Kuechenhoff, Wellisch:

Asymptotics for generalized linear segmented regression models with an unknown breakpoint

Sonderforschungsbereich 386, Paper 83 (1997)

Asymptotics for generalized linear segmented regression models with an unknown breakpoint

Helmut Küchenhoff, Ulrich Wellisch
University of Munich, Institute of Statistics,
Akademiestraße 1, D–80799 München

Abstract

We consider asymptotic theory for the maximum likelihood estimator in the generalized linear model with an unknown breakpoint. A proof for the asymptotic normality is given. The methods are based on the work of Huber (1967). The main problem is the non–differentiability of the likelihood and the score function, which requires non–standard methods. An example from epidemiology is presented, where confidence intervals for the parameters are calculated with the asymptotic results.

Keywords: *Asymptotic Normality, Breakpoint, Maximum Likelihood, Non-differentiable Score Function, Segmented Regression.*

1 Introduction

In segmented regression models the domain of the regressor X is divided in two or more intervals, where the regression function has a different form or different parameters. The endpoints of these intervals are called changepoints or breakpoints (Seber and Wild, 1989). In this paper, we consider generalized linear models with two segments, i.e. one breakpoint. We further assume that the regression function is continuous and that only the slope parameter differs in the two segments. The regression equation can be written as

$$E(Y|X = x) = G(\alpha + \beta_1(x - \tau)_- + \beta_2(x - \tau)_+), \qquad (1)$$

where $$t_+ := \max(0,t), t_- := \min(0,t).$$

Here, G denotes the link function, e.g. logistic, identity etc. A major task with respect to such models concerns parameter estimation if the breakpoint τ is treated as an unknown parameter. Then standard asymptotic theory is not applicable, because the regression function is not differentiable in $X = \tau$. Feder (1975) has shown asymptotic normality for the least square estimator in this model with continuous response Y. We address the more general case of maximum likelihood in generalized linear models. An important special case is the logistic regression, which has many applications in toxicology and epidemiology, see Ulm (1991), Küchenhoff and Carroll (1997).

We give a proof for the asymptotic normality of the maximum likelihood estimator (MLE) for the i.i.d. case. We use methods which differ from those used by Feder and which can be applied in the much more general context of estimating equations. The outline of the paper is as follows: In the second section, the model is introduced and the consistency of the MLE is pointed out. In the third section, the asymptotic normality of this estimator is shown. In the fourth section, the derived results are applied to a study from occupational epidemiology.

2 Model and assumptions

Let $(Y_i, X_i), i = 1, \dots, n$, be an independent sample from the generalized linear segmented regression model, which is given by the conditional expectation of $Y|X$ and the corresponding conditional density from an exponential family, see e.g. Fahrmeir and Tutz (1994),

$$E(Y|X = x) = G(\alpha_0 + \beta_{10}(x - \tau_0)_- + \beta_{20}(x - \tau_0)_+), \quad (2)$$

$$f(y|\vartheta, \xi) = \exp\left\{\frac{y\vartheta - b(\vartheta)}{\xi} + c(y, \xi)\right\}. \quad (3)$$

Here, ξ is the nuisance–parameter and $b'(\vartheta) = E(Y|X = x)$. Further we assume that G is the natural link function yielding

$$\vartheta = \alpha_0 + \beta_{10}(x - \tau_0)_- + \beta_{20}(x - \tau_0)_+.$$

We want to estimate the parameter vector $\theta_0 = (\alpha_0, \beta_{10}, \beta_{20}, \tau_0)$, where the

MLE is defined by

$$\hat{\theta}_n \equiv \hat{\theta} = \arg\max \sum_{i=1}^{n} \mathcal{G}(y_i, \alpha + \beta_1 (x_i - \tau)_- + \beta_2 (x_i - \tau)_+, \xi). \tag{4}$$

$\mathcal{G}$ denotes the log–likelihood function. We further define the score function by

$$S(Y, X, \theta) = (Y - G(\alpha + \beta_1 (X - \tau)_- + \beta_2 (X - \tau)_+)) \cdot \cdot \begin{pmatrix} 1 \\ (X - \tau)_- \\ (X - \tau)_+ \\ -\beta_1 I_{\{X \le \tau\}} - \beta_2 I_{\{X > \tau\}} \end{pmatrix}. \tag{5}$$

Note that the log–likelihood is not differentiable in $\tau = X$, so the score function is the gradient with respect to θ only for $X \neq \tau$. Also the MLE is not necessarily a solution of

$$\sum_{i=1}^{n} S(Y_i, X_i, \hat{\theta}) = 0.$$

In the following we need some regularity assumptions, which are similar to those in Fahrmeir and Kaufmann (1985, p. 356).

(R1) Let $(X_i, Y_i)_{i=1,\dots,n}$ be an independent sample from model (2) and (3) with natural link function. Identifiability of model (2) holds, i.e. $\beta_{10} \neq \beta_{20}$.

(R2) X has a twice differentiable density on $\mathbb{R}$. The first two moments of X exist.

(R3) The expectations

(i) $E(\mathcal{G}(Y, \alpha_0 + \beta_{10}(X - \tau_0)_- + \beta_{20}(X - \tau_0)_+, \xi))$,

(ii) $E(S(Y, X, \theta_0)^2)$

exist.

Note that (R3)(ii) is fulfilled if the moments $E(Y^4), E(X^4), E(G^4)$ exist.

Theorem 2.1 (*Consistency and asymptotic solution*) Under regularity conditions (R1) to (R3),

(a) $\hat{\theta}$ is consistent, i.e.

$$\text{plim}\ \hat{\theta} = \theta_0 \text{ for } n \to \infty.$$

(b) $\hat{\theta}$ is an asymptotic solution of the score equations, i.e.

$$\left(n^{-\frac{1}{2}} \sum_{i=1}^{n} S(Y_i, X_i, \hat{\theta}) \right) \xrightarrow{\mathbb{P}} 0 \text{ for } n \to \infty. \tag{6}$$

Proof:

a) The first proof of the consistency of the MLE without using differentiability conditions is due to Wald (1949). Using arguments from Huber (1967) we do not need compactness of the parameterspace yielding the result.

b) Since the score function is continuously differentiable in the first three components, we have

$$\sum_{i=1}^{n} S_j(Y_i, X_i, \hat{\theta}) = 0 \text{ for } j = 1, 2, 3.$$

The forth component is the right directional derivative having jumps in all points X_i. Using that this derivative changes sign in the MLE and that X has a density we get (6) for the forth component. Details can be found in Küchenhoff (1995).

3 Asymptotic normality of the MLE

In this section, we are concerned with the asymptotic normality of the estimator $\widehat{\theta}$, i.e. the convergence in distribution

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}_4(0, V) \tag{7}$$

for $n \to \infty$, where $V = -[\frac{\partial}{\partial\theta} E(S(Y, X, \theta))|_{\theta=\theta_0}]^{-1} \xi_0$. In the following, we are using the notation

$$\frac{\partial}{\partial\theta} E(S(Y, X, \theta))|_{\theta=\theta_0} \equiv E'(S(Y, X, \theta_0)).$$

Lemma 3.1 There exist strictly positive numbers a, d_0 such that

$$|E(S(Y, X, \theta))| \geq a \cdot |\theta - \theta_0| \text{ for } |\theta - \theta_0| \leq d_0. \tag{8}$$

Proof: Since $E(S(Y, X, \theta))$ is differentiable in some neighbourhood of θ_0 the mean value theorem yields the equation

$$|E(S(Y, X, \theta))| = |E'(S(Y, X, \theta^*)) \cdot (\theta - \theta_0)|, \tag{9}$$

where $\theta^* \equiv \theta_0 + t \cdot (\theta - \theta_0)$, $t \in (0, 1)$ and θ^* is possibly different for each row of $E'(S(Y, X, .))$.
Since the mapping $E'(S(Y, X, \theta))$ is continuous in θ the convergence

$$\lim_{\theta^* \to \theta_0} E'(S(Y, X, \theta^*)) \cdot (\theta - \theta_0) = E'(S(Y, X, \theta_0)) \cdot (\theta - \theta_0)$$

is valid. According to this convergence and (9) we can choose $d_0 > 0$ such that

$$|E'(S(Y, X, \theta^*)) \cdot (\theta - \theta_0)| \geq \frac{1}{2} \cdot |E'(S(Y, X, \theta_0)) \cdot (\theta - \theta_0)|,$$

for all θ^* with $|\theta^* - \theta_0| \leq d_0$. Using the fact that the matrix $E'(S(Y, X, \theta_0))$ is symmetric and regular we get the inequality

$$|E(S(Y, X, \theta))| \geq \frac{1}{2} \lambda_{min} \cdot |\theta - \theta_0|$$

for all θ with $|\theta - \theta_0| \leq d_0$, where λ_{min} is defined as follows

$$\lambda_{min} := \min\{|\lambda| : \lambda \text{ is eigenvalue of } E'(S(Y, X, \theta_0))\}.$$

Note that $\lambda_{min} \neq 0$ holds. Setting $a = \frac{1}{2}\lambda_{min}$ we obtain (8). □

Note that the result (8) is valid for any $d_0' > 0$, such that $d_0' \leq d_0$.

Theorem 3.1 (*Asymptotic normality*) We assume that there are strictly positive numbers K_1, K_2, d_0 such that

(N1)
$$E(\sup_{|\psi-\theta|\leq d} |S(Y,X,\psi) - S(Y,X,\theta)|) \leq K_1 \cdot d$$
$$\text{for } |\theta - \theta_0| + d \leq d_0,$$

(N2)
$$E[(\sup_{|\psi-\theta|\leq d} |S(Y,X,\psi) - S(Y,X,\theta)|)^2] \leq K_2 \cdot d$$
$$\text{for } |\theta - \theta_0| + d \leq d_0.$$

Then under the regularity conditions (R1) to (R3) the MLE $\widehat{\theta}$ is asymptotically normal, i.e. (7) is valid.

Proof: The basic idea in the proof is the application of a theorem by Huber (1967) and the fact that the mapping

$$E(S(Y,X,\theta)) : \Theta \to \mathbb{R}^4$$

is continuously differentiable in a suitable neighbourhood of θ_0.
Using the mean value theorem we get the vector equation

$$E(S(Y,X,\widehat{\theta})) = E(S(Y,X,\theta_0)) + E'(S(Y,X,\theta^*)) \cdot (\widehat{\theta} - \theta_0), \tag{10}$$

where $\theta^* \equiv \theta_0 + t \cdot (\widehat{\theta} - \theta_0)$, $t \in (0,1)$ and θ^* is possibly different for each component of (10).
Equation (10) can be written in the form

$$\sqrt{n}E(S(Y,X,\widehat{\theta})) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S(Y_i,X_i,\theta_0) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S(Y_i,X_i,\theta_0) =$$
$$E'(S(Y,X,\theta^*)) \cdot \sqrt{n}(\widehat{\theta} - \theta_0). \tag{11}$$

According to our regularity conditions (R1) to (R3), Theorem 2.1, the conditions (N1), (N2) and Lemma 3.1 the assumptions of the theorem of Huber (1967) (cf. appendix) are fulfilled and therefore

$$\sqrt{n}E(S(Y,X,\widehat{\theta})) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S(Y_i,X_i,\theta_0) \xrightarrow{\mathbb{P}} 0. \tag{12}$$

The central limit theorem for i.i.d. random vectors yields

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} S(Y_i, X_i, \theta_0) \xrightarrow{D} \mathcal{N}_4(0, -E'(S(Y, X, \theta_0)) \cdot \xi_0). \tag{13}$$

Note that the assumption of a natural link function yields

$$\mathrm{Var}(S(Y, X, \theta_0)) = -E'(S(Y, X, \theta_0)) \cdot \xi_0.$$

Further, the convergence

$$E'(S(Y, X, \theta^*)) \xrightarrow{\mathbb{P}} E'(S(Y, X, \theta_0)). \tag{14}$$

is valid. The convergences (12), (13), (14) together with an argument of the Crámer type (cf. Pruscha, 1996, Prop. B 3.9., p. 397) complete the proof. □

In the following two theorems we present sufficient conditions for the conditions (N1) and (N2). Let $S_i(\theta) \equiv S_i(Y, X, \theta)$, $i = 1, ..., 4$, denote the four components of the vector $S(Y, X, \theta)$.

Theorem 3.2 (*Sufficient conditions*) The set of conditions:

(S_1) The regressor variable X has a bounded density,

(S_2) There are strictly positive numbers d_0, a_1, a_2, a_3, a_4, such that

$$E \sup_{\rho \in B(\theta_0, d_0)} |S_i'(\rho)| \leq a_i, \; i = 1, ..., 4,$$

where the supremum is taken over all $\rho \in B(\theta_0, d_0)$ where S_i is differentiable,

is sufficient for (N1).

Proof: Let θ_4 denote the fourth component of the parameter vector θ. In the following various positive constants are denoted by C. Note that the four components of the score vector $S(Y, X, \theta)$ are not generally differentiable and the fourth component $S_4(\theta)$ can have jumps with respect to the variable $\tau \equiv \theta_4$. These circumstances are the main technical aspects in the proof.

For $|\theta-\theta_0|+d \le d_0$ the inequality

$$E \sup_{|\psi-\theta|\le d} |S(Y,X,\psi)-S(Y,X,\theta)| \le E \sup_{|\psi-\theta|\le d} |S_1(\psi)-S_1(\theta)| +$$
$$E \sup_{|\psi-\theta|\le d} |S_2(\psi)-S_2(\theta)| + E \sup_{|\psi-\theta|\le d} |S_3(\psi)-S_3(\theta)| +$$
$$E[I_{\{X\notin B(\theta_4,d)\}} \sup_{|\psi-\theta|\le d} |S_4(\psi)-S_4(\theta)|] +$$
$$E[I_{\{X\in B(\theta_4,d)\}} \sup_{|\psi-\theta|\le d} |S_4(\psi)-S_4(\theta)|] \tag{15}$$

holds. For the first four terms of the right hand side of (15) an application of the mean value theorem yields for $i=1,...,3$

$$E \sup_{|\psi-\theta|\le d} |S_i(\psi)-S_i(\theta)| \le C\cdot d\cdot E \sup_{\rho\in B(\theta_0,d_0)} |S_i'(\rho)|, \tag{16}$$

and

$$E[I_{\{X\notin B(\theta_4,d)\}} \sup_{|\psi-\theta|\le d} |S_4(\psi)-S_4(\theta)|] \le d\cdot E \sup_{\rho\in B(\theta_0,d_0)} |S_4'(\rho)|. \tag{17}$$

For the fifth term the inequalities

$$E[I_{\{X\in B(\theta_4,d)\}} \sup_{|\psi-\theta|\le d} |S_4(\psi)-S_4(\theta)|] \le$$
$$E[I_{\{X\in B(\theta_4,d)\}}\cdot C_{jump} + C\cdot d\cdot \sup_{\rho\in B(\theta_0,d_0)} |S_4'(\rho)|] \le$$
$$C\cdot d + C\cdot d\cdot E \sup_{\rho\in B(\theta_0,d_0)} |S_4'(\rho)| \tag{18}$$

hold, where C_{jump} is a positive constant, which is due to the possible jump in the fourth component of $S(Y,X,\theta)$. Note that we have used the boundedness of the density of X in the second inequality. The inequalities (15), (16), (17), (18) and the condition (S_2) finish the proof. □

Theorem 3.3 (*Sufficient conditions*) The pair of conditions:

(S_1') The regressor variable X has a bounded density,

(S_2') There are strictly positive numbers d_0, b_1, b_2, b_3, b_4, such that

$$E[\sup_{\rho \in B(\theta_0, d_0)} |S_i'(\rho)|]^2 \leq b_i,\ i = 1, ..., 4,$$

where the supremum is taken over all $\rho \in B(\theta_0, d_0)$ where S_i is differentiable,

is sufficient for (N2).

Proof: Arguing analogously as in the proof of Theorem 3.2 we obtain the inequality

$$E[\sup_{|\psi-\theta|\leq d} |S(Y, X, \psi) - S(Y, X, \theta)|]^2 \leq$$

$$C \cdot d^2 \cdot \left(\sum_{i=1}^{4} E[\sup_{\rho \in B(\theta_0, d_0)} |S_i'(\rho)|]^2 \right) + C \cdot d^2.$$

Choosing $d_0 \leq 1$ it yields $d \leq 1$. Therefore $d^2 \leq d$ holds and the theorem is proven. □

The conditions (S_1), (S_2) and (S_1'), (S_2') respectively are connected with the special form of the link function as well as with the distribution of the regressor X. For the identity link, i.e. the case of the least square estimation, the conditions (S_2) and (S_2') are obviously fulfilled, if $E(X^4)$ exists. In the following theorem we formulate sufficient conditions for the conditions (S_2) and (S_2'). These assumptions are fulfilled for the logistic regression model which is often used in practice (cf. Sec. 4).

Theorem 3.4 (*Logistic regression*) Under the conditions

- The mappings $|G|$ and $|G'|$ are bounded.
- $E(X^4) < \infty, \quad E(Y^2) < \infty.$

the assumptions on boundedness (S_2) and (S_2') hold.

Proof: The result follows immediately by computation of the derivatives S_i', $i = 1, ..., 4,$. □

4 The determination of a threshold limiting value: cement dust and chronic bronchitis

In occupational epidemiology it is often of interest to assess a so–called threshold limiting value (TLV) which is defined as the maximum concentration of a chemical substance at the workplace under which no negative impact on the employee's health is expected even if the employee is repeatedly exposed over long periods. Here, we are especially concerned with the concentration of cement dust in the workplace air which is regarded as a possible risk factor of chronic bronchitis. The investigation of this relationship caused several epidemiological studies with that topic which were conducted by the German research foundation (DFG) between 1972 and 1977 (DFG, 1981). The one we consider here involved 499 workers from a cement plant in Heidelberg. For each of these workers, four variables can be used for the analysis: chronic bronchitis, average exposure to cement dust concentration, smoking and duration. The dust concentration [mg/m^3] was calculated as a weighted average for each worker where at most five "measurements" taken at the workplace were included. The covariate smoking (*SM*) can only be taken into account as binary variable. The last covariate duration (*DUR*) means the age of a person at the first examination with respect to chronic bronchitis (*CBR*) minus the age when first exposed to cement dust, where *CBR* was also measured as binary. For a more detailed discussion of the considered variables and further details of the data set and its analysis we refer to Küchenhoff and Pigeot (1998).

The assessment of a TLV can be coped with modelling the TLV as a breakpoint in a segmented regression model, typically with G of model (2) chosen as the logistic function, where the slope of the first segment is fixed as zero. Since the response variable Y, i.e. *CBR*, is binary its expectation reads as $\mathbb{P}(CBR = 1)$ such that we obtain the following model equation

$$\mathbb{P}(CBR = 1) = G(\alpha + \beta_X(X - \tau)_+ + \beta_s SM + \beta_d DUR), \qquad (19)$$

where X denotes the logarithmically transformed dust concentration after having added 1.0 to its original value, i.e. $\log_{10}(1 + \text{dust concentration})$. The MLE of the unknown parameters can be calculated using an algorithm proposed by Küchenhoff (1997). Since the results obtained in Section 2 and 3 can be directly transfered to models with further covariates and a fixed slope parameter in the first segment, the estimated standard deviations are derived using the above asymptotic theory. The results are summarized in Table 1. The corresponding confidence intervals can be calculated from these results in the usual way. The resulting regression curve is depicted in Figure 1, where the additional covariates smoking and duration enter the model by their means.

Table 1: Parameter estimates of the threshold model for the Heidelberg data. The estimated standard deviations are given in column $\hat{\sigma}$.

parameter	estimate	$\hat{\sigma}$
τ	0.394	0.020
α	-2.169	0.401
β_X	52.390	31.474
β_s	0.470	0.262
β_d	0.026	0.009

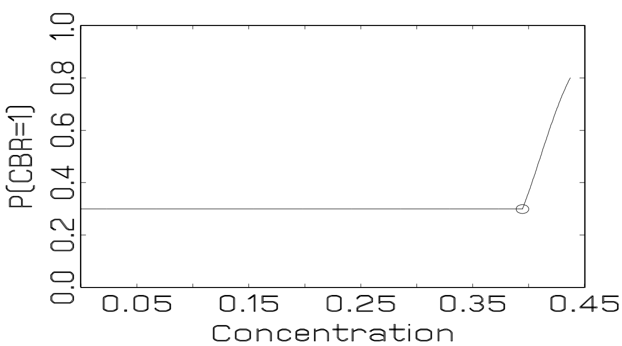

Figure 1: Segmented logistic regression model. Circle marks the estimated threshold limiting value.

The estimated TLV results in 1.48 mg/m^3 ($\hat{\tau} = 0.394$) where the estimated slope parameter for the second segment shows an extremely increasing relationship between dust concentration and *CBR* ($\hat{\beta}_X = 52.390$). Since the estimated threshold limiting value for cement dust takes a rather large value, i.e. it is located at the upper border of the observed dust concentrations, our results have to be interpreted very carefully.

5 Appendix

We present the following result from Huber (1967, 1981) in our notation and give some useful explanations.
Let Θ be an open subset of d-dimensional Euclidian space $\mathbb{R}^d$, $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ is a probability space, and $S : \mathcal{X} \times \Theta \to \mathbb{R}^d$ some mapping. Assume that $X_1, X_2, ...$ are independent random variables with values in $\mathcal{X}$ and common distribution $\mathbb{P}$. Further $\widehat{\theta} \equiv \widehat{\theta}_n(X_1, ..., X_n)$ denotes an estimator for $\theta \in \Theta$.

Theorem Under the assumptions

(H1) For each fixed $\theta \in \Theta$, $S(X, \theta)$ is measurable and $S(X, \theta)$ is separable,

(H2) There is a $\theta_0 \in \Theta$ with $ES(X, \theta_0) = 0$,

(H3) There are strictly positive numbers a, b, c, d_0 such that

- (i) $|E(S(X, \theta))| \geq a \cdot |\theta - \theta_0|$ for $|\theta - \theta_0| \leq d_0$,
- (ii) $E(\sup_{|\psi-\theta|\leq d} |S(X, \psi) - S(X, \theta)|) \leq b \cdot d$ for $|\theta - \theta_0| + d \leq d_0$,
- (iii) $E[(\sup_{|\psi-\theta|\leq d} |S(X, \psi) - S(X, \theta)|)^2] \leq c \cdot d$ for $|\theta - \theta_0| + d \leq d_0$,

(H4) $E(|S(X, \theta_0)|^2) < \infty$,

for an estimator $\widehat{\theta}$, which has the properties

$$\mathbb{P}(|\widehat{\theta} - \theta_0| \leq d_0) \to 1, \tag{E1}$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} S(X_i, \widehat{\theta}) \xrightarrow{\mathbb{P}} 0, \tag{E2}$$

the convergence

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} S(X_i, \theta_0) + \sqrt{n} E(S(X, \widehat{\theta})) \xrightarrow{\mathbb{P}} 0$$

holds.

Remarks:

(i) Note that in the conditions (H3) (ii) and (H3) (iii) the supremum is taken for fixed θ.

(ii) Condition (H3) (iii) is somewhat stronger than needed; it can be weakened to

$$E(\sup_{|\psi-\theta|\leq d} |S(X,\psi) - S(X,\theta)|)^2 \leq o(|\log d|^{-1}).$$

(iii) The expression $E(S(X,\widehat{\theta}))$ is a random vector. Here the notation means that we first compute the expectation of $S(X,\theta)$ with a fixed θ and then plug in the estimator $\widehat{\theta}$ in this function of θ.

(iv) A consistent estimation equation estimator (consistent asymptotic solution) of the estimation equation (cf. Pfanzagl, 1994, Sec. 7.4)

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} S(X_i,\theta) = 0$$

fulfills the assumptions (E1), (E2).

Acknowledgement: This work is supported by the Deutsche Forschungsgemeinschaft (SFB 386).

References

Deutsche Forschungsgemeinschaft: DFG-Forschungsbericht (1981). *Chronische Bronchitis und Staubbelastung am Arbeitsplatz; Teil 2*, Harald Boldt, Boppard.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* **13**, 342–368.

Feder, P. (1975). On asymptotic distribution theory in segmented regression problems – identified case. *Annals of Statistics* **3**, 49–83.

Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **1**, University of California Press, Berkeley.

Huber, P.J. (1981). *Robust Statistics.* John Wiley New York.

Küchenhoff, H. (1997). An exact algorithm for estimating breakpoints in segmented generalized linear models. *Computational Statistics* **12**, 235–247.

Küchenhoff, H. (1995). *Schätzmethoden in mehrphasigen Regressionsmodellen.* Habilitationsschrift, University of Munich.

Küchenhoff, H. and Carroll, R.J. (1997). Segmented regression with errors in predictors: semi-parametric and parametric methods. *Statistics in Medicine* **16**, 169-188.

Küchenhoff, H. and Pigeot, I. (1998). Comparison of statistical software packages for finding a threshold limiting value in dose–response–relationships. To appear in: *Computational Statistics.*

Pfanzagl, J. (1994). *Parametric Statistical Theory.* de Gruyter Berlin.

Pruscha, H. (1996). *Angewandte Methoden der Mathematischen Statistik.* Teubner Stuttgart.

Seber, G.A.F and Wild, C.J. (1989). *Nonlinear Regression.* John Wiley New York.

Ulm, K. (1991). A statistical method for assessing a threshold in epidemiological studies. *Statistics in Medicine* **10**, 341–349.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Annals of Statistical Mathematics* **20**, 595–601.