Biometrika (2004), **91**, 4, *pp.* 849–862

Printed in Great Britain

A semiparametric changepoint model

By ZHONG GUAN

Department of Mathematical Sciences, Indiana University South Bend, South Bend, Indiana 46634, U.S.A.

zhong.guan@yale.edu

Summary

A semiparametric changepoint model is considered and the empirical likelihood method is applied to detect the change from a distribution to a weighted distribution in a sequence of independent random variables. The maximum likelihood changepoint estimator is shown to be consistent. The empirical likelihood ratio test statistic is proved to have the same limit null distribution as that with parametric models. A data-based test for the validity of the models is also proposed. Simulation shows the sensitivity and robustness of the semiparametric approach. The methods are applied to some classical datasets such as the Nile River data and stock price data.

Some key words: Changepoint; Empirical likelihood; Exponential family; Limit theorem; Power; Resampling; Robustness; Semiparametric changepoint model; Weighted distribution.

1. Introduction

In parametric models and linear models, the test statistics for a changepoint are generally related to the likelihood ratio statistic. The general results about parametric changepoint models can be found in Csörgő & Horváth (1997).

Nonparametric changepoint models have also been studied by many authors, such as Dümbgen (1991), Carlstein (1988) and Darkhovskh (1976), and again much discussion can be found in Csörgő & Horváth (1997). However, unlike in the parametric cases, most nonparametric changepoint models assume no relationship between the two population distributions, whereas in practice it is often very natural to assume that there is some such link. This scenario is similar to the assumption of some link function, such as the logistic link, to relate case and control population distributions. The logistic link corresponds to the semiparametric exponential model of the proposed semiparametric changepoint model in this paper. This exponential model is used frequently in epidemiology just as logistic regression is used in case-control studies.

We consider the following semiparametric changepoint model and focus on the estimation of and testing for a changepoint to a biased sample. Using the empirical likelihood method (Owen, 1988; Qin & Lawless, 1994; Zhang, 1997), we can make efficient use of auxiliary information about the relationship between the two population distributions. We assume that $x_1, \ldots, x_n$ are independent vectors in $\mathbb{R}^m$. We wish to test between the following hypotheses:

H_0: $x_1, \ldots, x_n$ is a random sample from a population with distribution function F;

H_1: for some $n^{-1} \leqslant \theta_n < 1$, $x_1, \ldots, x_{n\theta_n}$ is a random sample from F and $x_{n\theta_n+1}, \ldots, x_n$ is a random sample from a weighted G, where

$$G(y) = \frac{1}{w}\int_{-\infty}^{y} w(x, \beta) dF(x), \quad w(.,.) > 0, \tag{1·1}$$

in which $w = \int_{-\infty}^{\infty} w(s, \beta) dF(s) < \infty$ is the normalising constant and the weight function $w(x, \beta)$ is of known form, but may depend on an unknown parameter $\beta \in \mathbb{R}^d$. We assume that $w(x, \beta)$ is positive and differentiable with respect to β. If $\beta \neq 0$, $w(x, \beta) \neq 1$ and $w(x, 0) = 1$ for all x. Here θ_n is an unknown parameter taking a value from $\Theta_n = \{k/n : k = 1, \ldots, n\}$. In the asymptotic theory of this paper, it is assumed that $\theta_n \to \theta_0 \in (0, 1]$ as $n \to \infty$.

A slight but significant extension of H_1 is the so-called epidemic alternative (Levin & Kline, 1985; Yao, 1993) that, for some $1 \leqslant k_1 < k_2 < n$, $x_1, \ldots, x_{k_1}, x_{k_2+1}, \ldots, x_n$ is a sample from F and $x_{k_1+1}, \ldots, x_{k_2}$ is a sample from G. Results corresponding to epidemic alternatives will be reported in a separate paper.

In this semiparametric model, the distributions F and G are treated nonparametrically except that the ratio of the density or probability functions has a known parametric form. We will focus on the change from F to the weighted G. Although most of the technical results of this paper are given for a general weight $w(x; \beta)$, the most important model in practice is the semiparametric exponential model, with weight $w(x, \beta) = \exp\{\beta^{\mathrm{T}} \tau(x)\}$. In this case, the log ratio of the two density or probability functions, f and g say, is of the form

$$\log \frac{g(x)}{f(x)} = \alpha + \beta^{\mathrm{T}} \tau(x), \tag{1·2}$$

where $\alpha = -\log w$. Typically, we use $\tau(x) = x$, or $\tau(x)$ is chosen to contain second- or higher-order powers of x. These choices correspond to the first-, second- or higher-order approximation of $\log\{g(x)/f(x)\}$. Of course, if we believe that both f and g belong to a parametric exponential family, then (1·2) is an exact parametric expression and we also have other choices for $\tau(x)$ such as $\tau(x) = \log x$. We shall show that this exponential model can be used to detect a possible change in $E\{\tau(X)\}$. In fact, from (1·2), it follows that the symmetrised Kullback–Leibler information distance $\bar{I}(f, g)$ between the two distributions in the exponential changepoint model measures the change in $E\{\tau(X)\}$, that is $\bar{I}(f, g) = \frac{1}{2}\beta^{\mathrm{T}}[E_F\{\tau(X)\} - E_G\{\tau(X)\}]$. The quantity $2\bar{I}(f, g)$ is also called J-divergence (Jeffreys, 1946). In § 7 we show that the model perfectly fits some classic changepoint datasets. Dümbgen (1991) also considered this model as an example of a general nonparametric model, but did not use information about the relationship between F and G. The general parametric exponential family changepoint model, studied by Worsley (1986), covers most of the parametric models used in applications.

2. Methodology

Let $x_1, \ldots, x_{n\theta_n}$ and $x_{n\theta_n+1}, \ldots, x_n$ be independent and identically distributed observations from population F and a weighted population G as in (1·1) respectively. The likelihood L of the data is

$$L = \prod_{i=1}^{n} p_i \prod_{j=n\theta_n+1}^{n} w(x_j) \left\{ \sum_{i=1}^{n} p_i w(x_i) \right\}^{-n(1-\theta_n)}, \tag{2·1}$$

where $p_i = dF(x_i)$ $(i = 1, \ldots, n)$. For each fixed $\theta_n \in \Theta_n \cap (0, 1)$, let $l(\theta_n, w, \beta)$ be the maximum value of $\log L + n \log n$ with respect to p_i $(i = 1, \ldots, n)$ subject to the constraints

$$\sum_{i=1}^{n} p_i w(x_i) = w, \quad \sum_{i=1}^{n} p_i = 1 \quad (p_i \geqslant 0, i = 1, \ldots, n).$$

As in Qin (1993), see also Vardi (1982), the Lagrange multiplier method leads to

$$l(\theta_n, w, \beta) = -\sum_{i=1}^{n} \log\{\theta_n w + (1-\theta_n) w(x_i, \beta)\} + \sum_{i=n\theta_n+1}^{n} \log\{w(x_i, \beta)\} + n\theta_n \log w. \tag{2·2}$$

Therefore, the score functions are

$$\psi_1(\theta_n, w, \beta) = \frac{\partial l(\theta_n, w, \beta)}{\partial w} = \frac{\theta_n(1-\theta_n)}{w} \sum_{i=1}^{n} \frac{w(x_i, \beta) - w}{(1-\theta_n) w(x_i, \beta) + \theta_n w}, \tag{2·3}$$

$$\psi_2(\theta_n, w, \beta) = \frac{\partial l(\theta_n, w, \beta)}{\partial \beta} = \sum_{i=n\theta_n+1}^{n} \frac{w'_\beta(x_i, \beta)}{w(x_i, \beta)} - \sum_{i=1}^{n} \frac{(1-\theta_n) w'_\beta(x_i, \beta)}{(1-\theta_n) w(x_i, \beta) + \theta_n w}, \tag{2·4}$$

where $w'_\beta(x, \beta) = \partial w(x, \beta)/\partial \beta$. Let $(\hat{w}, \hat{\beta}^{\mathrm{T}}) = (\hat{w}(\theta_n), \hat{\beta}^{\mathrm{T}}(\theta_n))$ be the solution to

$$\psi_1(\theta_n, w, \beta) = 0, \quad \psi_2(\theta_n, w, \beta) = 0. \tag{2·5}$$

Consequently,

$$\hat{p}_i = \frac{1}{n} \frac{\hat{w}}{(1-\theta_n) w(x_i, \hat{\beta}) + \theta_n \hat{w}} \quad (i = 1, \ldots, n) \tag{2·6}$$

and the profile loglikelihood function of the unknown changepoint θ_n is given by

$$\begin{aligned} l(\theta_n) &\equiv l\{\theta_n, \hat{w}(\theta_n), \hat{\beta}(\theta_n)\} \\ &= -\sum_{i=1}^{n\theta_n} \log\left\{(1-\theta_n)\frac{w(x_i, \hat{\beta})}{\hat{w}} + \theta_n\right\} - \sum_{i=n\theta_n+1}^{n} \log\left\{1 - \theta_n + \theta_n \frac{\hat{w}}{w(x_i, \hat{\beta})}\right\}, \end{aligned} \tag{2·7}$$

with $l(0) = l(1) = 0$.

The changepoint estimator $\hat{\theta}_n$ can be defined as

$$\hat{\theta}_n = \min\left[\arg\max\{l(\theta_n) : \theta_n \in \Theta_n\}\right]. \tag{2·8}$$

When $w(x, \beta) = \exp\{\beta^{\mathrm{T}} \tau(x)\}$, this model includes the exponential family and the 'partially exponential' family of the form $dG(x) = \exp\{\beta^{\mathrm{T}} \tau(x) + c(x, \phi) + b(\beta, \phi)\} dx$. Write $w = e^{-\alpha}$. If $\theta_n \in (0, 1)$ is known, this is a two-sample semiparametric model related to the logistic regression model

$$\mathrm{pr}(Z = 1 | X = x) = \frac{\exp\{\alpha^* + \beta^{\mathrm{T}} \tau(x)\}}{1 + \exp\{\alpha^* + \beta^{\mathrm{T}} \tau(x)\}}, \tag{2·9}$$

where Z is a binary response variable. Based on case-control data and the empirical likelihood method, Qin & Zhang (1997) and Zhang (1999) respectively proposed a Kolmogorov–Smirnov-type test and a chi-squared-type test for testing the goodness-of-fit. Let $z_i = 0$ if $i = 1, \ldots, n\theta_n$, and $z_i = 1$ if $i = n\theta_n + 1, \ldots, n$. For each fixed θ_n, data (z_i, x_i) $(i = 1, \ldots, n)$ can be fitted by the model (2·9). Using statistical packages such as R and S-Plus, one can easily obtain estimates $\hat{\alpha}^*(\theta_n)$ and $\hat{\beta}(\theta_n)$ of the coefficients of the logistic regression so that $\hat{w}(\theta_n) = \exp\{-\hat{\alpha}(\theta_n)\}$, where $\hat{\alpha}(\theta_n) = \hat{\alpha}^*(\theta_n) + \log\{\theta_n/(1-\theta_n)\}$.

As pointed out by Albert & Anderson (1984), if, for some k, $x_1, \ldots, x_k$ and $x_{k+1}, \ldots, x_n$ are completely or quasi-completely separated, the maximum likelihood estimate $\hat{\beta}(k/n)$ in the above logistic regression does not exist. This may well happen if F and G are very different. The methods of Albert & Anderson (1984) and Santner & Duffy (1986) can be applied, for each $k=1, \ldots, n-1$, to check the separation status. If there exist some k's such that there is a complete or quasi-complete separation, we can use a two-sample test method to see whether the change is significant or not and define the changepoint as the smallest k for which a significant change has occurred. Other methods such as non-parametric methods can also be used to locate the changepoint in this case. Iterative methods are usually required to find the maximum likelihood estimates in the semi-parametric model. When the true value of θ_n is close to $1/n$ or 1, the convergence of the iteration is not guaranteed. However, the simulation study of § 5 shows that the semiparametric maximum likelihood ratio test performs very well in this case.

A significantly large value of the empirical log likelihood ratio statistic

$$S_n = 2l(\hat{\theta}_n) = 2 \max_{\theta_n} l(\theta_n) \tag{2·10}$$

will lead to rejection of H_0 in favour of H_1, as in § 1.

The algorithm for finding the changepoint estimate and calculating the test statistic is very simple. First, for each $\theta_n = k/n$ $(1 \leqslant k < n)$ solve the system of equations (2·5) to obtain $\hat{w}(\theta_n)$ and $\hat{\beta}(\theta_n)$. Secondly, calculate the profile loglikelihood $l(\theta_n)$ as in (2·7) for each θ_n. Finally, obtain $\hat{\theta}_n$ as in (2·8) and S_n as in (2·10). R and S-Plus libraries for exponential changepoint models are available from the author upon request.

3. Some asymptotic results

Define $\Lambda_0 = \Lambda_n = 0$ and, for $k=1, \ldots, n-1$, $n\Lambda_k = l(k/n)$ so that $S_n = 2\max_{0 \leqslant k \leqslant n}\{n\Lambda_k\}$. For $k=1, \ldots, n-1$, write $\hat{w}_k = \hat{w}(k/n)$ and $\hat{\beta}_k = \hat{\beta}(k/n)$, the maximum semiparametric likelihood estimators of w and β, respectively, when $\theta_n = k/n$.

Let $\eta = (\theta, w, \beta^{\mathrm{T}})$, $a(x, \eta) = (1-\theta)w(x, \beta) + \theta w$ and $a(x, \eta_0) = (1-\theta_0)w(x, \beta_0) + \theta_0 w_0$, where $w_0 = E_F\{w(X, \beta_0)\}$, and β_0 is the true value of β. Clearly, under H_0, $w_0 = 1$ and $\beta_0 = 0$. For $0 < \theta < 1$, define

$$g_1(\theta, w, \beta) = \frac{\theta(1-\theta)}{w_0 w} \int_{-\infty}^{\infty} \frac{a(x, \eta_0)\{w(x, \beta) - w\}}{a(x, \eta)} dF(x), \tag{3·1}$$

$$g_2(\theta, w, \beta) = \frac{1}{w_0} \int_{-\infty}^{\infty} \frac{\{(\theta_0 - \theta_0 \wedge \theta)w_0 + (1 - \theta_0 \vee \theta)w(x, \beta_0)\}w'_\beta(x, \beta)}{w(x, \beta)} dF(x) - \frac{1-\theta}{w_0} \int_{-\infty}^{\infty} \frac{a(x, \eta_0)w'_\beta(x, \beta)}{a(x, \eta)} dF(x). \tag{3·2}$$

Set

$$A = \begin{pmatrix} -\dfrac{\partial g_1(\theta, w, \beta)}{\partial w} & -\dfrac{\partial g_1(\theta, w, \beta)}{\partial \beta^{\mathrm{T}}} \\ -\dfrac{\partial g_2(\theta, w, \beta)}{\partial w} & -\dfrac{\partial g_2(\theta, w, \beta)}{\partial \beta^{\mathrm{T}}} \end{pmatrix}.$$

We need the following assumptions for our main theoretical results.

Assumption 1. The function $w(x, \beta)$ is twice differentiable with respect to β and

$$E_F\left\{\frac{a(\eta_0)}{a^2(\eta)}\left|\frac{\partial^i w(X,\beta)}{\partial\beta_u^i}\frac{\partial^j w(X,\beta)}{\partial\beta_v^j}\right|\right\} < \infty,$$

for $1 \leqslant u$, $v \leqslant d$, $i + j \leqslant 2$ and $\theta \in (0, 1)$.

Assumption 2. The matrix A is positive definite, that is $A > 0$.

Assumption 3. There are functions $M_i(x)$ such that $\{w(x,\beta)\}^{-1}|\partial^i w(x,\beta)/\partial\beta_j^i| \leqslant M_i(x)$, and $E_F\{M_i(X)\} < \infty$, $E_F\{M_1^3(X)\} < \infty$ and $E_F\{M_1(X)M_2(X)\} < \infty$, where $1 \leqslant i \leqslant 3$, $1 \leqslant j \leqslant d$.

It is easy to see that Assumption 2 is satisfied for exponential changepoint models when $\tau(x)$ is a nonconstant vector of real-valued functions. The following result about the approximate null distribution of the test statistic is the same as that for the parametric case (Csörgő & Horváth, 1997, Theorem 1.3.1). Proofs of the theorems are given in the Appendix.

THEOREM 1. *Suppose that Assumptions* 1–3 *are fulfilled. If* H_0 *is true, then*

$$\lim_{n\to\infty} \operatorname{pr}\{C(\log n)Z_n \leqslant t + D_d(\log n)\} = \exp(-2e^{-t}), \tag{3·3}$$

for all t, *where* $Z_n = S_n^{\frac{1}{2}}$, $C(x) = (2\log x)^{\frac{1}{2}}$, $D_d(x) = 2\log x + \frac{1}{2}d\log\log x - \log\Gamma(\frac{1}{2}d)$ *and* $\Gamma(t)$ *is the gamma function.*

Define $\tau(x) = w'_\beta(x, \beta_0)$ and $\tau_i = \tau(x_i)$ $(i = 1, \ldots, n)$. The proof of this theorem indicates that the proposed test statistic S_n is asymptotically equivalent to the nonparametric test statistic

$$\tilde{S}_n = \max_{1\leqslant k<n} \frac{n}{k(n-k)} G^{\mathrm{T}}(k)\Sigma_n^{-1}G(k), \tag{3·4}$$

which in turn is equivalent to the statistic

$$\tilde{S}_n^* = \max_{1\leqslant k<n} \frac{n}{k(n-k)} G^{\mathrm{T}}(k)\Sigma_k^{-1}G(k) \tag{3·5}$$

(Csörgő & Horváth, 1997, p. 76, eqn (2.1.66)) for the change of $E\{\tau(X)\}$ with unknown constant var $\{\tau(X)\}$, where

$$\Sigma_k = \begin{cases} n^{-1}\{\sum_{i=1}^k(\tau_i - \bar{\tau}_k)(\tau_i - \bar{\tau}_k)^{\mathrm{T}} + \sum_{i=k+1}^n(\tau_i - \tilde{\tau}_k)(\tau_i - \tilde{\tau}_k)^{\mathrm{T}}\}, & \text{if } k < n, \\ n^{-1}\sum_{i=1}^n(\tau_i - \bar{\tau}_n)(\tau_i - \bar{\tau}_n)^{\mathrm{T}}, & \text{if } k = n, \end{cases} \tag{3·6}$$

$$\bar{\tau}_k = \frac{1}{k}\sum_{i=1}^k \tau_i, \quad \tilde{\tau}_k = \frac{1}{n-k}\sum_{i=k+1}^n \tau_i, \quad G(k) = \sum_{i=1}^k \tau_i - \frac{k}{n}\sum_{i=1}^n \tau_i. \tag{3·7}$$

For example, if $w(x, \beta) = \exp\{\beta^{\mathrm{T}}\tau(x)\}$, then $\tau(x) = w'_\beta(x, \beta_0)$ with $\beta_0 = 0$. Unfortunately, as in the parametric case, (3·3) only gives a conservative rejection region, because of the slow convergence rate of (3·3). However, in § 4, Theorem 1 plays a key role in obtaining the approximation of critical values as in Theorem 1.3.2 and Corollary 1.3.1 of Csörgő & Horváth (1997).

THEOREM 2. *Suppose that Assumptions* 1 *and* 2 *are satisfied. If* $0<\theta_0\leqslant 1$ *and* $k/n\to\theta\in(0,1)$, *then, as* $n\to\infty$, $\Lambda_k\to\lambda(\theta)$ *almost surely, where the function* $\lambda(\theta)$ *is strictly convex on* $(0,\theta_0)$ *and* $(\theta_0,1)$, $\max_{\theta\in[0,1]}\lambda(\theta)=\lambda(\theta_0)$ *and* $\lambda(\theta)<\lambda(\theta_0)$ *for any* $\theta\neq\theta_0$.

This theorem shows that the maximum likelihood estimator $\hat{\theta}_n$ is asymptotically unique when $0<\theta_0<1$. Consistency is established in the following result.

THEOREM 3. *Suppose that Assumptions* 1 *and* 2 *are satisfied.*
(i) *Under* H_0, *as* $n\to\infty$, $\hat{\theta}_n\to\xi_0$ *in distribution and* $\mathrm{pr}(\xi_0=0)=\mathrm{pr}(\xi_0=1)=\frac{1}{2}$.
(ii) *If* H_1 *is true, then, as* $n\to\infty$, $\hat{\theta}_n-\theta_n\to 0$ *in probability.*

4. CALCULATION OF CRITICAL VALUES OF THE LIKELIHOOD RATIO TEST STATISTIC

Let $0<\alpha<1$, and define $Z_{n,\alpha}=\sup\{x:\mathrm{pr}(Z_n\leqslant x)\leqslant 1-\alpha\}$. Similarly to Theorem 1.3.2 and Corollary 1.3.1 of Csörgő & Horváth (1997), it is easy to see that the critical value $Z_{n,\alpha}$ can be approximated by

$$u(h,l,d;1-\alpha)=\sup\left[x:\kappa(x^2;d)\left\{(x^2-d)\log\frac{(1-h)(1-l)}{hl}+4\right\}=\alpha\right],\tag{4·1}$$

for suitable h and l, where $\kappa(t;d)$ is the $\chi^2(d)$ density function. From a simulation study, we found that a good choice for h and l is $h_\alpha(n)=l_\alpha(n)=(\log n/n)^{(2d+1)(1-\alpha)/2}$.

As pointed out by Csörgő & Horváth (1997), (4·1) usually provides very good approximations of critical values for small α. Tables 1 and 2 compare the critical values u^*, which are determined by Theorem 1, and $\hat{u}=u\{h_\alpha(n),l_\alpha(n),d;1-\alpha\}$ with empirical critical values based on 10 000 replicates from certain distributions, namely normal, exponential, binomial and Poisson distributions with change in mean. The model used corresponds to $d=1$ and $w(x,\beta)=\exp(\beta x)$. Table 2 displays critical values when

Table 1. *Comparison of critical values of the likelihood ratio statistic, with* $d=1$ *and* $w(x,\beta)=\exp(\beta x)$

n	$1-\alpha$	u^*	$\hat{u}$	z_{norm}	z_{exp}	z_{binom}	z_{pois}
20	0·90	3·1133	2·8290	2·8179	2·8179	2·8179	2·8179
	0·95	3·5993	3·1146	3·1062	3·1298	3·0306	3·0306
	0·99	4·6996	3·6519	3·6061	3·6246	3·6060	3·6060
50	0·90	3·1813	2·9440	3·1311	3·1311	2·8841	2·8944
	0·95	3·6171	3·2177	3·1311	3·1647	3·1311	3·1311
	0·99	4·6039	3·7386	3·6840	3·6931	3·6173	3·6280
100	0·90	3·2256	3·0140	3·0560	3·0004	2·9375	2·9032
	0·95	3·6374	3·2808	3·3467	3·3467	3·1895	3·1886
	0·99	4·5701	3·7921	3·7954	3·7123	3·6901	3·7552
500	0·90	3·3096	3·1401	3·0037	3·0401	2·9957	3·0287
	0·95	3·6862	3·3954	3·2768	3·3087	3·2501	3·2573
	0·99	4·5389	3·8897	3·7985	3·7985	3·7042	3·7984

Table 2. *Comparison of critical values of the likelihood ratio statistic, with $d=2$ and $w(x, \beta) = \exp\{\beta^{\mathrm{T}}\tau(x)\}$*

n	$1-\alpha$	u^*	$\hat{u}$	z_{norm}	z_{binorm}
20	0·90	3·5310	3·4994	3·6061	3·6061
	0·95	4·0169	3·7482	3·7234	3·6601
	0·99	5·1173	4·2290	4·2037	4·2589
50	0·90	3·6218	3·5980	3·6656	3·3701
	0·95	4·0576	3·8381	4·0980	3·7184
	0·99	5·0444	4·3061	4·2702	4·2152
100	0·90	3·6742	3·6594	3·6613	3·3731
	0·95	4·0861	3·8944	3·9855	3·6283
	0·99	5·0187	4·3548	4·4283	4·2312
500	0·90	3·7667	3·7715	3·7983	3·5448
	0·95	4·1432	3·9977	4·0203	3·7983
	0·99	4·9960	4·4446	4·6138	4·7488

$d=2$ and $w(x, \beta) = \exp\{\beta^{\mathrm{T}}\tau(x)\}$, where $\tau(x) = (x, x^2)^{\mathrm{T}}$, when samples are drawn from univariate normal distributions and both mean and variance may have changed, and where $\tau(x) = x = (x_1, x_2)^{\mathrm{T}}$ when samples come from bivariate normal distributions with a fixed covariance matrix and only the means may have changed.

Since the null hypothesis H_0 is equivalent to $F = G$ for any $0 < \theta_n < 1$, both permutation and bootstrap methods could be used to compute p-values. Clearly, the permutation method is preferred for small sample size since it provides an exact p-value if the number of permutation repetitions is sufficiently large (Efron & Tibshirani, 1993, Ch. 15). The permutation (bootstrap) algorithm is as follows. First, for each $b = 1, \ldots, B$, independently choose a permutation (bootstrap) sample $x^*_{b1}, \ldots, x^*_{bn}$ which is obtained without(with) replacement from $x_1, \ldots, x_n$. Then calculate the permutation (bootstrap) version $S^*_{bn} = S_n(x^*_{b1}, \ldots, x^*_{bn})$ of S_n. The p-values of S_n can be approximated by $\hat{P} = \#\{S^*_{bn} \geqslant S_n\}/B$. Examples of using resampling methods to approximate p-values can be found in § 7.

5. Power comparison: A simulation study

Consider a change in the mean. The proposed semiparametric test is compared with the nonparametric test $\tilde{S}^*_n$ in (3·5) based on the tied-down partial sum (Csörgő & Horváth, 1997, §§ 1.4, 2.1). The latter is equivalent to the parametric maximum likelihood ratio test when samples are drawn from normal distributions with unknown constant variance. Again normal, exponential, binomial and Poisson distributions were considered, with changes in the mean value at $k = n\theta_n = 0$, 10, 20, 30, 40 and 50, respectively, and 5000 samples are drawn from each of the four distributions. Using the critical values given in Table 1 with observed p-value 0·05, we simulated powers of the tests. The results are presented in Table 3, which shows that the semiparametric test is sensitive and robust. If samples are from normal distributions, then the proposed test is a little less powerful than the nonparametric test $\tilde{S}^*_n$, but otherwise the semiparametric test is more powerful than $\tilde{S}^*_n$.

Table 3. *Comparison of powers of the tests, with* $n=100$, $\alpha=0{\cdot}05$ *and* $E_F(X)=1$, $E_G(X)=2$

	Normal		Exponential		Binomial		Poisson	
$n\theta_n$	Semip	Nonp	Semip	Nonp	Semip	Nonp	Semip	Nonp
0	0·0500	0·0500	0·0500	0·0500	0·0500	0·0500	0·0500	0·0500
10	0·5076	0·5860	0·1746	0·0370	0·4302	0·2884	0·3690	0·1862
20	0·8222	0·8822	0·3552	0·0806	0·7680	0·6990	0·6624	0·5312
30	0·9304	0·9582	0·5398	0·2078	0·8900	0·8628	0·8236	0·7460
40	0·9628	0·9778	0·6358	0·3666	0·9314	0·9218	0·8724	0·8356
50	0·9676	0·9814	0·6846	0·5030	0·9456	0·9408	0·8994	0·8792
z	3·3467	3·1821	3·3467	3·7108	3·1895	3·2805	3·1886	3·3301

Semip, the semiparametric test; Nonp, the nonparametric test

6. Test of validity of the changepoint model

For each $k=1,\ldots,n-1$, let $\hat F_k(t)$ and $\hat G_k(t)$ denote the empirical distribution functions of samples $x_1,\ldots,x_k$ and $x_{k+1},\ldots,x_n$ respectively. Define semiparametric empirical distribution functions

$$\tilde F_k(x)=\sum_{i=1}^{n}\hat p_{ki}I_{[x_i\leqslant x]},\quad \tilde G_k(x)=\sum_{i=1}^{n}\hat q_{ki}I_{[x_i\leqslant x]}, \tag{6·1}$$

where

$$\hat p_{ki}=\frac{\hat w(k/n)}{(n-k)w\{x_i,\hat\beta(k/n)\}+k\hat w(k/n)},\quad \hat q_{ki}=\frac{w\{x_i,\hat\beta(k/n)\}}{\hat w(k/n)}\hat p_{ki}\quad (i=1,\ldots,n).$$

Let $\hat k=n\hat\theta_n$. Similarly to Qin & Zhang (1997), we propose to use the statistic

$$K_n=\sqrt{n}\left\{\hat\theta_n\sup_t|\tilde F_{\hat k}(t)-\hat F_{\hat k}(t)|+(1-\hat\theta_n)\sup_t|\tilde G_{\hat k}(t)-\hat G_{\hat k}(t)|\right\} \tag{6·2}$$

to test the validity of the semiparametric changepoint model. A significantly large value of K_n would indicate the invalidity of the model. To obtain p-values of K_n, the following bootstrap method can be used. If $\hat\theta_n=0$ or $\hat\theta_n=1$, then a bootstrap sample $x_1^*,\ldots,x_n^*$ is drawn from $x_1,\ldots,x_n$ with replacement. Otherwise, if $0<\hat\theta_n<1$, bootstrap samples $x_1^*,\ldots,x_{n\hat\theta_n}^*$ and $x_{n\hat\theta_n+1}^*,\ldots,x_n^*$ are drawn independently as samples from $\tilde F_{n\hat\theta_n}$ and $\tilde G_{n\hat\theta_n}$ respectively. By replacing the original sample $x_1,\ldots,x_n$ with the bootstrap sample $x_1^*,\ldots,x_n^*$, we can obtain the bootstrap version K_n^* of K_n, and p-values of K_n can be approximated by those of K_n^*. From the consistency of $\hat\theta_n$, the validity of the bootstrap method follows easily.

For univariate distributions, a visual diagnostic is to plot $\{\tilde F_{\hat k}(t),\hat F_{\hat k}(t)\}$ and $\{\tilde G_{\hat k}(t),\hat G_{\hat k}(t)\}$. If the changepoint is detected in the middle of the observations, substantial discrepancy in one of the plots would indicate the inadequateness of the model. If $\hat k$ is close to 1 or n, we may have to use the bootstrap method to approximate the p-value for the test.

7. Applications to some well-known datasets

Example 1. The Nile River data (Cobb, 1978) have been studied by, among many others, Cobb (1978), Dümbgen (1991) and Csörgő & Horváth (1997). All these authors obtained

$\hat{k} = n\hat{\theta}_n = 100 \times 0{\cdot}28 = 28$, which corresponds to the year 1898. The loglikelihood function $l(k/n) - n \log n$ and the estimated distribution functions $\tilde{F}_{\hat{k}}(t)$, $\hat{F}_{\hat{k}}(t)$, $\tilde{G}_{\hat{k}}(t)$ and $\hat{G}_{\hat{k}}(t)$ are plotted in Fig. 1. Figure 1 (b) indicates that the selected model fits the data perfectly. Based on 10 000 bootstrap replicates, the observed p-value of the model test with $w(x, \beta) = \exp(\beta x)$ and $d = 1$ is 0·6264, also indicating that the model is valid. We obtain the same changepoint estimate as above. The test statistic $Z_n = 7{\cdot}2085$ is also highly significant, with p-value $9{\cdot}9907 \times 10^{-5}$ from Theorem 1. From Table 1, Z_n is significant at level $\alpha = 0{\cdot}01$. An approximation of the p-value from (4·1) is $1{\cdot}3608 \times 10^{-10}$. Since Theorem 1 always gives a conservation rejection region, the actual p-value is not greater than $9{\cdot}9907 \times 10^{-5}$.

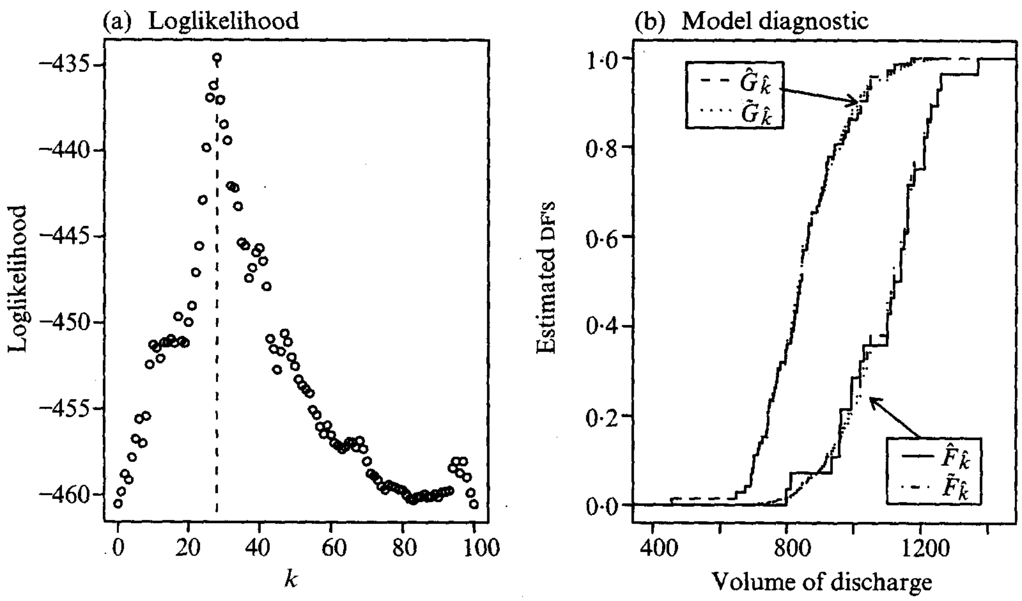

Fig. 1: Nile River data. (a) shows loglikelihood function and (b) estimated distribution functions, $\hat{F}_{\hat{k}}$, solid, $\tilde{F}_{\hat{k}}$, dash-dotted, $\hat{G}_{\hat{k}}$, dashed, and $\tilde{G}_{\hat{k}}$, dotted, plotted against Nile River annual volume of discharge (10^8 m^3).

Example 2. The stock-market price data, first studied by Hsu (1977, 1979) and then by Lee (1996) and Chen & Gupta (1997), consists of weekly closing values, $n = 162$, of the Dow Jones Industrial Average from 1 July 1971 to 2 August 1974. Based on a gamma distribution model and a nonparametric model for the rates of return, all these authors treated the rates of return as a series of independent random variables and arrived at the same conclusion that one variance shift has occurred, during the 89th week, 19–23 March 1973. The independence assumption was examined by Hsu (1977, 1979). We have fitted three exponential models with different $\tau(x)$ to this dataset. All these models have passed the model test. Figure 2 shows that this semiparametric model fits the data perfectly. Model 1, with $\tau(x) = x$, aims at detecting a possible change in the mean value. Model 2 is the same as Model 1, but is based on the data $x_1^2, \ldots, x_n^2$ instead of $x_1, \ldots, x_n$. It tests whether or not $E(X^2)$ has changed. Model 3, with $\tau(x) = (x, x^2)^{\mathrm{T}}$, can detect changes in mean and variance. The results are summarised in Table 4. The number of repetitions is 5000. Table 4 shows that there was no significant change in the mean rate of stock return but that the variance changed significantly.

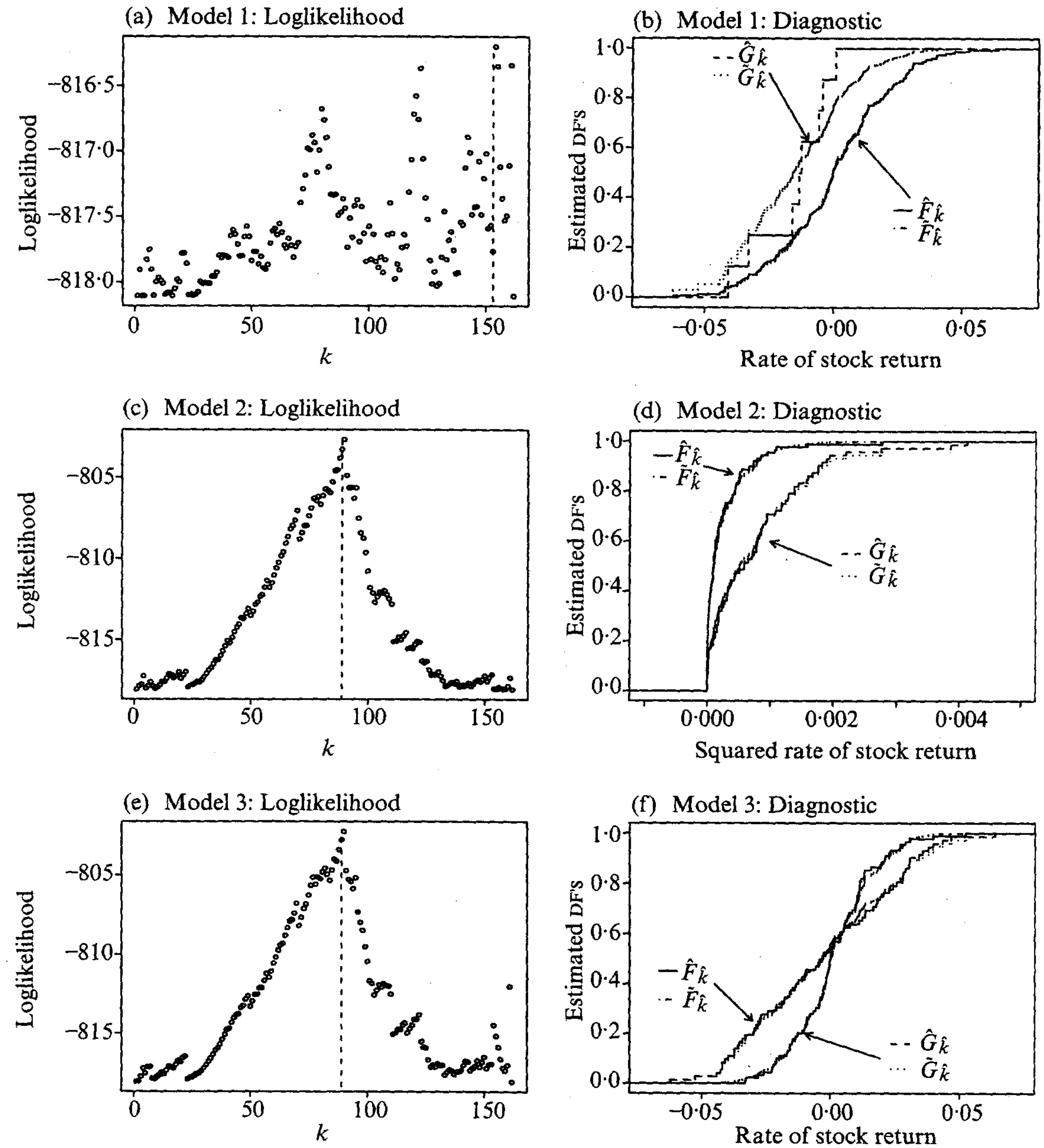

Fig. 2. Stock price data. (a), (c) and (e) show loglikelihood functions and (b), (d) and (f) show estimated distribution functions, $\hat{F}_{\hat{k}}$, solid, $\tilde{F}_{\hat{k}}$, dash-dotted, $\hat{G}_{\hat{k}}$, dashed, and $\tilde{G}_{\hat{k}}$, dotted.

Table 4: *Stock price data. Results of data analysis*

Model	K_n	Pv-K_n	$n\hat{\theta}_n$	Z_n	Pv-(3·3)	Pv-(4·1)	Pv-P	Pv-B
1	0·2882	0·4314	153	1·9516	0·6674	0·5263	0·6318	0·6410
2	0·5851	0·4584	89	5·5493	0·0017	$1{\cdot}0 \times 10^{-6}$	0	0
3	0·5529	0·2760	89	5·6269	0·0033	$6{\cdot}2 \times 10^{-5}$	0	0

Pv-K_n, p-value of bootstrap test of the model; Pv-(3·3), Pv-(4·1), Pv-P and Pv-B are the p-values based on (3·3), (4·1), the permutation test and the bootstrap test respectively.

ACKNOWLEDGEMENT

Part of this work is from the author's Ph.D. thesis at the University of Toledo. The author is grateful to his advisor Professor Biao Zhang for his advice and help. The author would like to thank the editor, an associate editor and a referee for the valuable and helpful comments and constructive suggestions about the paper.

APPENDIX

Proofs

Define

$$\lambda_1(\theta) = -w_0^{-1}E_F\{a(X,\eta_0)\log a(X,\eta)\} + \theta\log w(\theta),$$

$$\lambda_2(\theta) = (\theta_0 - \theta\wedge\theta_0)E_F[\log w\{X,\beta(\theta)\}] + (1-\theta_0\vee\theta)E_F\left[\frac{w(X,\beta_0)}{w_0}\log w\{X,\beta(\theta)\}\right],$$

and $\lambda(\theta) = \lambda_1(\theta) + \lambda_2(\theta)$, where, for fixed $\theta_0 \in (0,1]$ and $\theta \in (0,1)$, $w(\theta)$ and $\beta(\theta)$ satisfy the equations

$$g_1\{\theta, w(\theta), \beta(\theta)\} = 0, \quad g_2\{\theta, w(\theta), \beta(\theta)\} = 0. \tag{A·1}$$

The existence, uniqueness and differentiability of $w(\theta)$ and $\beta(\theta)$ are ensured by the implicit function theorem and Assumption 2. Thus we can define $\{w(0), \beta^{\mathrm{T}}(0)\} = \{w(0+), \beta^{\mathrm{T}}(0+)\}$ and $\{w(1), \beta^{\mathrm{T}}(1)\} = \{w(1-), \beta^{\mathrm{T}}(1-)\}$. In the above and the following notation, we suppress the dependence of quantities on θ_0. It is evident that $\lambda(\theta)$ is a continuous function of θ and is also twice differentiable with respect to θ on $(0,1)$ except at θ_0.

In order to prove Theorem 1, we need the following lemmas whose proofs are similar to, but a little less tedious than, those of Lemmas 1.2.1 and 1.2.2 of Csörgő & Horváth (1997), respectively, and are therefore omitted.

LEMMA A·1. *If H_0 holds, then for any $\varepsilon > 0$ and $\delta > 0$ there exist $K_1 = K_1(\varepsilon,\delta)$ and $N_1 = N_1(\varepsilon,\delta)$ such that, if $K > K_1$ and $N > N_1$,*

$$\mathrm{pr}\left\{\max_{K\leqslant k\leqslant n-K} |(\hat{w}_k, \hat{\beta}) - (1,0)| > \varepsilon\right\} \leqslant \delta. \tag{A·2}$$

LEMMA A·2. *If H_0 and Assumptions 1–3 hold, then for all $\delta > 0$ we can find $C_1 = C_1(\delta)$, $K_2 = K_2(\delta)$ and $N_2 = N_2(\delta)$ such that*

$$\mathrm{pr}\left\{(n/\log\log n)^{1/2}\max_{K\leqslant k\leqslant n-K}|\hat{w}_k - 1| > C_1\right\} \leqslant \delta, \tag{A·3}$$

$$\mathrm{pr}\left\{(n/\log\log n)^{1/2}\max_{K\leqslant k\leqslant n-K}|\hat{\beta}_k - 0| > C_1\right\} \leqslant \delta, \tag{A·4}$$

$$\mathrm{pr}\left\{\max_{K\leqslant k\leqslant n-K}|\hat{w}_k - 1| > \frac{C_1}{n^{1/2}}\right\} \leqslant \delta, \quad \mathrm{pr}\left\{\max_{K\leqslant k\leqslant n-K}|\hat{\beta}_k - 0| > \frac{C_1}{n^{1/2}}\right\} \leqslant \delta, \tag{A·5}$$

if $K \geqslant K_2$ and $n \geqslant N_2$. Also

$$|(\hat{w}_k, \hat{\beta}_k) - (1,0)| = \mathcal{O}_P(n^{-1/2}). \tag{A·6}$$

Proof of Theorem 1. Two-term Taylor expansion of $\psi_1(k/n, \hat{w}_k, \hat{\beta}_k)$ and $\psi_2(k/n, \hat{w}_k, \hat{\beta}_k)$ at $(\omega_0, \beta_0) = (1, 0)$ yields

$$\begin{pmatrix} \hat{w}_k - w_0 \\ \hat{\beta}_k - \beta_0 \end{pmatrix} = A_n^{-1}(k/n, w_0, \beta_0)\begin{pmatrix} n^{-1}\psi_1(k/n, w_0, \beta_0) + \varepsilon'_{nk} \\ n^{-1}\psi_2(k/n, w_0, \beta_0) + \varepsilon''_{nk}\end{pmatrix}, \tag{A·7}$$

where both ε'_{nk} and $\|\varepsilon''_{nk}\|$ are $k/n(1-k/n)o_P(|\hat{w}_k - w_0| + \|\hat{\beta}_k - \beta_0\|)$ and

$$A_n(\theta, w, \beta) = -\frac{1}{n}\begin{pmatrix} \dfrac{\partial\psi_1(\theta, w, \beta)}{\partial w} & \dfrac{\partial\psi_1(\theta, w, \beta)}{\partial\beta^{\mathrm{T}}} \\ \dfrac{\partial\psi_2(\theta, w, \beta)}{\partial w} & \dfrac{\partial\psi_2(\theta, w, \beta)}{\partial\beta^{\mathrm{T}}} \end{pmatrix}.$$

Define $D(x) = \Delta(x, \beta_0) - E_F\{\Delta(x, \beta_0)\}$ and

$$B_n = \begin{pmatrix} 1 & \bar{\tau}_n^{\mathrm{T}} \\ \bar{\tau}_n & n^{-1}\sum_{i=1}^n \tau_i\tau_i^{\mathrm{T}} + r_{nk} \end{pmatrix},$$

where $r_{nk} = \{n\sum_{i=1}^k D(x_i) - k\sum_{i=1}^n D(x_i)\}/\{k(n-k)\}$,

$$\Delta(x, \beta) = w''_{\beta\beta}(x, \beta)w(x, \beta) - w'_\beta(x, \beta)\{w'_\beta(x, \beta)\}^{\mathrm{T}}, \quad w''_{\beta\beta}(x, \beta) = \frac{\partial^2 w(x, \beta)}{\partial\beta\,\partial\beta^{\mathrm{T}}}.$$

It is easy to see that

$$A_n\left(\frac{k}{n}, w_0, \beta_0\right) = \frac{k}{n}\left(1 - \frac{k}{n}\right)B_n, \quad B_n^{-1} = \begin{pmatrix} U_1 & U_1\bar{\tau}_n^{\mathrm{T}}U_2 \\ U_2\bar{\tau}_n U_1 & (\Sigma_n + r_{nk})^{-1} \end{pmatrix},$$

where

$$U_1 = \{1 - \bar{\tau}_n^{\mathrm{T}}U_2\bar{\tau}_n\}^{-1}, \quad U_2 = \left\{\frac{1}{n}\sum_{i=1}^n \tau_i\tau_i^{\mathrm{T}} + r_{nk}\right\}^{-1}$$

and Σ_n is given in (3·6). Thus

$$\begin{pmatrix} \hat{w}_k - w_0 \\ \hat{\beta}_k - \beta_0 \end{pmatrix} = \frac{n}{k(n-k)}B_n^{-1}\begin{pmatrix} n\varepsilon'_{nk} \\ G(k) + n\varepsilon''_{nk} \end{pmatrix}, \tag{A·8}$$

where $G(k)$ is defined in (3·7). It is also easy to see that, for any $\varepsilon > 0$, there exists a $K = K(\varepsilon)$ such that, when $n > n(\varepsilon)$, $\max_{K \leqslant k \leqslant n-K}|(r_{nk})_{uv}| < \varepsilon$ for $1 \leqslant u, v \leqslant d$. Three-term Taylor expansion of $2n\Lambda_k$ at $(w_0, \beta_0) = (1, 0)$, (A·7) and some tedious algebra yield

$$2n\Lambda_k = \frac{n}{k(n-k)}G^{\mathrm{T}}(k)\Sigma_n^{-1}G(k) + R_{nk}, \tag{A·9}$$

where

$$R_{nk} = \frac{k}{n}\left(1 - \frac{k}{n}\right)o_P(|\hat{w}_k - w_0|^2 + \|\hat{\beta}_k - \beta_0\|^2).$$

This together with Lemma A·2 ensures that, for all $0 \leqslant \eta < \frac{1}{2}$,

$$n^\eta \max_{1 \leqslant k < n}\left\{\frac{k}{n}\left(1 - \frac{k}{n}\right)\right\}^\eta\left|2n\Lambda_k - \frac{n}{k(n-k)}G^{\mathrm{T}}(k)\Sigma_n^{-1}G(k)\right| = \mathcal{O}_P(1), \tag{A·10}$$

$$\max_{1 \leqslant k < n}\frac{k(n-k)}{n^2}\left|2n\Lambda_k - \frac{n}{k(n-k)}G^{\mathrm{T}}(k)\Sigma_n^{-1}G(k)\right| = \mathcal{O}_P\{n^{-1/2}(\log\log n)^{3/2}\}. \tag{A·11}$$

Using the same arguments as used in the proof of Theorem 1.3.1 of Csörgő & Horváth (1997), we can prove the theorem. □

Proof of Theorem 2. Based on one- and two-term Taylor expansions of Λ_k as a function of $(\hat{\beta}_k, \hat{w}_k)$ at $(\beta_k, w_k) = \{\beta(k/n), w(k/n)\}$ and very extensive algebra, we can show that

$$\Lambda_k + \frac{1}{n}\left[\sum_{i=1}^k \log\left\{\frac{k}{n} + \frac{1-k}{n}\frac{w(x_i, \beta_k)}{w_k}\right\} + \sum_{i=k+1}^n \log\left\{\frac{1-k}{n} + \frac{k}{n}\frac{w_k}{w(x_i, \beta_k)}\right\}\right] = \mathcal{O}_P(n^{-1}).$$

By the law of large numbers and the central limit theorem, the almost sure convergence of Λ_k follows from the above expansion. From $g_1\{w(\theta), \beta(\theta), \theta\} = 0$, it follows that

$$E_F\left[\frac{a(X, \eta_0)w(\theta)}{a\{X, \eta(\theta)\}}\right] = E_F\left[\frac{a(X, \eta_0)w\{X, \beta(\theta)\}}{a\{X, \eta(\theta)\}}\right] = w_0. \tag{A·12}$$

Clearly, $w(\theta_0) = w_0$ and $\beta(\theta_0) = \beta_0$. For fixed θ_0, we write $w_k = w(k/n)$ and $\beta_k = \beta(k/n)$. From the implicit function theorem, it follows that $w'(\theta)$ and $\beta'(\theta)$ are determined by

$$\begin{pmatrix} w'(\theta) \\ \beta'(\theta) \end{pmatrix} = A^{-1}\begin{pmatrix} \partial g_1/\partial\theta \\ \partial g_2/\partial\theta \end{pmatrix}. \tag{A·13}$$

Since $A > 0$,

$$A_{11} = \frac{\theta(1-\theta)}{w_0 w}E_F\left\{\frac{a(X, \eta_0)}{a^2(X, \eta)}w(X, \beta)\right\}, \quad A_{21} = A_{12}^{\mathrm{T}} = -\frac{\theta(1-\theta)}{w_0}E_F\left\{\frac{a(X, \eta_0)}{a^2(X, \eta)}w'_\beta(X, \beta)\right\},$$

$$A_{22} = -\frac{\partial g_2}{\partial\beta^{\mathrm{T}}} = \frac{1}{w_0}\left[w\theta(1-\theta)E_F\left\{\frac{a(X, \eta_0)}{a^2(X, \eta)}w''_{\beta\beta}(X, \beta)\right\} - E_F\{B(X)\}\right] > 0,$$

where

$$B(X) = \left\{\frac{(\theta_0 - \theta_0 \wedge \theta)w_0 + (1 - \theta_0 \vee \theta)w(X, \beta_0)}{w^2(X, \beta)} - \frac{(1-\theta)^2 a(X, \eta_0)}{a^2(X, \eta)}\right\}\Delta(X, \beta).$$

Thus

$$\left.\frac{\partial g_1(\theta, w, \beta)}{\partial\theta}\right|_{w = w(\theta), \beta = \beta(\theta)} = \frac{1}{w(\theta)}\left\{1 - \frac{w^2(\theta)}{\theta(1-\theta)}A_{11}\right\} \tag{A·14}$$

and consequently

$$1 \geqslant \frac{w(\theta)}{w_0}E_F\left[\frac{a(X, \eta_0)}{a^2(X, \eta)}w\{X, \beta(\theta)\}\right] = \frac{w^2(\theta)}{\theta(1-\theta)}A_{11}. \tag{A·15}$$

From (A·1), (A·12), (3·2) and simple algebra, it follows that

$$\lambda''(\theta) = \frac{w'(\theta)}{w(\theta)} + \{\beta'(\theta)\}^{\mathrm{T}}V, \tag{A·16}$$

where

$$V = -E_F\left[\frac{\{I(\theta < \theta_0)w_0 + I(\theta > \theta_0)w(X, \beta_0)\}w'_\beta(X, \beta)}{w_0 w(X, \beta)}\right]. \tag{A·17}$$

Clearly

$$\frac{\partial g_2(\theta, w, \beta)}{\partial\theta} = V + \frac{w(\theta)}{w_0}E_F\left\{\frac{a(X, \eta_0)}{a^2(X, \eta)}w'_\beta(X, \beta)\right\} = V - \frac{w(\theta)}{\theta(1-\theta)}A_{21}. \tag{A·18}$$

Combining (A·14)–(A·18), we obtain

$$\lambda''(\theta) = \left(\frac{1}{w(\theta)}, V^{\mathrm{T}}\right)A^{-1}\begin{pmatrix} 1/w(\theta) \\ V \end{pmatrix} - \frac{1}{\theta(1-\theta)} > \frac{1}{w^2(\theta)}A_{11}^{-1} - \frac{1}{\theta(1-\theta)} \geqslant 0.$$

Thus $\lambda''(\theta) \geqslant 0$ on $(0, \theta_0)$ and $(\theta_0, 1)$. It is easy to see that $\lambda(0) = \lambda(1) = 0$ and $\lambda(\theta_0) > 0$. Therefore, the theorem is proved. □

Proof of Theorem 3. (i) The proof is based on Theorem 2 and is similar to the proof of Theorem 1.6.1 of Csörgő & Horváth (1997, p. 51).

(ii) It follows from Theorem 2 that $\max_{1 \leqslant k < n} \Lambda_k \to \lambda(\theta_0)$ almost surely as $n \to \infty$ and, for any δ such that $0 < \delta < \min\{\theta_0, 1 - \theta_0\}$,

$$\max_{|k - n\theta_0| > \delta n} \Lambda_k \to \max_{|\theta - \theta_0| > \delta} \lambda(\theta) < \lambda(\theta_0)$$

almost surely. Therefore, $\lim_{n\to\infty} \mathrm{pr}\{|\hat{\theta}_n - \theta_0| > \delta\} = 0$. This proves that $\hat{\theta}_n - \theta_n \to 0$ in probability since we have assumed that $\lim_{n\to\infty} \theta_n = \theta_0$. □

References

ALBERT, A. & ANDERSON, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71**, 1–10.

CARLSTEIN, E. (1988). Nonparametric changepoint estimation. *Ann. Statist.* **16**, 188–97.

CHEN, J. & GUPTA, A. (1997). Testing and locating variance changepoints with application to stock prices. *J. Am. Statist. Assoc.* **92**, 739–47.

COBB, G. W. (1978). The problem of the Nile: conditional solution to a changepoint problem. *Biometrika* **65**, 243–51.

CSÖRGŐ, M. & HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. New York: John Wiley.

DARKHOVSKH, B. S. (1976). A non-parametric method for the posterior detection of the 'disorder' time of a sequence of independent random variables. *Theory Prob. Applic.* **21**, 178–83.

DÜMBGEN, L. (1991). The asymptotic behaviour of some nonparametric changepoint estimators. *Ann. Statist.* **19**, 1471–95.

EFRON, B. & TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

HSU, D. A. (1977). Tests for variance shift at an unknown time point. *Appl. Statist.* **26**, 279–84.

HSU, D. A. (1979). Detecting shifts of parameter in gamma sequences with applications to stock prices and air traffic flow analysis. *J. Am. Statist. Assoc.* **74**, 31–46.

JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond.* A **186**, 453–61.

LEE, C.-B. (1996). Nonparametric multiple changepoint estimators. *Statist. Prob. Lett.* **27**, 295–304.

LEVIN, B. & KLINE, J. (1985). The cusum test of homogeneity with an application in spontaneous abortion epidemiology. *Statist. Med.* **4**, 469–88.

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–49.

QIN, J. (1993). Empirical likelihood in biased sample problems. *Ann. Statist.* **21**, 1182–96.

QIN, J. & LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300–25.

QIN, J. & ZHANG, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika* **84**, 609–18.

SANTNER, T. J. & DUFFY, D. E. (1986). A note on A. Albert and J. A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **73**, 755–8.

VARDI, Y. (1982). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10**, 616–20.

WORSLEY, K. J. (1986). Confidence regions and test for a changepoint in a sequence of exponential family random variables. *Biometrika* **73**, 91–104.

YAO, Q. (1993). Tests for changepoints with epidemic alternatives. *Biometrika* **80**, 179–91.

ZHANG, B. (1997). Empirical likelihood confidence intervals for *M*-functionals in the presence of auxiliary information. *Statist. Prob. Lett.* **32**, 87–97.

ZHANG, B. (1999). A chi-squared goodness-of-fit test for logistic regression models based on case-control data. *Biometrika* **86**, 531–9.

[*Received February* 2003. *Revised February* 2004]