν) for the infant growth data is in Figure 2. Taken together, Figures 1 and 2 support the idea that inclusion of some degree of saltation provides the best explanation for the data. We continue this discussion in Section 6, applying the new methods which are developed below. It is clear from Figure 1 that a major distinction exists between smooth ($\nu = 0$) and nonsmooth ($\nu > 0$) models, demonstrating that the model selection issues which are addressed next are of interest beyond this particular example.

Additional pertinent details on this preliminary change-point approach can be found in Appendix A.2.

3. Asymptotic linear model and proposed estimators. As outlined in the introduction, we assume that the regression function is in $S_{\gamma,\delta}$. We then consider the null hypothesis H_0: $\gamma = 0$, that the function is smooth, which is to be tested against the alternative H_A: $\gamma > 0$, that the function contains jump discontinuities.

The inference procedures which we propose are based on sums of squared differences of the data. These differences are formed with various span sizes. Specifically, we consider the statistics

$$(3.1)\qquad Z_k = \sum_{j=1}^{n-L} (y_{j+k} - y_j)^2/(n-L), \qquad 1 \le k \le L.$$

Here, $L = L(n) \ge 1$ is a sequence of integers depending on n. Theoretical and practical choice of L is discussed in Sections 4 and 5.2.

As will be shown, the statistics Z_k can be interpreted as dependent variables within the following asymptotic simple linear model, which contains the parameters of interest σ^2 and γ as intercept and slope parameters:

$$(3.2)\qquad Z_k = 2\sigma^2 + (k/(n-L))\gamma + \eta_k, \qquad 1 \le k \le L.$$

The asymptotic linear model (3.2) is characterized by the behavior of the residual errors

$$(3.3)\qquad \eta_k = Z_k - 2\sigma^2 - (k/(n-L))\gamma, \qquad 1 \le k \le L.$$

We write $\eta = (\eta_1, \dots, \eta_L)^T$, 1_L for a $L \times L$-matrix with all entries being 1, and I_L for the $L \times L$ identity matrix, and list the following additional assumptions:

$$(\text{A1})\qquad \min_{1 \le j \le m} (\tau_j - \tau_{j-1}) \ge \xi_n = 2L/n,$$

where τ_j and m are as in (1.2). This condition ensures that different change-points and their associated jumps do not get too close asymptotically so that they can be separated by the proposed method. Furthermore, as $n \to \infty$,

$$(\text{A2})\qquad L/n \to 0,$$

$$(\text{A3})\qquad L^2/n \to \infty.$$

For all of the following, we assume that (A1) and (A2) hold. We obtain

$$(3.4)\qquad E\eta_k = O(k^2/n^2),$$

and setting $\delta_{l,k} = 1$ if $l = k$ and $\delta_{l,k} = 0$ otherwise,

$$(3.5)\qquad \operatorname{cov}(\eta_k, \eta_l) = \frac{4}{n}\big((\mu_4 - \sigma^4) + \sigma^4\delta_{k,l}\big) + O\left(\frac{L}{n^2}\right)$$

uniformly in $1 \le k, l \le L$.

Hence one may identify a leading covariance matrix for η,

$$C_n = \frac{4}{n}\left(\left(\mu_4 - \sigma^4\right)1_L + \sigma^4 I_L\right) = \frac{4}{n}C_0, \tag{3.6}$$

where C_0 is seen to be a nonsingular $L \times L$ matrix.

For a proof of (3.4), (3.5), we refer to the proof of Theorem 4.1 in Section 7; compare also the remark after Theorem 4.1. Note that in the case of normal errors, $\varepsilon_i \sim \mathscr{N}(0, \sigma^2)$, $\mu_4 - \sigma^4 = 2\sigma^4$ and therefore $\operatorname{cov}(\eta) = (4\sigma^4/n)\{2_L + I_L\}$, ignoring terms of smaller order.

Setting $\beta = 2\sigma^2$, we may rewrite the linear regression model (3.2) as

$$Z_k = \beta + \frac{k}{n-L}\gamma + \eta_k, \qquad 1 \le k \le L, \tag{3.7}$$

with the design matrix

$$A = (a_{ij}) = \begin{cases} 1, & \text{if } j = 1, \\ i/(n-L), & \text{if } j = 2, \end{cases} \qquad 1 \le i \le L. \tag{3.8}$$

One finds immediately

$$A^T A = \begin{pmatrix} L & \dfrac{L(L+1)}{2(n-L)} \\ \dfrac{L(L+1)}{2(n-L)} & \dfrac{L(L+1)(2L+1)}{6(n-L)^2} \end{pmatrix} \tag{3.9}$$

and

$$\begin{aligned}(A^T A)^{-1} &= \frac{12(n-L)^2}{L(L^2-1)} \\ &\quad\times \begin{pmatrix} (L+1)(2L+1)/(6(n-L))^2 & -(L+1)/(2(n-L)) \\ -(L+1)/(2(n-L)) & 1 \end{pmatrix}.\end{aligned} \tag{3.10}$$

The least squares estimator for $\binom{\beta}{\gamma}$ is then given by

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma}_1 \end{pmatrix} = (A^T A)^{-1} A^T Z = \begin{pmatrix} \dfrac{2}{L(L-1)} \displaystyle\sum_{k=1}^{L} (2L + 1 - 3k) Z_k \\ \dfrac{6(n-L)}{L(L^2-1)} \displaystyle\sum_{k=1}^{L} (2k - (L+1)) Z_k \end{pmatrix} \tag{3.11}$$

and for $\binom{\sigma^2}{\gamma}$ by

$$\begin{pmatrix} \hat{\sigma}_1^2 \\ \hat{\gamma}_1 \end{pmatrix} = B \begin{pmatrix} \hat{\beta} \\ \hat{\gamma}_1 \end{pmatrix} \quad \text{with } B = \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.12}$$

Since the data are not i.i.d., but have covariance structure (3.5), one should actually benefit from using the weighted least squares estimator with weights determined by the asymptotic covariance matrix of Z, which is given by C_n (3.6). However, according to a result of McElroy (1967), matrices C_n (3.6) happen to be of a type where the weighted and ordinary least squares estimators coincide. Therefore, (3.11) and (3.12) simultaneously provide weighted and unweighted least squares estimators.

The covariance matrix of $\binom{\hat\sigma_1^2}{\hat\gamma_1}$ is given by

$$\operatorname{Cov}\begin{pmatrix}\hat\sigma_1^2\\ \hat\gamma_1\end{pmatrix} = B(A^TA)^{-1}\left(A^T\tilde C_n A\right)A^TAB^T, \tag{3.13}$$

where $\tilde C_n = \operatorname{Cov}(Z) = \operatorname{Cov}(\eta)$.

4. Asymptotic results and testing for jumps. Our first result provides the asymptotic behavior of the first two moments of estimator (3.12).

THEOREM 4.1. *Under* (A1)–(A3), *we have that*

$$E\begin{pmatrix}\hat\sigma_1^2\\ \hat\gamma_1\end{pmatrix} = \begin{pmatrix}\sigma^2\\ \gamma\end{pmatrix} + \begin{pmatrix}O(L^2/n^2)\\ O(L/n)\end{pmatrix} \tag{4.1}$$

and

$$\operatorname{Cov}\begin{pmatrix}\sqrt{n}\,\hat\sigma_1^2\\ \sqrt{L}\,\hat\gamma_1\end{pmatrix} \to (\mu_4 - \sigma^4)\begin{pmatrix}1 & 0\\ 0 & \dfrac{12}{5}\end{pmatrix} \quad \text{as } n \to \infty, \tag{4.2}$$

provided that $\gamma = 0$. *If* $E\varepsilon_1^3 = 0$, *then we have for arbitrary* $\gamma \ge 0$ *that*

$$\operatorname{Cov}\begin{pmatrix}\sqrt{n}\,\hat\sigma_1^2\\ \sqrt{L}\,\hat\gamma_1\end{pmatrix} \to \begin{pmatrix}\mu_4 - \sigma^4 & 0\\ 0 & \dfrac{12}{5}(\mu_4 - \sigma^4) + \dfrac{48}{5}\gamma\sigma^2\end{pmatrix} \quad \text{as } n \to \infty. \tag{4.3}$$

The proof is in Section 7. Observe that this result does not follow directly from (3.11)–(3.13), by combining the usual variance formula for a linear regression estimator such as (3.13) with the leading term $(4/n)C_0$ in $\operatorname{Cov}(Z) = \operatorname{Cov}(\eta)$. That this intuition does not work has two reasons: first the dimension of the sequence of $L \times 2$ matrices $A = A_n$ increases with n [as $L = L(n)$], that is, the estimates $\binom{\hat\sigma_1^2}{\hat\gamma_1}$ are obtained from a sequence of linear models with changing dimensions. Second, there are cancellation effects due to the fact that the sum of the weights in $\hat\gamma_1$ is zero; see the first two paragraphs of Section 7.1 and Lemma A1 in the Appendix.

It is an immediate consequence that both $\hat{\sigma}_1^2$, $\hat{\gamma}_1$ are consistent. The mean squared error rates of convergence are seen to be $O((1/n) + (L^4/n^4))$ for $\hat{\sigma}_1^2$ and $O((1/L) + (L^2/n^2))$ for $\hat{\gamma}_1$. On sets $S_{\gamma,\delta}$ with $\delta > 0$, our proofs show that these rates cannot be improved. To verify this, use $E(\eta_k) \sim \delta k^2/n^2$ [see (4.11) below and Lemma A1(vi)] to obtain

$$E(\hat{\gamma}_1) - \gamma = \delta(L/n)(1 + o(1)). \tag{4.4}$$

Therefore, the optimal mean squared error rates are n^{-1} for $\hat{\sigma}_1^2$ and $n^{-2/3}$ for $\hat{\gamma}_1$. The asymptotically optimal choice for L with respect to mean squared error according to (4.3), (4.4) is obtained by minimizing $[\delta(L/n)]^2 + \Lambda/L$ with respect to L, which yields

$$L^* = \left((\Lambda/2\delta^2)n^2\right)^{1/3}, \tag{4.5}$$

where $\Lambda = 12(\mu_4 - \sigma^4)/5 + 48\gamma\sigma^2/5$.

We next discuss the asymptotic limit distribution for the estimates $(\hat{\sigma}_1^2, \hat{\gamma}_1)^T$ [see (3.12)]. Note that throughout this paper the case that the errors $\varepsilon_i = \varepsilon_{i,n}$ come from a triangular array is included. This assumption is often more realistic to describe how the designs vary with n. For the following result, we need in addition to (A1)–(A3),

$$L^3/n^2 \to 0 \quad \text{as } n \to \infty. \tag{A4}$$

THEOREM 4.2. *Under the assumptions* (A1)–(A4), $E\varepsilon_1^3 = 0$ *and* $\mu_4 > \sigma^4$, *we have*

$$\begin{pmatrix} \sqrt{n}\,(\hat{\sigma}_1^2 - \sigma^2) \\ \sqrt{L}\,(\hat{\gamma}_1 - \gamma) \end{pmatrix} \to_d \mathscr{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, (\mu_4 - \sigma^4)\begin{pmatrix} 1 & 0 \\ 0 & \dfrac{12}{5} \end{pmatrix}\right). \tag{4.6}$$

For the proof see Section 7. We note that for the degenerate case $\mu_4 = \sigma^4$ (i.e., $\varepsilon_1 \sim \sigma(B_{1,1/2} - 1)$, where $B_{1,1/2}$ denotes a Bernoulli random variable with $p = 1/2$), one can also derive asymptotic normality with, however, different rates and variance. We will not pursue this case here further [for more details, see Dubowik (1996)]. We note that if we use the optimal L^* (4.5) which is of order $n^{2/3}$, a bias term must be included in the limit distribution. A similar effect is known to occur in curve estimation when optimal smoothing parameters are inserted.

One important application is the construction of an asymptotic level α test for the null hypothesis of no change,

$$H_o: \gamma = 0 \text{ versus } H_a: \gamma > 0,$$

when we assume that $\mu_4 - \sigma^4 > 0$. The test statistic

$$\Gamma = \frac{\sqrt{L}\,\hat{\gamma}_1}{\left((12/5)\left(\tilde{\mu}_4 - \tilde{\sigma}^4\right)\right)^{1/2}}, \tag{4.7}$$

targets a standardized version of $\hat{\gamma}_1$ (3.12). It is asymptotically normal by Slutsky's theorem, if we insert consistent estimators $\tilde{\mu}_4$, $\tilde{\sigma}^2$ for μ_4, σ^2. We suggest applying the following consistent estimators: $\tilde{\sigma}^2 = \hat{\sigma}_1^2$ as defined by (3.12) and $\tilde{\mu}_4 = 3\tilde{\sigma}^2$ in case the errors can be assumed to be approximately normally distributed.

If the errors are not normal and $\tilde{\mu}_4$ must be estimated independently of $\tilde{\sigma}^2$, asymptotically consistent estimators for μ_4 can still be found in the case where the regression function f has only a fixed finite number of jump discontinuities, irrespective of the sample size n. This restriction is not necessary for our other results, where the number of discontinuities may change with n and could diverge. One then may show that

$$\tilde{\mu}_4 = \frac{1}{2n} \sum_{j=1}^{n-1} (y_{j+1} - y_j)^4 - 3\tilde{\sigma}^4 \tag{4.8}$$

provides such an estimate with $\tilde{\mu}_4 \to_p \mu_4$ as $n \to \infty$. The reason is that the bias induced by the jumps is $O(n^{-1})$ if there are only finitely many jumps. In a finite sample situation, however, this bias may be nonnegligible.

Analogously, as an alternative estimate of σ^2, one could consider the difference-based estimate [compare Rice (1984)]

$$\tilde{\sigma}_D^2 = \frac{1}{2n} \sum_{j=1}^{n-1} (y_{j+1} - y_j)^2, \tag{4.9}$$

which is also consistent for σ^2 if there are only finitely many jumps. For both estimates (4.8) and (4.9), improved versions for practical applications are possible after locating the change-points as illustrated in the growth example below. One then would omit differences of y's in (4.8), (4.9) which cut across a jump.

Then, if $\gamma = 0$, we have asymptotically,

$$\Gamma \sim \mathcal{N}(0,1) \tag{4.10}$$

from which level α-tests for H_o: $\gamma = 0$ can be derived. We note that by a more refined analysis using Riemann sum approximation and Taylor expansion, we can write

$$Z_k = 2\sigma^2 + \frac{k}{n-L}\gamma + \left(\frac{k}{n-L}\right)^2 \delta + \tilde{\eta}_k, \tag{4.11}$$

where $\delta = (\int_0^1 g'^2(t)\,dt + 2\int_0^1 g'(t)\,dh(t))$ and $E(\tilde{\eta}_k) = o((\frac{L}{n})^2)$. This motivates a three-parameter asymptotic linear model with parameters β, γ,

δ. Similar arguments as before lead to the following least squares estimates for σ^2 and γ:

$$\hat{\sigma}_2^2 = \frac{3}{2L(L-1)(L-2)} \times \sum_{j=1}^{L} (3L^2 + 3L + 2 - 6(2L+1)j + 10j^2) Z_j,$$

$$\hat{\gamma}_2 = \frac{6(n-L)}{L(L^2-1)(L^2-4)} \times \sum_{j=1}^{L} (-3(L+1)(L+2)(2L+1) + 2(8L+11)(2L+1)j - 30(L+1)j^2) Z_j. \tag{4.12}$$

Plugging estimates (4.12) into model (4.11), we obtain

$$E\begin{pmatrix} \hat{\sigma}_2^2 \\ \hat{\gamma}_2 \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ \gamma^2 \end{pmatrix} + \begin{pmatrix} o(L^2/n^2) \\ o(L/n) \end{pmatrix}, \tag{4.13}$$

an improvement over (4.1). Furthermore, one can show that

$$\begin{pmatrix} \sqrt{n}\,(\hat{\sigma}_2^2 - \sigma^2) \\ \sqrt{L}\,(\hat{\gamma}_2 - \gamma) \end{pmatrix} \xrightarrow{d} \mathscr{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (\mu_4 - \sigma^4) & 0 \\ 0 & \frac{384}{35}(\mu_4 - \sigma^4) \end{pmatrix} \right), \tag{4.14}$$

provided that $L^3/n^2 = O(1)$ and $\gamma = 0$. A basic result for the proof of (4.13) and (4.14) is Lemma A5 in the Appendix and further calculations can be found in Dubowik (1996).

Note that the bias of $\hat{\gamma}_2$ is $o(L/n)$ rather than being asymptotically L/n as is the bias of $\hat{\gamma}_1$. This stems from the fact that while $\hat{\gamma}_2$ is based on the second-order Taylor expansion (4.11), $\hat{\gamma}_1$ is based on the first-order Taylor expansion (3.7). Therefore, we can let L increase up to order $L \sim n^{2/3}$ without incurring an asymptotic bias. In place of assumption (A4), we then only need

$$L^2/n^3 \le C < \infty; \tag{A5}$$

also compare the discussion after (4.6).

Note that (4.14) implies that the asymptotic variances of $\hat{\gamma}_2$ and of $\hat{\gamma}_1$ (obtained from the two-parameter or asymptotic simple linear regression model) are related by $\mathrm{var}(\hat{\gamma}_2)/\mathrm{var}(\hat{\gamma}_1) = (5/12)(384/35) = 32/7$. Thus the estimate $\hat{\gamma}_2$ is more variable when compared to $\hat{\gamma}_1$. Still we recommend using this asymptotic quadratic model in particular for the cases where $\delta > 0$, that is, where the smooth part of the function cannot be neglected. In such cases, for sizable $\delta > 0$, $\hat{\gamma}_1$ is contaminated by nonnegligible bias in particular for large L, whereas $\hat{\gamma}_2$ is less affected by such bias; compare (4.13). See Section 5 for simulation comparisons of $\hat{\gamma}_1$ and $\hat{\gamma}_2$.

5. Finite sample behavior. The proposed tests and estimates depend on the choice of the maximal span size L, which assumes a role similar to that of a smoothing parameter in smoothing methods. The asymptotic variances of these estimates become smaller if L is increased, according to Theorem 4.1. However, large choices of L may lead to violation of (A1) whenever the regression function has several jumps. Conditions (A4) and (A5) effectively place an upper bound on L. If L is too large, this may lead to biases in $\hat{\gamma}$. Thus L must be chosen to negotiate a compromise between variance and bias; theoretically the mean squared error provided by Theorem 4.1 can be minimized by using L^* (4.5). However, this is not feasible in practice and a data-based "plateau" method is proposed in Section 5.2. We first report the results of simulation studies regarding the behavior of the estimates of γ and σ^2 when varying L.

5.1. *Monte Carlo results.* In order to study the behavior of estimates $\hat{\gamma}_1$, $\hat{\sigma}_1^2$ (3.12) obtained by fitting the asymptotic simple linear regression model to the data $(k/n, Z_k)$ and of estimates $\hat{\gamma}_2$, $\hat{\sigma}_2^2$ (4.12) obtained by fitting the asymptotic quadratic model, we looked at various simulated settings. For the summands $g \in S_C$ and $h \in S_D$ of the function $f \in S_{\gamma,\delta}$, where $f = g + h$, we chose the following functions, all on support $[0, 1]$. For the step function h,

$$h(x) = c_0 1_{[0,0.25]} + c_1 1_{[0.25,0.5]} + c_2 1_{[0.5,1]}. \tag{5.1}$$

The constants c_0, c_1, c_2 in h were chosen to achieve the following values for the sum of squared jump sizes $\gamma = (c_1 - c_0)^2 + (c_2 - c_1)^2$: (1) $\gamma = 0$, with the choice $c_0 = c_1 = c_2 = 0$ (no jump); (2) $\gamma = 1.0$, choosing $c_0 = 0$, $c_1 = c_2 = 1$ (one jump) and (3) $\gamma = 3.25$, choosing $c_0 = 0$, $c_1 = 1$, $c_2 = -0.5$ (two jumps). For the smooth part g, the three functions $g_1(x) \equiv 0$, $g_2(x) \equiv x$ and $g_3(x) \equiv 4x(1-x)$ were investigated. The error variance in the basic model was chosen to be either $\sigma^2 = 0.25$ or $\sigma^2 = 1.0$, and the errors ε_i were generated as normal random variables.

By varying the parameters, we obtained various scatterplots of data generated from model (1.6). A variety of examples are shown in Figures 3 and 4 for $n = 100$ and $\sigma^2 = 0.25$ and in Figures 5 and 6 for $n = 1000$ and $\sigma^2 = 1.0$. The upper two panels display the scatterplots, and the lower panels the estimates $\hat{\gamma}_1$, respectively, $\hat{\gamma}_2$ of γ (solid lines) as well as $\hat{\sigma}_1^2$, respectively, $\hat{\sigma}_2^2$ of σ^2 (dashed lines) in dependency on the auxiliary parameter L. The left panels are for the case of no jumps, $\gamma = 0$, and the right panels for $\gamma = 3.25$ (two jumps).

The main findings are as follows: for all examples, the estimates $\hat{\sigma}_1^2$, respectively, $\hat{\sigma}_2^2$ are quite stable, that is, do not vary much with L and are also quite accurate. This indicates that the estimation of σ^2 is relatively easy and does not require any sophisticated choice of L. Some exceptions to this rule do exist, though, as exemplified in Figure 4, right panels, where for the case $n = 100$, $\gamma = 3.25$, $g \equiv g_3 \equiv 4x(1-x)$ and $\sigma^2 = 0.25$, the estimates $\hat{\sigma}_2^2$ turn negative for larger values of L.

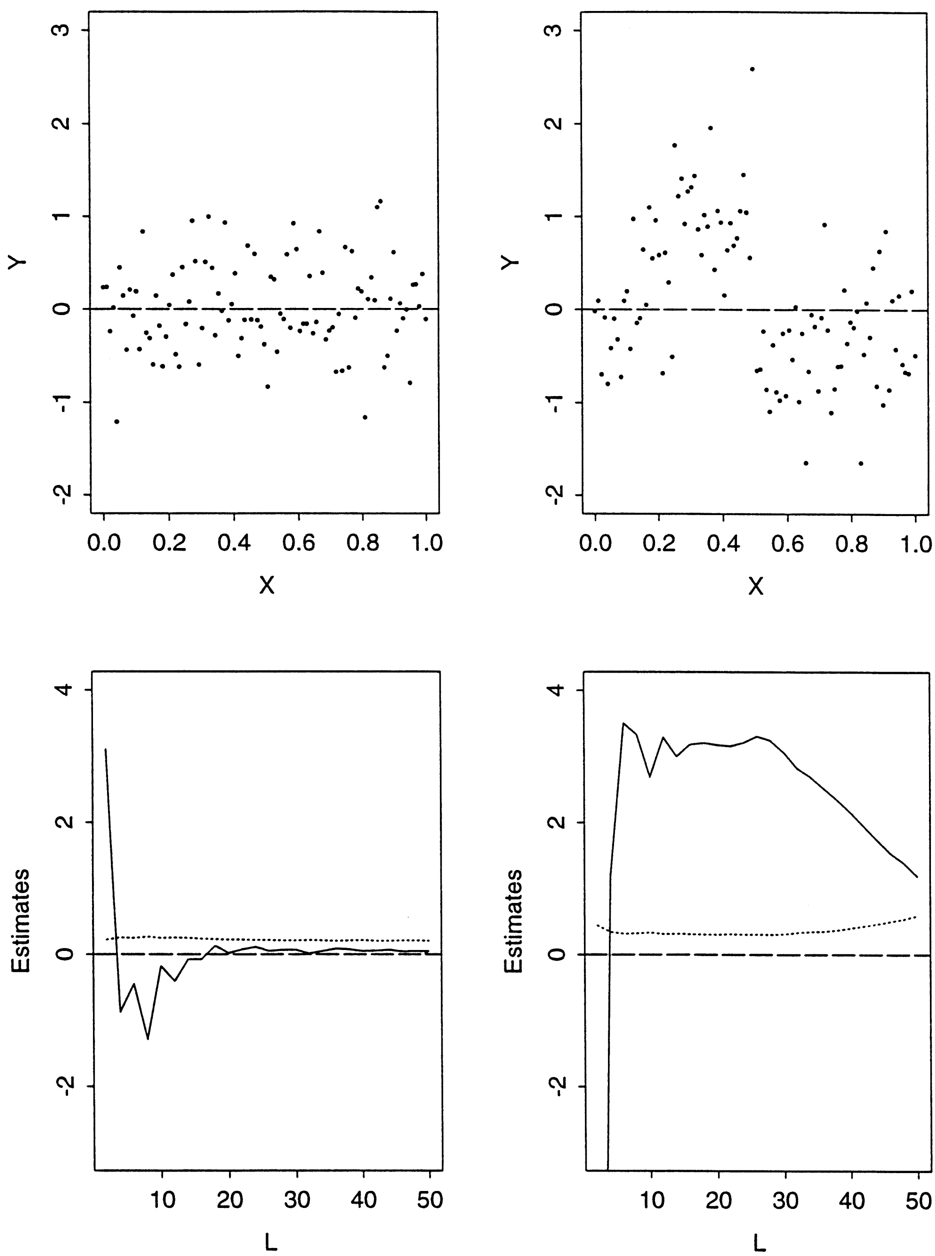

FIG. 3. *Simulated examples. The upper panels show scatterplots for* $n = 100$ *data, generated with smooth functions* $g_1 \equiv 0$ *and error variance* $\sigma^2 = 0.25$. *The lower panels display estimates* $\hat{\gamma}_1$ *(solid) and* $\hat{\sigma}_1^2$ *(dashed)* (3.12), *obtained by fitting the asymptotic simple linear model, in dependency on L. The two left panels are for the case of no jumps,* $\gamma = 0$, *and the two right panels are for the case* $\gamma = 3.25$, *with superimposed step function h* (5.1).

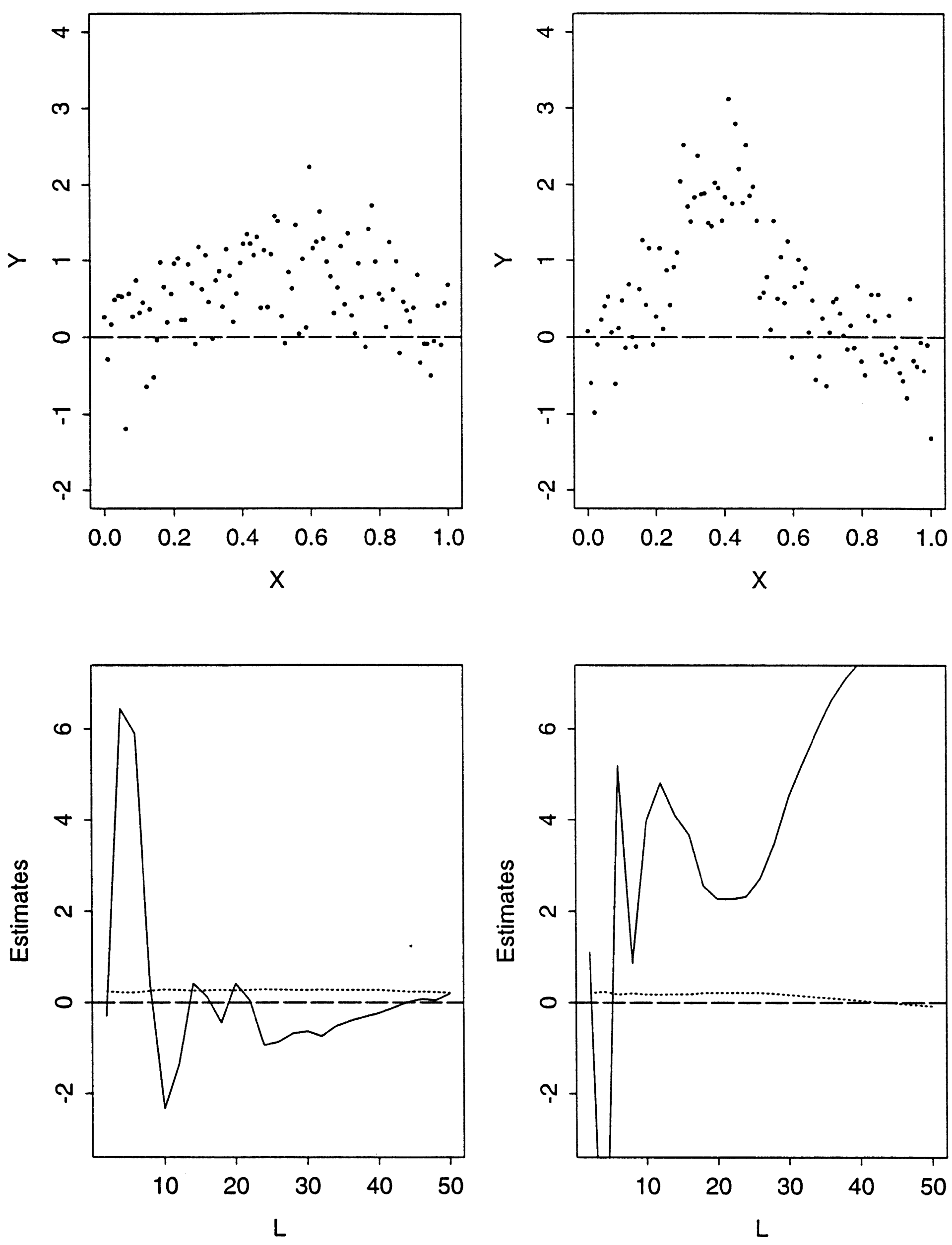

FIG. 4. *Simulated examples. The upper panels show scatterplots for* $n = 100$ *data, generated with smooth functions* $g_3 \equiv 4x(1-x)$ *and error variance* $\sigma^2 = 0.25$. *The lower panels display estimates* $\hat{\gamma}_2$ *(solid) and* $\hat{\sigma}_2^2$ *(dashed)* (4.12), *obtained by fitting the asymptotic quadratic model, in dependency on L. The two left panels are for the case of no jumps,* $\gamma = 0$, *and the two right panels are for the case* $\gamma = 3.25$, *with superimposed step function h* (5.1).

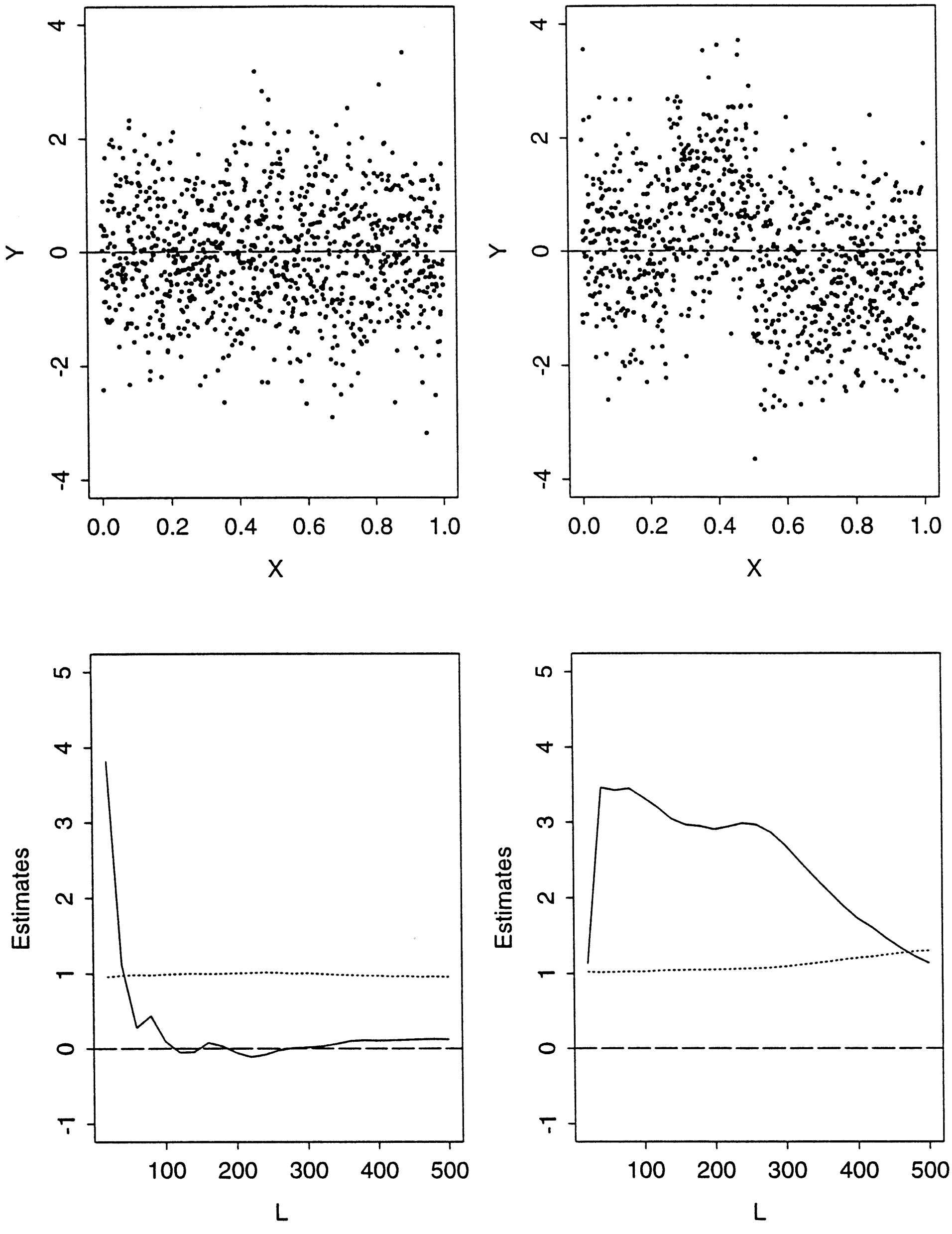

FIG. 5. *Simulated examples. The upper panels show scatterplots for n = 1000 data, generated with smooth functions $g_1 \equiv 0$ and error variance σ^2 = 1.0. The lower panels display estimates $\hat{\gamma}_1$ (solid) and $\hat{\sigma}_1^2$ (dashed)* (3.12), *obtained by fitting the asymptotic simple linear model, in dependency on L. The two left panels are for the case of no jumps, γ = 0, and the two right panels are for the case γ = 3.25, with superimposed step function h* (5.1).

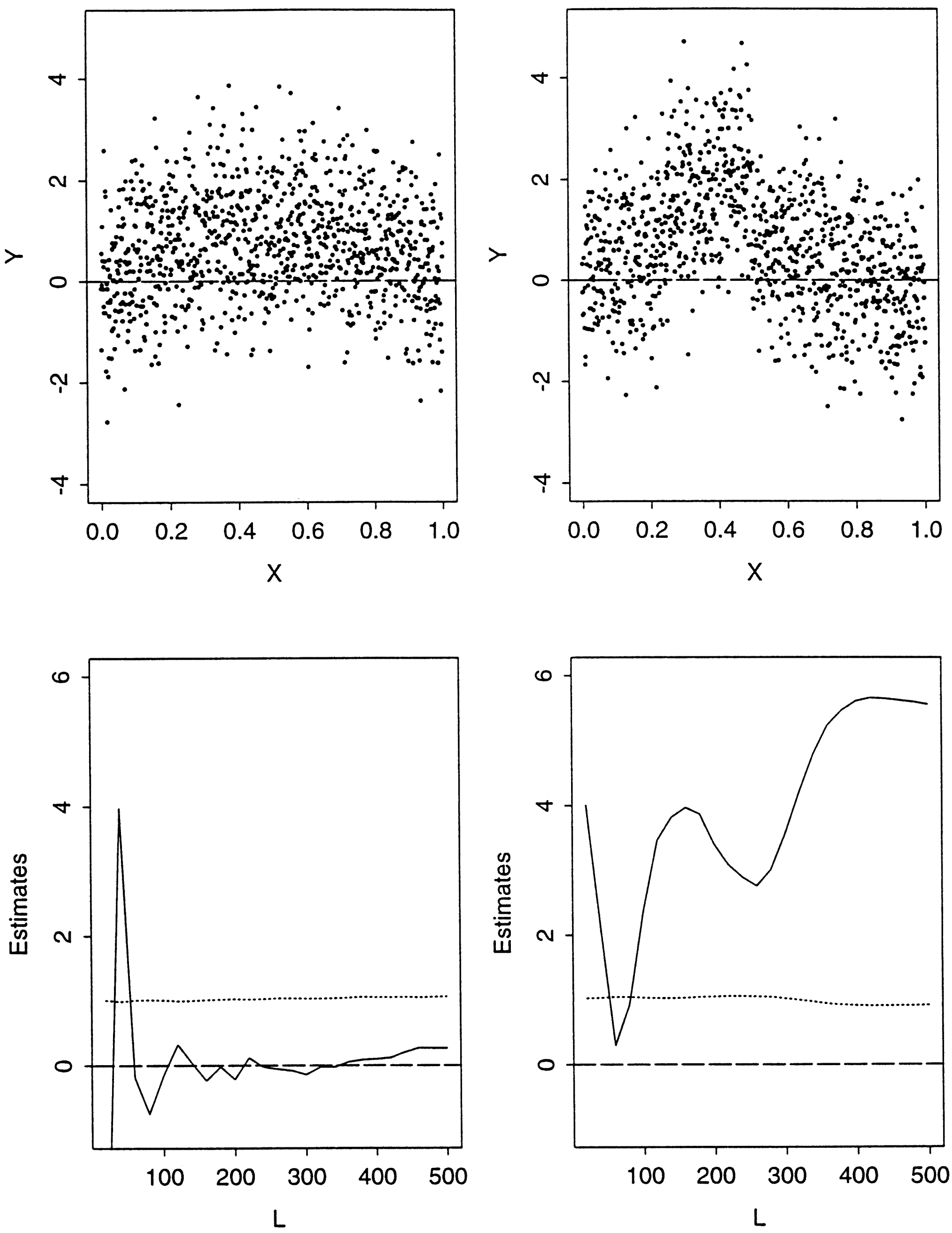

FIG. 6. *Simulated examples. The upper panels show scatterplots for $n = 1000$ data, generated with smooth functions $g_3 \equiv 4x(1-x)$ and error variance $\sigma^2 = 1.0$. The lower panels display estimates $\hat{\gamma}_2$ (solid) and $\hat{\sigma}_2^2$ (dashed) (4.12), obtained by fitting the asymptotic quadratic model, in dependency on L. The two left panels are for the case of no jumps, $\gamma = 0$, and the two right panels are for the case $\gamma = 3.25$, with superimposed step function h (5.1).*

The estimation of γ is seen to be far less stable, especially for the cases with sample size $n = 100$ (Figures 3 and 4). These and other examples indicate that for small values of L, the functions $\hat{\gamma}(L)$ typically display oscillatory behavior whereas for very large values of L, the estimates are trailing off, and become biased. In nearly all situations, a value for L which leads to reasonably good estimates lies in between these two extremes.

These findings are corroborated when looking at mean estimates for $\hat{\gamma}_2$ and $\hat{\sigma}_2^2$ obtained from 500 simulations as shown in Figure 7. The left panels are always for $\gamma = 0$, the right panels for $\gamma = 1$, and the estimates for γ in dependency on L are solid curves, while the estimates for σ^2 are dashed curves. The top two panels in (a) are for $n = 100$; all the others are for $n = 1000$. These average estimates are seen to be well on target irrespective of the value of L for σ^2 and in middle ranges of L for γ as well. Similar observations can be made for the mean squared error of $\hat{\gamma}_1$, $\hat{\gamma}_2$ as functions of L (results not reported).

5.2. *Choice of L.* The simulations indicate that the choice of the span size L is critical for the estimation of γ in most cases, while it does not matter much for the estimation of σ^2. For the applications, an empirical, data-dependent method for the choice of L is needed. We propose the *plateau method* which is inspired by the simulation results, in particular, Figure 7. The best results are obtained in a range of L's where the estimates $\hat{\gamma}(L)$ are relatively stable. This "plateau" is reached in most cases after an initial period of rapid oscillations for small L and prior to a trend which sets in for large values of L and which is biasing the estimates upward from the target values.

The plateau method can be implemented in a variety of ways. We found the following version to be particularly successful: First define a function

$$\Xi(L) = \sum_{i=L-L_0}^{L+L_0} (i - L)\hat{\gamma}(i),$$

where $L_0 = \max([n/50], 2)$ for a sample size n, and $\hat{\gamma}(i)$ stands for $\hat{\gamma}_1(i)$ or $\hat{\gamma}_2(i)$, choosing i as the span size. This can be interpreted as a derivative estimate of $\hat{\gamma}(\cdot)$ at L, using a window smoother with window $[L - L_0, L + L_0]$, and ignoring normalizing constants. Then obtain

$$\hat{L} = \min\{L: \Xi(L - i) > 0 \text{ for } 0 \le i \le L_0\}. \tag{5.2}$$

Similar criteria which are also useful but in our simulations came out somewhat inferior are

$$\hat{L} = \min\{L: |\Xi(L - i)| > 0 \text{ for } 0 \le i \le L_0\} \tag{5.3}$$

and

$$\hat{L} = \arg\min_L \left\{ \frac{1}{2L_0 + 1} \sum_{i=L-L_0}^{L+L_0} \hat{\gamma}(i)^2 - \left[\frac{1}{2L_0 + 1} \sum_{i=L-L_0}^{L+L_0} \hat{\gamma}(i) \right]^2 \right\}, \tag{5.4}$$

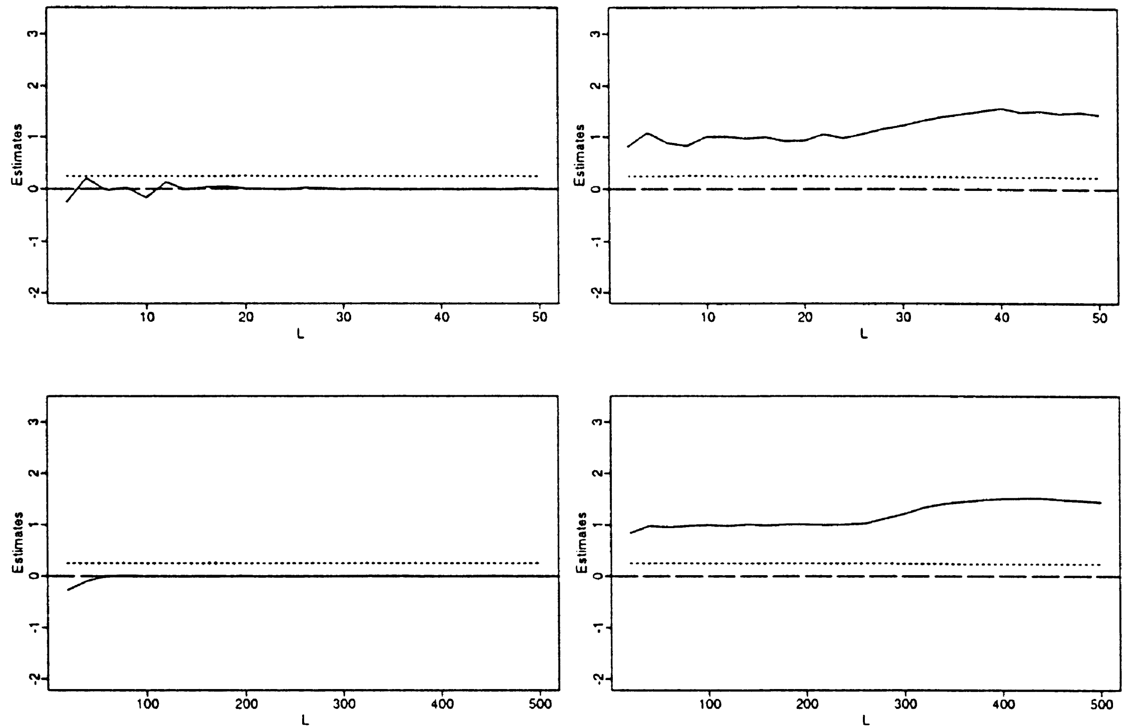

FIG. 7. *Average values of estimates* $\hat{\gamma}_2$ *(solid) and* $\hat{\sigma}_2^2$ *(dashed)* (4.12), *obtained by fitting the asymptotic quadratic model, in dependency on L. Curves* $\hat{\gamma}_2(L)$ *and* $\hat{\sigma}_2^2(L)$ *are based on averages over* 500 *Monte Carlo runs. The parameters under which the curves were obtained are as follows for the various panels*: (*a*) *Upper left*: $n = 100$, $g \equiv g_1 \equiv 0$, $\sigma^2 = 0.25$, $\gamma = 0$; *Upper right*: $n = 100$, $g \equiv g_1 \equiv 0$, $\sigma^2 = 0.25$, $\gamma = 1$; *Lower left*: $n = 1000$, $g \equiv g_1 \equiv 0$, $\sigma^2 = 0.25$, $\gamma = 0$; *Lower right*: $n = 1000$, $g \equiv g_1 \equiv 0$, $\sigma^2 = 0.25$, $\gamma = 1$.

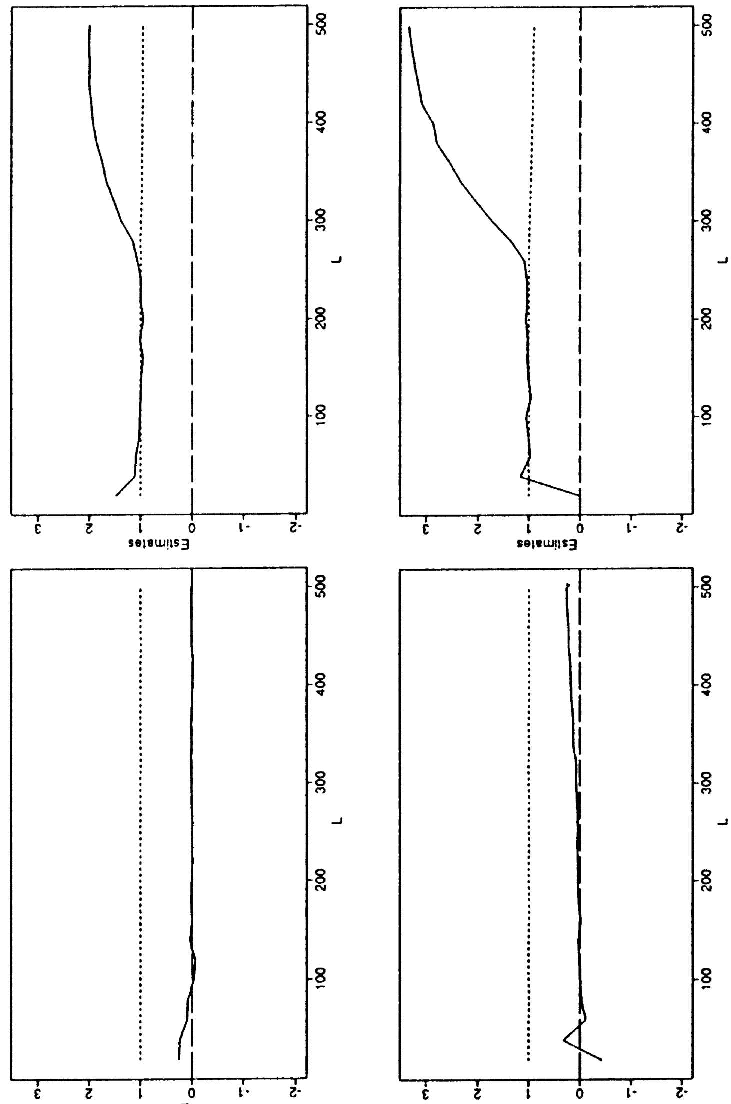

FIG. 7. (*Continued*). (*b*) *Upper left*: $n = 1000$, $g \equiv g_2 \equiv x$, $\sigma^2 = 1$, $\gamma = 0$; *Upper right*: $n = 1000$, $g \equiv g_2 \equiv x$, $\sigma^2 = 1$, $\gamma = 1$; *Lower left*: $n = 1000$: $g \equiv g_3 \equiv 4x(1-x)$, $\sigma^2 = 1$, $\gamma = 0$; *Lower right*: $n = 1000$: $g \equiv g_3 \equiv 4x(1-x)$, $\sigma^2 = 1$, $\gamma = 1$.

TABLE 1
*Simulation results for MSE's of estimates $\hat{\gamma}_1$, $\hat{\gamma}_2$, $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ **

Case		**Selected $\hat{L}$ (with std. dev.) for**		**MSE $\times 10^3$ for**		**MSE $\times 10^4$ for**	
Parameter γ	**Function g**	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
0	1	196(70)	187(52)	2.45	17.9	1.58	1.51
0	2	156(22)	187(54)	16.9	16.3	1.37	1.72
0	3	157(45)	177(41)	242	19.2	2.10	1.55
1.0	1	235(128)	180(41)	65.4	75.7	9.61	18.3
1.0	2	148(20)	180(40)	195	74.8	2.04	1.85
1.0	3	145(19)	180(39)	1172	78.2	3.93	1.68
3.25	1	223(121)	185(42)	688	221	88.6	2.24
3.25	2	252(137)	183(41)	1682	210	2.00	2.06
3.25	3	146(63)	182(40)	1225	207	4.03	2.06

*In various situations with auxiliary parameter L selected by the plateau method (5.2).
Error variance $\sigma^2 = 0.25$, sample size $n = 1000$. Results are for 1000 simulations. Parameter γ is the sum of squared jump sizes, and the smooth part of the regression function was selected as $g_1 \equiv 0$, $g_2 \equiv x$ and $g_3 \equiv 4x(1-x)$.
Besides MSE's, the table also contains means and standard deviations for the selected values $\hat{L}$ (5.2) for both $\hat{\gamma}_1$ and $\hat{\gamma}_2$.

the latter choosing L in such a way as to minimize the variance of $\hat{\gamma}(\cdot)$ in the local window $[L - L_0, L + L_0]$. The idea behind criteria (5.2), (5.3) is that (1) $\hat{L}$ will not be chosen from within the early oscillation period (too small L's), as there the derivative will change sign over a span of size L_0 and therefore Ξ will not satisfy $\Xi(L-i) > 0$ for $0 < i \le L_0$ for L's chosen from a region where oscillations occur; (2) $\hat{L}$ will also not be chosen from within the right-hand region where a monotone (upward or downward) trend and therefore bias occurs (too large L's). Instead, it is likely that $\hat{L}$ will be chosen from somewhere in the area of the plateau, where flat but sustained upward trends occur fairly regularly.

This approach worked well in simulations; the resulting fully data-based estimates for γ and σ^2 yielded mean squared errors as listed in Table 1. It is clear from this table that data-based estimation of the error variance σ^2 in the presence of jump discontinuities works well with this method in almost all cases. Regarding the estimation of γ, a more mixed picture emerges. As a rule, estimates $\hat{\gamma}_2$ (4.12) based on the asymptotic quadratic model have smaller bias but in some cases substantially larger variance than estimates $\hat{\gamma}_1$ (3.12), which are based on the asymptotic simple linear model. The overall results for mean squared errors show that the estimates $\hat{\sigma}_2^2$ for the error variance σ^2 are mostly better than estimates $\hat{\sigma}_1^2$. Also, with the exception of the cases where $g \equiv g_1 \equiv 0$, $\gamma = 0$ and $g \equiv g_1 \equiv 0$, $\gamma = 1.0$, $\hat{\gamma}_2$ is better than $\hat{\gamma}_1$ in terms of MSE. For the cases where n was small or σ^2 was large, the comparison between the two estimators was not as clear-cut. Nevertheless,

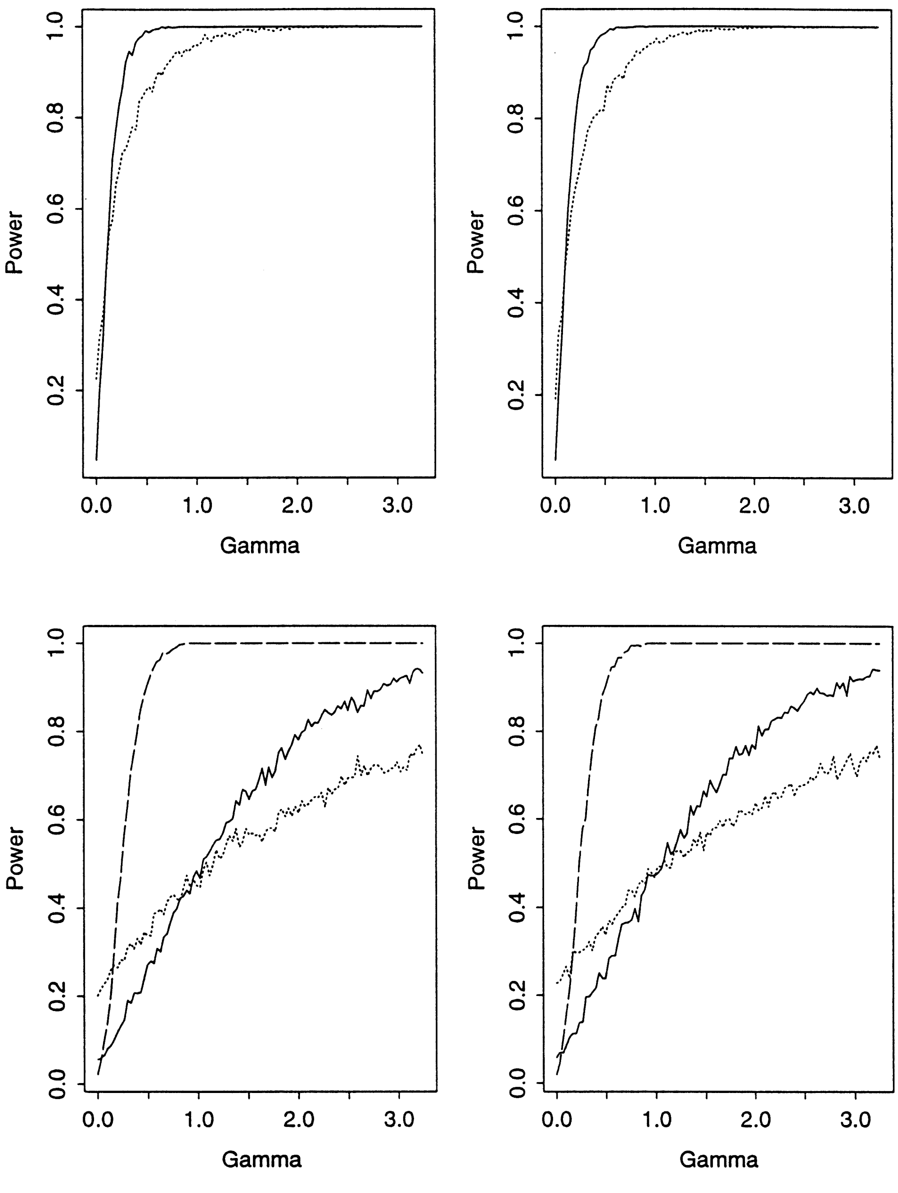

FIG. 8. *Empirical power for the asymptotic level α test H_0: $\gamma = 0$ versus H_a: $\gamma > 0$, $\alpha = 0.1$, computed from 1000 simulations, in dependency on $\gamma > 0$ and based on estimates $\hat{\gamma}_2$, $\hat{\sigma}_2^2$ (4.12). Sample sizes $n = 30$ (short dashed), $n = 100$ (solid) and $n = 1000$ (long dashed). The parameters for the panels are upper left: $g_1 \equiv 0$, $\sigma^2 = 0.025$, upper right: $g_2 \equiv x$, $\sigma^2 = 0.025$, lower left: $g_1 \equiv 0$, $\sigma^2 = 0.25$, lower right: $g_2 \equiv x$, $\sigma^2 = 0.25$.*

we recommend using the estimates $\hat{\gamma}_2$ and $\hat{\sigma}_2^2$. We also note the remarkable stability of the average chosen L by means of the plateau method for $\hat{\gamma}_2$.

It is also of interest to investigate the empirical power of the test for H_0: $\gamma = 0$ as a function of γ, using the plateau method of choosing L and thus a completely data-driven procedure. The results are shown in Figure 8 for asymptotic level α tests with $\alpha = 0.1$, $n = 30$, 100 and 1000 and four different situations, including different smooth functions and error variances. These results indicate that the test has good power for $n = 1000$ but relatively low power when $n = 30$ or $n = 100$ and the error variance is high.

The code implementing estimators (3.12) and (4.12) and the plateau method (5.2) has been written in Fortran and is available from the authors, along with the infant growth data discussed in the next section.

6. Further analysis of infant growth data. A preliminary analysis with nonparametric regression methods incorporating jump discontinuities was presented in Section 2. The methods described there provide a variety of fits, but cannot resolve the model selection problem, namely whether a smooth or rather a discontinuous model is appropriate for these data. This model selection problem is at the heart of the scientific controversy on the existence of saltatory growth. The estimates $\hat{\gamma}_2(L)$ and $\hat{\sigma}_2^2(L)$, using the asymptotic quadratic model for these data, are shown in Figure 9; we note that the asymptotic simple linear model did not provide reasonable estimates for this case.

Figure 9 shows a weakly expressed plateau. We choose $L = 10$ and obtain estimates $\hat{\gamma}_2 = 0.70$ and $\hat{\sigma}_2^2 = 0.024$. Assuming normal errors, and under the null hypothesis $\gamma = 0$, the corresponding estimated variances according to (4.14) are $\widehat{\mathrm{var}}(\hat{\gamma}_2) = 1.2 \cdot 10^{-3}$, $\widehat{\mathrm{var}}(\hat{\sigma}_2^2) = 1.5 \cdot 10^{-4}$. This provides for a highly significant z-value of $z = 20.08$ in the test statistic, and thus the smooth model would be rejected in favor of a model containing discontinuities. These results have to be viewed with a grain of salt, as they rely on asymptotic approximations and the normality and independence assumptions for the errors. One needs to keep in mind that the sample size of $n = 30$ available here is quite small.

Nevertheless, this analysis in conjunction with the findings in Section 2 points toward the existence of periods of fast growth which occurs in a relatively short time. The fast growth can be modelled as a discontinuity when the analysis is based on daily measurements. This does not mean that a mathematical jump discontinuity exists in reality, and a much finer grid of measurements is likely to resolve the apparent discontinuity into a short period of fast growth. Furthermore, one cannot draw broad conclusions based on limited data from one subject.

The smooth and discontinuous fits in Figure 2 have been monotonized as growth curves of course must be monotone; see Appendix A.2. Sum of squared jump sizes as well as error variances were calculated for both unrestricted (not necessarily monotone) as well as monotonized versions. Error variances were estimated by the average squared residuals according to the formula

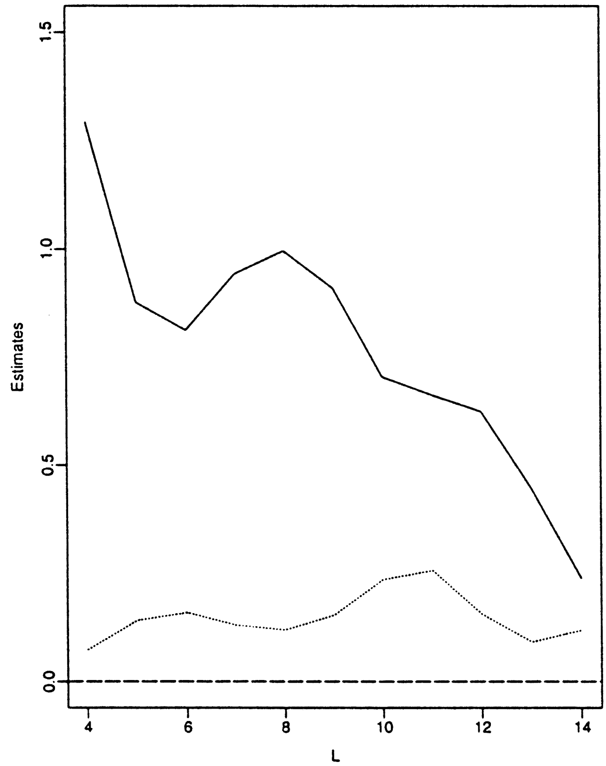

FIG. 9. *Estimates $\hat{\gamma}_2$ (solid) and $10\hat{\sigma}_2^2$ (dashed) for the crown-heel lengths data as a function of L.*

(A.2) given in Appendix A.2. These values and also individual jump sizes for nonmonotone and monotone fits are listed in Table 2.

Since the simulations have shown that even for modest sample sizes the error variance estimate $\hat{\sigma}_2^2$ is quite stable and has low variance, we can compare $\hat{\sigma}_2^2 = 0.024$ with the error variances produced from the residuals of the various models. As it happens, $\hat{\sigma}_2^2$ falls in between the values obtained for the smooth fit $\hat{\sigma}^2 = 0.030$ and for the smooth fit with one discontinuity ($\hat{\sigma}^2 = 0.019$). The estimate $\hat{\gamma}_2 = 0.70$ is somewhat larger than the estimated sum of squared jump sizes with three discontinuities ($\hat{\gamma} = 0.62$). We cautiously conclude from this and the test results rejecting H_0: $\gamma = 0$ that these data are better represented by models containing jump discontinuities rather than models assuming a smooth or linear function. We note that Figure 8 indicates that for $\sigma^2 = 0.025$ and $n = 30$ the power of the proposed test is reasonable.

TABLE 2

*Estimated error variances $\hat{\sigma}^2$ and jump sizes $\hat{\Delta}$ for the crown-heel lengths data**

Fit	$\hat{\sigma}^2(\times 10^2)$ Nonmonotone	$\hat{\sigma}^2(\times 10^2)$ Monotone	$\hat{\Delta}$ Nonmonotone at 72.5	77.5	86.5	$\hat{\Delta}$ Monotone at 72.5	77.5	86.5	$\hat{\gamma}$ nonmonotone	$\hat{\gamma}$ monotone
Linear	4.2	4.2	0	0	0	0	0	0	0	0
Smooth	3.0	3.0	0	0	0	0	0	0	0	0
Smooth with one jump	1.9	1.9	0	0	0.70	0	0	0.57	0.49	0.32
Smooth with two jumps	1.8	1.8	0	0.28	0.70	0	0.28	0.57	0.57	0.40
Smooth with three jumps	1.6	1.7	0.22	0.28	0.70	0.21	0.28	0.57	0.62	0.44
Asymptotic quadratic model $\hat{\gamma}_2$, $\hat{\sigma}_2^2$	2.4								0.70	

*Based on various monotone and nonmonotone fits.
Here, $\hat{\gamma}$ denotes the estimated sum of squared jump sizes.

7. Proofs of the main results.

7.1. *Proof of Theorem* 4.1. In order to calculate the first two moments of our estimator, we define for $1 \le j \le L$ and $\lambda, \rho \in \mathbf{R}$ the quantities

$$\alpha_j = \frac{2}{L(L-1)}(2L+1-3j),$$

(7.1)

$$\beta_j = \frac{6(n-L)}{L(L^2-1)}(2j-(L+1)), \qquad w_j = \lambda\alpha_j + \rho\beta_j,$$

and obtain

$$\hat{\beta} = \sum_{j=1}^{L} \alpha_j Z_j, \qquad \hat{\gamma}_1 = \sum_{j=1}^{L} \beta_j Z_j.$$

Furthermore we find from (3.3),

(7.2)

$$\begin{aligned}\eta_k &= \frac{1}{n-L}\sum_{j=1}^{n-L}\left\{(\varepsilon_{j+k}-\varepsilon_j)^2 - 2\sigma^2\right\}\\ &\quad + \frac{2}{n-L}\sum_{j=1}^{n-L}\left(f\left(\frac{j+k}{n}\right) - f\left(\frac{j}{n}\right)\right)(\varepsilon_{j+k}-\varepsilon_j)\\ &\quad + \frac{1}{n-L}\sum_{j=1}^{n-L}\left(g\left(\frac{j+k}{n}\right) - g\left(\frac{j}{n}\right)\right)^2\\ &\quad + \frac{2}{n-L}\sum_{j=1}^{n-L}\left(g\left(\frac{j+k}{n}\right) - g\left(\frac{j}{n}\right)\right)\left(h\left(\frac{j+k}{n}\right) - h\left(\frac{j}{n}\right)\right),\end{aligned}$$

where g is the smooth and h the nonsmooth part of f. Obviously, $E(\eta_k) = O(k^2/n^2)$, which is (3.4).

In order to find the expectations of $\hat{\beta}$ and $\hat{\gamma}_1$, observe that by Lemma A1 below, $\sum \alpha_k = 1$, $\sum \alpha_k k = 0$ and $\sum \beta_k = 0$, $\sum \beta_k k = n - L$, $\sum |\alpha_k| k^2 = O(L^2)$, $\sum |\beta_k| k^2 = O(Ln)$. Hence

$$E(\hat{\beta}) = \beta + O\left(\sum_{k=1}^{L} |\alpha_k E(\eta_k)| \right) = \beta + O(L^2/n^2),$$

$$E(\hat{\gamma}_1) = \gamma + O\left(\sum_{k=1}^{L} |\beta_k E(\eta_k)| \right) = \gamma + O(L/n),$$

which had to be shown.

To obtain the covariance matrix, we calculate for arbitrary $\lambda, \rho \in \mathbf{R}$ the variance of $(\lambda\hat{\beta} + \rho\hat{\gamma}_1)$,

$$S^2 := \operatorname{var}\left(\lambda \sum_{j=1}^{L} \alpha_j Z_j + \rho \sum_{j=1}^{L} \beta_j Z_j \right) = \operatorname{var}\left(\sum_{k=1}^{L} w_k Z_k \right),$$

and obtain

$$S^2 = \operatorname{var}\left(\sum_{k=1}^{L} w_k \eta_k \right) = \sum_{k,l=1}^{L} w_k w_l \operatorname{cov}(\eta_k, \eta_l).$$

Furthermore, for $1 \le k, l \le L$,

$$\begin{aligned}
&\operatorname{cov}(\eta_k, \eta_l)\\
&\quad = \frac{1}{(n-L)^2} \sum_{i,j=1}^{n-L} \Big[(\mu_4 - \sigma^4)(\sigma_{j+k,i+l} + \delta_{j,i+l} + \delta_{j+k,i} + \delta_{j,i}) \\
&\qquad\qquad + 4\sigma^4 \delta_{i,j} \delta_{k,l} \Big] \\
&\qquad + \frac{4\sigma^2}{(n-L)^2} \sum_{i,j=1}^{n-L} \left(f\left(\frac{j+k}{n}\right) - f\left(\frac{j}{n}\right) \right) \left(f\left(\frac{i+l}{n}\right) - f\left(\frac{i}{n}\right) \right) \\
&\qquad\qquad \times (\delta_{j+k,i+l} - \delta_{j+k,i} - \delta_{j,i+l} + \delta_{i,j}) \\
&\quad = \frac{4\sigma^4}{n-L} \delta_{k,l} + \frac{\mu_4 - \sigma^4}{(n-L)^2} [4(n-L) - (k + l + |k-l|)] \\
&\qquad + \frac{8}{(n-L)^2} \gamma\sigma^2 \min\{k, l\} + O\left(\frac{L^{3/2}}{n^{5/2}} \right).
\end{aligned}$$

Here we used the fact that $f = g + h$ and, observing (A1),

$$\begin{aligned}
&\sum_{i,j=1}^{n-L} \left(h\left(\frac{j+k}{n}\right) - h\left(\frac{j}{n}\right) \right) \left(h\left(\frac{i+l}{n}\right) - h\left(\frac{i}{n}\right) \right) \\
&\qquad \times (\delta_{j+k,i+l} - \delta_{j+k,i} - \delta_{j,i+l} + \delta_{j,i}) = 2\gamma \min\{k, l\}.
\end{aligned}$$

For instance,

$$\begin{aligned}
&\sum_{j=1}^{n-L}\left(h\left(\frac{j+k}{n}\right)-h\left(\frac{j}{n}\right)\right)\left(g\left(\frac{j+k}{n}\right)-g\left(\frac{j}{n}\right)\right)\\
&\quad\le c\left(\sum_{j=1}^{n-L}\left(h\left(\frac{j+k}{n}\right)-h\left(\frac{j}{n}\right)\right)^2\sum_{j=1}^{n-L}\left(g'\left(\frac{j}{n}\right)\right)^2\left(\frac{k}{n}\right)^2\right)^{1/2}\\
&\quad= O\left(\sqrt{\frac{L^3}{n}}\right),\qquad 1\le k\le L,
\end{aligned}$$

and similarly for the other terms.

Thus we obtain (3.5) and

$$\begin{aligned}
S^2 &= \sum_{k,l=1}^{L} w_k w_l \operatorname{cov}(\eta_k,\eta_l)\\
&= \frac{4\sigma^4}{n-L}\sum_{j=1}^{L} w_j^2 + \frac{4(\mu_4-\sigma^4)}{n-L}\left(\sum_{j=1}^{L} w_j\right)^2\\
&\quad -\frac{\mu_4-\sigma^4}{(n-L)^2}\left\{4\sum_{j=1}^{L} w_j\sum_{k=j+1}^{L} k w_k + 2\sum_{j=1}^{L} w_j^2 j\right\}\\
&\quad +\frac{16\gamma}{(n-L)^2}\sum_{k=1}^{L} w_k\sum_{l=1}^{k} w_l l - \frac{8\gamma}{(n-L)^2}\sum_{k=1}^{L} w_k^2 k + O\left(\frac{L^{3/2}}{n^{5/2}}\right)\left(\sum_{k=1}^{L}|w_k|\right)^2.
\end{aligned}$$

Now using Lemma A1, we conclude that

$$\begin{aligned}
S^2 &= \frac{4\lambda^2(\mu_4-\sigma^4)}{n}(1+o(1))\\
&\quad + 12\rho^2\left\{\frac{4\sigma^4 n}{L^3}+\frac{1}{5}\frac{(\mu_4-\sigma^4)}{L}+\frac{4}{5}\frac{\gamma\sigma^2}{L}\right\}(1+o(1)).
\end{aligned}$$

From $\hat\sigma_1^2 = \frac{1}{2}\hat\beta$ and (A3) we find that

$$\begin{aligned}
\operatorname{var}\left(\lambda\hat\sigma_1^2+\rho\hat\gamma_1\right) &= (\mu_4-\sigma^4)\left(\frac{\lambda^2}{n}(1+o(1)) + \frac{12}{5}\frac{\rho^2}{L}(1+o(1))\right)\\
&\quad + \frac{48\rho^2\gamma\sigma^2}{5L}(1+o(1))
\end{aligned}$$

for any $\lambda, \rho \in \mathbf{R}$, which yields Theorem 4.1. □

7.2. *Proof of Theorem* 4.2. We start by showing the asymptotic normality of $(\hat\sigma_1^2, \hat\gamma_1)^T$ for the case $\gamma = 0$ in detail. This will follow from the asymptotic normality of $(\hat\beta, \hat\gamma_1)^T = B^{-1}(\hat\sigma_1^2, \hat\gamma_1)^T$ where

$$B = \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}.$$

By the Cramér-Wold device it is sufficient to show for any pair $(\lambda, \rho) \in \mathbf{R}^2$ that

$$X_n = \lambda\sqrt{n-L}\left(\hat{\beta} - E(\hat{\beta})\right) + \rho\sqrt{L}\left(\hat{\gamma}_1 - E(\hat{\gamma}_1)\right)$$

is asymptotically normal. We will show that

$$X_n \to_d \mathcal{N}\left(0, 4\lambda^2(\mu_4 - \sigma^4) + \frac{12}{5}\rho^2(\mu_4 - \sigma^4)\right), \tag{7.3}$$

that is,

$$\begin{pmatrix} \sqrt{n} & (\hat{\beta} - \beta) \\ \sqrt{L} & (\hat{\gamma}_1 - \gamma) \end{pmatrix} \to_d \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4(\mu_4 - \sigma^4) & 0 \\ 0 & \frac{12}{5}(\mu_4 - \sigma^4) \end{pmatrix}\right).$$

The strategy for the proof is now as follows. After identifying the leading terms of η and thus X_n we are left with $\tilde{X}_n$, which is essentially a weighted U-statistic as discussed, for example, in Lee (1990). However, $\tilde{X}_n$ does not conform to the usual assumptions. Adopting the projection method [see, e.g., Serfling (1980), Section 5.3], we therefore project $\tilde{X}_n$ onto the ε_i, obtaining in the process a weighted sum of independent random variables which is shown to be asymptotically normal in Lemma A3. In a last step, it is shown that this projection is sufficiently close to $\tilde{X}_n$ (Lemma A4).

Now, as in the previous proof,

$$\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma}_1 - \gamma \end{pmatrix} = \frac{12(n-L)^2}{L(L^2-1)} \begin{pmatrix} \cdots & \frac{L+1}{6(n-L)^2}(2L+1-3k) & \cdots \\ \cdots & \frac{2k-(L+1)}{2(n-L)} & \cdots \end{pmatrix} \eta.$$

Therefore,

$$X_n = \sum_{k=1}^{n} \omega_{k,n}\eta_k,$$

with weights (observe these are the suitably normalized w_k from above)

$$\begin{aligned} \omega_{k,n} = {} & \frac{2\lambda\sqrt{n-L}}{L(L-1)}(2L+1-3k) \\ & + \frac{6\rho\sqrt{L}(n-L)}{L(L^2-1)}(2k-(L+1)), \qquad 1 \le k \le L. \end{aligned} \tag{7.4}$$

Defining

$$\bar{\eta}_{k,n} = \sum_{\nu=1}^{n-L} \left\{(\varepsilon_{k+\nu} - \varepsilon_\nu)^2 - 2\sigma^2\right\} \Big/ (n-L),$$

we decompose r.v.'s η_k (7.2),

$$\eta_k = \overline{\eta}_{k,n} + r_{k,n}. \tag{7.5}$$

By calculations as in the proof of Theorem 4.1, we find

$$E\left(\sum_{k=1}^{L} \omega_{k,n} r_{k,n}\right) = O\left(L^{3/2}/n\right) \to 0,$$

using assumption (A4). Furthermore we observe in the proof of Theorem 4.1 that $\operatorname{var}(\sum_{k=1}^{L} \omega_k r_{k,n}) = o(1)$ (note that $\gamma = 0$), hence this is a remainder term and it is sufficient to investigate the term

$$\tilde{X}_n = \sum_{k=1}^{L} \omega_{k,n} \overline{\eta}_{k,n}. \tag{7.6}$$

Now consider the projection

$$\hat{T}_n = n^{-1} \sum_{\mu=1}^{n} E\left(\tilde{X}_n | \varepsilon_\mu\right). \tag{7.7}$$

Lemmas A3 and A4 demonstrate asymptotic normality of $\hat{T}_n$ and $E(\hat{T}_n - \tilde{X}_n)^2 \to 0$ as $n \to \infty$, and thus imply (7.3). □

APPENDIX

A.1. Technical lemmas. In the following, we use the notation

$$s_\mu^2 = \varepsilon_\mu^2 - \sigma^2, \qquad c_{\nu,n} = n^{-1} \sum_{k=1}^{\nu-1} \omega_{k,n} \tag{A.1}$$

for indices $\nu \geq 2$.

Using the notation of (7.1), somewhat tedious but simple algebra yields the following lemma.

LEMMA A1.

(i) $$\sum_{j=1}^{L} \alpha_j^2 \sim \frac{4}{L}, \sum_{j=1}^{L} \beta_j^2 \sim \frac{12n^2}{L^3}, \sum_{j=1}^{L} \alpha_j \beta_j \sim -\frac{6n}{L^2},$$

and hence

(ii) $$\sum_{j=1}^{L} w_j^2 = \frac{4\lambda^2}{L}(1+o(1)) + \frac{12n^2\rho^2}{L^3}(1+o(1)) - \frac{12\lambda\rho n}{L^2}(1+o(1));$$

(iii) $$\sum_{j=1}^{L} \alpha_j^2 j \sim 1, \sum_{j=1}^{L} \beta_j^2 j \sim \frac{6n^2}{L^2}, \sum_{j=1}^{L} \alpha_j \beta_j j \sim -\frac{2n}{L},$$

and hence

$$\text{(iv)}\qquad \sum_{j=1}^{L} w_j^2 j = \lambda^2(1+o(1)) + \frac{6\rho^2 n^2}{L^2}(1+o(1)) - \frac{4\lambda\rho n}{L}(1+o(1));$$

$$\text{(v)}\qquad \sum_{j=1}^{L} w_j = \lambda,\ \sum_{j=1}^{L} w_j j = \rho(n-L);$$

$$\text{(vi)}\qquad \sum_{j=1}^{L} w_j \sum_{k=j+1}^{L} w_k k = -\frac{\lambda^2 L}{15}(1+o(1)) + \frac{11\lambda\rho n}{10}(1+o(1)) - \frac{3\rho^2 n^2}{5L}(1+o(1));$$

$$\text{(vii)}\qquad \sum_{j=1}^{L} w_j \sum_{k=1}^{j} w_k k = \frac{\lambda^2 L}{15}(1+o(1)) - \frac{\lambda\rho n}{10}(1+o(1)) + \frac{3\rho^2 n^2}{5L}(1+o(1));$$

$$\text{(viii)}\qquad \sum_{k=1}^{L} w_k k^2 = (-\lambda L^2/6 + \rho n L)\left(1 + O\left(\frac{1}{L}\right)\right);$$

$$\text{(ix)}\qquad \sum_{k=1}^{L} |w_k| = O\left(\frac{n}{L}\right).$$

LEMMA A2. *For the quantities* $c_{\nu,n}$, $1 \le \nu \le L$, *defined in* (A.1), *we have*

$$\text{(i)}\qquad c_{L+1,n} = \lambda\sqrt{N-L}/n \sim \lambda/\sqrt{n};$$

$$\text{(ii)}\qquad c_{\nu,n} = \frac{\sqrt{n-L}}{n}\frac{\lambda(\nu-1)}{L(L-1)}(4L-3\nu+2) + \frac{n-L}{n}\frac{6\rho\sqrt{L}(\nu-1)}{L(L^2-1)}(\nu-(L+1));$$

$$\text{(iii)}\qquad \sum_{\nu=1}^{L} c_{\nu,n} = \frac{\lambda L}{\sqrt{n}}(1+o(1)) - \rho\sqrt{L}(1+o(1));$$

$$\text{(iv)}\qquad \sum_{\nu=1}^{L} c_{\nu,n}^2 = \frac{6}{5}\rho^2(1+o(1));$$

$$\text{(v)}\qquad \max_{1\le\nu\le L}\{|c_{\nu,n}|\}\Bigg/\left(\sum_{\nu=1}^{L} c_{\nu,n}^2\right)^{1/2} = O(L^{-1/2}).$$

PROOF. From (7.4),

$$\begin{aligned}
nc_{L+1,n} &= \sum_{k=1}^{L}\left(\frac{2\lambda\sqrt{n-L}}{L(L-1)}(2L+1-3k)+\frac{6\rho\sqrt{L}(n-L)}{L(L^2-1)}(2k-(L+1))\right)\\
&= \lambda\sqrt{n-L};\\
nc_{\nu,n} &= \sum_{k=1}^{\nu-1}\left(\frac{2\lambda\sqrt{n-L}}{L(L-1)}(2L+1-3k)+\frac{6\rho\sqrt{L}(n-L)}{L(L^2-1)}(2k-(L+1))\right)\\
&= \frac{\lambda\sqrt{n-L}(\nu-1)}{L(L-1)}(4L+2-3\nu)\\
&\quad+\frac{6\rho\sqrt{L}(n-L)(\nu-1)}{L(L^2-1)}(\nu-(L+1));\\
\sum_{\nu=1}^{L}c_{\nu,n} &= \frac{\lambda\sqrt{n-L}}{nL(L-1)}\sum_{\nu=1}^{L}((4L+2)(\nu-1)-3\nu^2+3\nu)\\
&\quad+\frac{6\rho\sqrt{L}(n-L)}{nL(L^2-1)}\sum_{\nu=1}^{L}(\nu^2-\nu-(L+1)(\nu-1))\\
&= \frac{\lambda L^3}{\sqrt{n}L^2}(1+o(1))-\frac{\rho\sqrt{L}L^3}{L^3}(1+o(1))\\
&= \frac{\lambda L}{\sqrt{n}}(1+o(1))-\rho\sqrt{L}(1+o(1));
\end{aligned}$$

and

$$\begin{aligned}
\sum_{\nu=1}^{L}c_{\nu,n}^2 &= \frac{\lambda^2(n-L)}{n^2L^2(L-1)^2}\sum_{\nu=1}^{L}(\nu-1)^2(4L-3\nu+2)^2\\
&\quad+\frac{36\rho^2L(n-L)^2}{n^2L^2(L^2-1)^2}\sum_{\nu=1}^{L}(\nu-1)^2(\nu-(L+1))^2\\
&\quad+\frac{12\rho\lambda\sqrt{L}(n-L)^{3/2}}{n^2L^2(L-1)(L^2-1)}\sum_{\nu=1}^{L}(4L-3\nu+2)(\nu-(L+1))(\nu-1)^2\\
&= O\left(\frac{L}{n}\right)+\frac{36\rho^2L^5}{L^5}\left(\frac{1}{5}-\frac{1}{2}+\frac{1}{3}\right)+O\left(\sqrt{\frac{L}{n}}\right)\\
&\sim \frac{6}{5}\rho^2.
\end{aligned}$$

Then (v) is obvious from (ii) and (iv).

LEMMA A3. *For $\hat{T}_n$ as in* (7.7),

$$\hat{T}_n \to_d \mathscr{N}\left(0, 4\lambda^2(\mu_4-\sigma^4)+\tfrac{12}{5}\rho^2(\mu_4-\sigma^4)\right).$$

PROOF. By the independence of the residuals we obtain

$$\begin{aligned}
\hat{T}_n &= n^{-1} \sum_{\mu=1}^{n} \sum_{\nu=1}^{n-L} \sum_{k=1}^{L} \omega_{k,n} E\Big((\varepsilon_{k+\nu} - \varepsilon_\nu)^2 - 2\sigma^2 | \varepsilon_\mu \Big) \\
&= n^{-1} \sum_{\mu=1}^{n} \sum_{\nu=1}^{n-L} \sum_{k=1}^{L} \omega_{k,n} (\delta_{k+\nu,\mu} + \delta_{\nu,\mu}) s_\mu^2 \\
&= n^{-1} \Bigg\{ \sum_{\mu=L+1}^{n-L} \sum_{k=1}^{L} \omega_{k,n} s_\mu^2 + \sum_{\mu=1}^{L} \sum_{k=1}^{\mu-1} \omega_{k,n} s_\mu^2 \\
&\qquad + \sum_{\mu=n-L+1}^{n} \sum_{k=\mu-n+L}^{L} \omega_{k,n} s_\mu^2 + \sum_{\mu=1}^{n-L} \sum_{k=1}^{L} \omega_{k,n} s_\mu^2 \Bigg\} \\
&= 2 c_{L+1,n} \sum_{\mu=L+1}^{n-L} s_\mu^2 + \left(\sum_{\mu=1}^{L} c_{\mu,n} s_\mu^2 + \sum_{\mu=1}^{L} (c_{L+1,n} - c_{\mu,n}) s_{\mu+n-L}^2 \right) \\
&\quad + c_{L+1,n} \left(\sum_{\mu=1}^{L} s_\mu^2 \right) \\
&=: I_n + II_n + III_n \quad \text{say.}
\end{aligned}$$

From Lemma A2, $c_{L+1,n} \sim \lambda/\sqrt{n}$ from which we conclude that $\mathrm{var}(III_n) = O(L/n)$ and thus $III_n \to_p 0$. Since I_n and II_n are independent for each n, we obtain that $\hat{T}_n$ is asymptotically normal if I_n, II_n are asymptotically normal. Obviously we have by the usual CLT that

$$I_n \to_d \mathscr{N}\big(0, 4\lambda^2(\mu_4 - \sigma^4)\big).$$

From Lemma A2,

$$\max_{1 \le \mu \le L} \big\{ |c_{\mu,n}| \big\} \Bigg/ \left(\sum_{\mu=1}^{L} c_{\mu,n}^2 \right)^{1/2} = O(L^{-1/2}) \to 0$$

and

$$\sum_{\mu=1}^{L} c_{\mu,n}^2 \to 6\rho^2/5.$$

This implies asymptotic normality for a weighted sum of i.i.d. random variables [see, e.g., Billingsley (1986), page 380], and therefore

$$\sum_{\mu=1}^{L} c_{\mu,n} s_\mu^2 \to_d \mathscr{N}\big(0, 6\rho^2(\mu_4 - \sigma^4)/5\big).$$

Observing that $c_{L+1,n} \sum_{\mu=1}^{L} s_{\mu+n-L}^2 \to_p 0$, we conclude

$$II_n \to_d \mathscr{N}\big(0, 12\rho^2(\mu_4 - \sigma^4)/5\big),$$

and this implies the result for the case $\gamma = 0$. For the case $\gamma > 0$, we have to replace $\overline{\eta}_{k,n}$ by $\hat{\eta}_{k,n} := \overline{\eta}_{k,n} + 2\sum_{j=1}^{n-L} (h((j+k)/n) - h((j/n))(\varepsilon_{j+k} - \varepsilon_j)/$

$(n - L)$ in the proof above and exercise more care in the partitioning of $\hat{T}_n$. We omit the details.

LEMMA A4. *For* $\tilde{X}_n$ *as in* (7.6) *and* $\hat{T}_n$ *as in* (7.7),

$$E\left(\hat{T}_n - \tilde{X}_n\right)^2 \to 0 \quad \textit{as } n \to \infty.$$

PROOF. From the proof of Lemma A3, it follows that $\operatorname{var}(\hat{T}_n) \sim \operatorname{var}(\tilde{X}_n) = O(1)$ and hence it suffices to prove that $\operatorname{cov}(\hat{T}_n, \tilde{X}_n) \sim \operatorname{var}(\hat{T}_n)$. Using the above representations for $\hat{T}_n$, Lemma A2, and noting that

$$\sum_{\mu=n-L+1}^{n} \sum_{k=\mu-n+L}^{L} \omega_{k,n} = \sum_{\rho=1}^{L} \sum_{k=\rho}^{L} \omega_{k,n} = \sum_{\mu=1}^{L} \left(c_{L+1,n} - c_{\mu,n}\right),$$

we find

$$\begin{aligned}
&\left(\mu_4 - \sigma^4\right)^{-1} \operatorname{cov}\left(\hat{T}_n, \tilde{X}_n\right) \\
&\quad = \left(\mu_4 - \sigma^4\right)^{-1} \Bigg[\Bigg(E \frac{2\lambda\sqrt{n-L}}{n} \sum_{\mu=1}^{n-L} s_\mu^2 + \sum_{\mu=1}^{L} c_{\mu,n} s_\mu^2 \\
&\qquad\qquad + \sum_{\mu=1}^{L} \left(c_{L+1,n} - c_{\mu,n}\right) s_{n-L+\mu}^2 - \frac{\lambda\sqrt{n-L}}{n} \sum_{\mu=1}^{L} s_\mu^2\Bigg) \\
&\qquad\qquad\qquad \times \left(\sum_{k=1}^{L} \sum_{\nu=1}^{n-L} \left(\varepsilon_{k+\nu} - \varepsilon_\nu\right)^2 \omega_{k,n}\right)\Bigg] \\
&\quad = \frac{2\lambda\sqrt{n-L}}{n} \sum_{\mu=1}^{n-L} \sum_{\nu=1}^{n-L} \sum_{k=1}^{L} \omega_{k,n}\left(\delta_{k+\nu,\mu} + \delta_{\nu,\mu}\right) \\
&\qquad + \sum_{\mu=1}^{L} c_{\mu,n} \sum_{\nu=1}^{n-L} \sum_{k=1}^{L} \omega_{k,n}\left(\delta_{k+\nu,\mu} + \delta_{\nu,\mu}\right) \\
&\qquad + \sum_{\mu=1}^{L} \left(c_{L+1,n} - c_{\mu,n}\right) \sum_{\nu=1}^{n-L} \sum_{k=1}^{L} \omega_{k,n}\left(\delta_{k+\nu,n-L+\mu} + \delta_{\nu,n-L+\mu}\right) \\
&\qquad - \frac{\lambda\sqrt{n-L}}{n} \sum_{\mu=1}^{n-L} \sum_{\nu=1}^{n-L} \sum_{k=1}^{L} \omega_{k,n}\left(\delta_{k+\nu,\mu} + \delta_{\nu,\mu}\right) \\
&\quad = \frac{2\lambda\sqrt{n-L}}{n} \left(\sum_{\mu=L+1}^{n-L} c_{L+1,n} + \sum_{\mu=1}^{L} c_{\mu,n} + \sum_{\mu=1}^{L} \left(c_{L+1,n} - c_{\mu,n}\right) + nc_{L+1,n}\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{\mu=1}^{L} c_{\mu,n}^2 + \sum_{\mu=1}^{L} c_{\mu,n} c_{L+1,n} + \sum_{\mu=1}^{L} c_{L+1,n}(c_{L+1,n} - c_{\mu,n}) \\
&- \sum_{\mu=1}^{L} c_{\mu,n}(c_{L+1,n} - c_{\mu,n}) - \frac{\lambda\sqrt{n-L}}{n}\left(\sum_{\mu=1}^{L} c_{\mu,n} + \sum_{\mu=1}^{L} c_{L+1,n} \right) \\
= {}& 4\lambda^2(1 + o(1)) + 2\sum_{\mu=1}^{L} c_{\mu,n}^2 + O\left(\frac{L}{n}\right) + O\left(\left(\frac{L}{n}\right)^{1/2}\right) \\
\sim {}& \operatorname{var}(\hat{T}_n).
\end{aligned}
$$

Considering now the more complicated case of estimators $\hat{\sigma}_2^2$, $\hat{\gamma}_2$ [see (4.12)], we define in analogy to (7.1) the quantities

$$
\begin{aligned}
\tilde{\alpha}_j &= \frac{3}{2L(L-1)(L-2)}(3L^2 + 3L + 2 - 6(2L+1)j + 10j^2), \\
\tilde{\beta}_j &= \frac{6(n-L)}{L(L^2-1)(L^2-4)}(-3(L+1)(L+2)(2L+1) \\
&\qquad + 2(8L+11)(2L+1)j - 30(L+1)j^2).
\end{aligned}
$$

Lemma A1 is replaced by the following result, which we state without proof. Analogous developments to those given in Section 7, then lead to results (4.13) and (4.14).

LEMMA A5. *Under* (A1)–(A3) *we have*

(i) $$\sum_{j=1}^{L} \tilde{\alpha}_j = \frac{1}{2}, \sum_{j=1}^{L} \tilde{\beta}_j = 0;$$

(ii) $$\sum_{j=1}^{L} |\tilde{\alpha}_j| = O(1), \sum_{j=1}^{L} |\tilde{\beta}_j| = O\left(\frac{n}{L}\right);$$

(iii) $$\sum_{j=1}^{L} \tilde{\alpha}_j^2 \sim \frac{9}{4L}, \sum_{j=1}^{L} \tilde{\beta}_j^2 \sim 172\frac{n^2}{L^3}, \sum_{j=1}^{L} \tilde{\alpha}_j \tilde{\beta}_j \sim -108\frac{n}{L^2};$$

(iv) $$\sum_{j=1}^{L} \tilde{\alpha}_j j = O(1); \sum_{j=1}^{L} \tilde{\beta}_j j \sim n;$$

(v) $$\sum_{j=1}^{L} \tilde{\alpha}_j^2 j \sim \frac{3}{8}, \sum_{j=1}^{L} \tilde{\beta}_j^2 j \sim 312\frac{n^2}{L^3};$$

(vi) $$\sum_{j=1}^{L} \tilde{\alpha}_j \sum_{\nu=1}^{j} \tilde{\alpha}_\nu \nu = O(L) \sum_{j=1}^{L} \tilde{\beta}_j \sum_{\nu=1}^{j} \tilde{\beta}_\nu \nu \sim \frac{96n^2}{35L}.$$

A.2. More on the preliminary change-point analysis. We provide here a brief description of some further details of the method used in Section

2; compare also Braun and Müller (1998), Sections 3.4 and 3.5.

Let $I = [a_1, a_2]$ denote the domain of the data x_i, let $\zeta > 0$ be a small constant and let $I_\zeta = [a_1 + \zeta, a_2 - \zeta]$. For $t \in I_\zeta$, local lines are fitted on the segments $S_-(t) = (-\infty, t] \cap I$ and $S_+(t) = [-\infty, t) \cap I$. These fits are obtained by locally weighted least squares [compare Fan and Gijbels (1996)]. The fit at a point $x \in S_\pm(t)$ is given by the estimated intercept $\hat{\alpha}_\pm(x)$ of the local fitted line, centered at x and taking into account only data (x_i, y_i) where $x_i \in S_\pm(t)$. Obtaining the minimizers $\tilde{\alpha}_\pm, \tilde{\beta}_\pm$ of the weighted least squares criterion

$$\sum_{x_i \in S_\pm(t)} \{y_i - (\alpha + \beta(x - x_i))\}^2 K\left(\frac{x - x_i}{b}\right)$$

with respect to α and β, we find $\hat{\alpha}_\pm(x) = \tilde{\alpha}_\pm$. Here, $b = b_n$ is a sequence of bandwidths and K is a nonnegative weight (kernel) function, often chosen as $K(z) = (1 - z^2)_+$. We then define

$$\hat{\Delta}(t) = \hat{\alpha}_+(t) - \hat{\alpha}_-(t).$$

Fixing the number ν of change-points and choosing a small constant $\rho > 0$, denoting $N_\rho(x) = \{z \in I: |x - z| \leq \rho\}$, we then obtain the estimated change-point locations

$$\hat{\tau}_1 = \arg\max_{t \in I_\zeta} \hat{\Delta}(t),$$

$$\hat{\tau}_j = \arg\max_{t \in I_j} \hat{\Delta}(t), \qquad 1 < j \leq \nu \text{ where } I_j = I_\zeta \setminus \bigcup_{i=1}^{j-1} N_\rho(\hat{\tau}_i).$$

Setting $\hat{\tau}_{(0)} = a_1$, $\hat{\tau}_{(\nu+1)} = a_2$, and denoting the ordered sample of $\hat{\tau}_i$'s by $(\hat{\tau}_{(1)}, \hat{\tau}_{(2)}, \dots, \hat{\tau}_{(\nu)})$, the $\hat{\tau}_i$ induce a partition $I = \bigcup_{i=1}^{\nu+1} S_i$, where $S_i = [\hat{\tau}_{(i-1)}, \hat{\tau}_{(i)}]$, $i = 1, \dots, \nu + 1$.

In a next step, we fit smooth functions on the segments S_j, by choosing $x \in S_j$ and using only data (x_i, y_i) for which $x_i \in S_j$, leading to the fitted values $\hat{y}(x_i)$. The automatic boundary adjustment feature of local polynomial fitting coupled with fast rates for the estimated change-points then ensures consistent estimation in the L^p sense [extending results of Müller (1992) to the situation of multiple change-points].

For various values of ν, this procedure produces results as those shown in Figure 1. For the two applications of the smoothing method (first to determine $\hat{\tau}_1, \dots, \hat{\tau}_\nu$, then to smooth the data on the induced segments), bandwidths need to be selected. The bandwidth choice for estimating the function $\hat{\Delta}$ can be taken from a global bandwidth choice for smoothing the data. Alternatively, this choice could be incorporated into the cross-validation criterion as in Braun and Müller (1998). For the bandwidth choices for smoothing on the segments, we adopted a version of a pilot method [analogous to Müller (1985)]. This method estimates and minimizes the finite integrated mean squared error of a linear smoother by substituting separate estimates for variance and bias.

Another special feature for the growth application is that we included a simple monotonization step for the estimated curves on the segments, in order to ensure that the estimated growth curve would be monotone increasing [using a method proposed by Friedman and Tibshirani (1984)]. This monotonization has the effect of possibly altering (in general, reducing) estimated jump size estimates given by $\hat{\Delta}(\hat{\tau}_j)$, $j = 1, \ldots, \nu$. These effects can be seen in Table 2, where the estimated jump sizes $\hat{\Delta}(\hat{\tau}_j)$ are indeed found to be smaller for the monotonized estimates as compared to the unrestricted estimates. Also, not unexpectedly, the fits under monotony occasionally lead to higher residual error variances, which are determined by

$$\hat{\sigma}_\nu^2 = \frac{1}{n} \sum_{i=1}^{\nu+1} \sum_{x_i \in S_i} \big(y_i - \hat{y}(x_i) \big)^2. \tag{A.2}$$

Acknowledgments. We thank a referee for detailed and helpful comments which led to numerous improvements in the paper.

REFERENCES

BHATTACHARYA, P. K. and BROCKWELL, P. J. (1976). The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrsch. Verw. Gebiete* **37** 51–75.

BILLINGSLEY, P. (1986). *Probability and Measure*, 2nd ed. Wiley, New York.

BRAUN, J. V. and MÜLLER, H. G. (1998). Statistical methods for the segmentation of DNA sequences. *Statist. Sci.* **13** 142–162.

CARLSTEIN, E., MÜLLER, H. G. and SIEGMUND, D. eds. (1994). *Change-Point Problems*. IMS, Hayward, CA.

CHURCHILL, G. A. (1992). Hidden Markov chains and the analysis of genome structure. *Computers in Chemistry* **6** 107–115.

COBB, G. W. (1978). The problem of the Nile: conditional solution to a change-point problem. *Biometrika* **62** 243–251.

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc. B* **57** 301–369.

DUBOWIK, C. (1996). Entdeckung von Changepoints und gewichtete *U*-Statistiken. Diploma thesis, Univ. Ulm.

EUBANK, R. L. and SPECKMAN, P. L. (1994). Nonparametric estimation of functions with jump discontinuities. In *Change-Point Problems* (E. Carlstein, H. G. Müller and D. Siegmund, eds.) IMS, 130–144. Hayward, CA.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

FREDKIN, D. R. and RICE, J. (1992). Bayesian restoration of single-channel patch clamp recordings. *Biometrika* **48** 427–448.

FRIEDMAN, J. and TIBSHIRANI, R. (1984). The monotone smoothing of scatterplots. *Technometrics* **26** 243–250.

GASSER, T., SROKA, L. and JENNEN-STEINMETZ, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528.

HALL, P. and TITTERINGTON, D. M. (1992). Edge-preserving and peak-preserving smoothing. *Technometrics* **34** 429–440.

HEINRICHS, C., MUNSON, P. D., COUNTS, D. R., CUTLER, G. B. and BARON, J. (1995). Patterns of human growth. *Science* **268** 442–445.

HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–16.

JARRETT, R. G. (1979). Time intervals between coal mining disasters. *Biometrika* **66** 191–193.

LAMPL, M., CAMERON, N., VELDHUIS, J. D. and JOHNSON, M. L. (1995). Patterns of human growth. *Science* **268** 445–447.

LAMPL, M., VELDHUIS, J. D. and JOHNSON, M. L. (1992). Saltation and stasis: A model of human growth. *Science* **258** 801–803.

LEE, A. J. (1990). *U-Statistics*. Dekker, New York.

LOADER, C. (1996). Change-point estimation using nonparametric regression. *Ann. Statist.* **24** 1667–1678.

MCELROY, F. W. (1967). A necessary and sufficient condition that the ordinary least-squares estimator be best linear unbiased. *J. Amer. Statist. Assoc.* **62** 1302–1304.

MÜLLER, H. G. (1985). Empirical bandwidth choice for nonparametric kernel regression by means of pilot estimators. *Statist. Decisions* **2** 193–206.

MÜLLER, H. G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761.

MÜLLER, H. G. (1993). Comment on "Local regression: Automatic kernel carpentry," by T. Hastie and C. Loader. *Statist. Sci.* **8** 134–139.

MÜLLER, H. G. and SONG, K. S. (1997). Two-stage change-point estimators in smooth regression models. *Statist. Probab. Lett.* **34** 323–335.

RICE, J. (1984). Bandwidth choice for nonparametric kernel regression. *Ann. Statist.* **12** 1215–1230.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

WANG, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82** 385–397.

WU, J. S. and CHU, C. K. (1993). Kernel-type estimators of jump points and values of a regression function. *Ann. Statist.* **21** 1545–1566.

YAO, Y.-C. (1984). Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *Ann. Statist.* **12** 1117–1123.

YAO, Y.-C. and AU, S. T. (1989). Least-squares estimation of a step function. *Sankhyā Ser. A* **51** 370–381.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616
E-MAIL: mueller@wald.ucdavis.edu

ABT. MATHEMATIK III
UNIVERSITÄT ULM
D-89069 ULM
GERMANY
E-MAIL: stamue@mathematik.uni.ulm.de