STATISTICS IN MEDICINE
Statist. Med. **19**, 335–351 (2000)

PERMUTATION TESTS FOR JOINPOINT REGRESSION WITH APPLICATIONS TO CANCER RATES

HYUNE-JU KIM[1]*, MICHAEL P. FAY[2], ERIC J. FEUER[2] AND DOUGLAS N. MIDTHUNE[3]

[1] *Syracuse University, Department of Mathematics, 215 Carnegie Building, Syracuse University, Syracuse, NY 13244-1150, U.S.A.*
[2] *National Cancer Institute, Executive Plaza North, Suite 313, 6130 Executive Boulevard, Bethesda, MD 20892, U.S.A.*
[3] *National Cancer Institute, Executive Plaza North, Suite 344, 6130 Executive Boulevard, Bethesda, MD 20892, U.S.A.*

SUMMARY

The identification of changes in the recent trend is an important issue in the analysis of cancer mortality and incidence data. We apply a joinpoint regression model to describe such continuous changes and use the grid-search method to fit the regression function with unknown joinpoints assuming constant variance and uncorrelated errors. We find the number of significant joinpoints by performing several permutation tests, each of which has a correct significance level asymptotically. Each *p*-value is found using Monte Carlo methods, and the overall asymptotic significance level is maintained through a Bonferroni correction. These tests are extended to the situation with non-constant variance to handle rates with Poisson variation and possibly autocorrelated errors. The performance of these tests are studied via simulations and the tests are applied to U.S. prostate cancer incidence and mortality rates. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

In studying trend data such as cancer mortality and incidence data one is frequently concerned with detecting a change in the recent trend. The joinpoint regression model, which is composed of a few continuous linear phases, is often useful to describe changes in trend data. The joinpoint regression model for the observations, $(x_1, y_1), \ldots, (x_n, y_n)$, where $x_1 \leqslant \cdots \leqslant x_n$ without loss of generality, may be written as

$$E[y|x] = \beta_0 + \beta_1 x + \delta_1 (x - \tau_1)^+ + \cdots + \delta_k (x - \tau_k)^+ \tag{1}$$

where the τ_k's are the unknown joinpoints and $a^+ = a$ for $a > 0$ and 0 otherwise.

This type of non-linear regression model has been studied by many authors and has been named in the literature as piecewise regression, segmented regression, broken line regression, and multi-phase regression with the continuity constraint.[1–6]

Assuming that $k = 1$ and ρ such that $\tau_1 \in (x_\rho, x_{\rho+1})$ is known, Sprent[3] derived the likelihood ratio statistic for testing $H_0: \tau_1 = \tau_{1,0}$ with applications in biometry. Hinkley[1] studied asymptotic properties of the maximum likelihood estimators of the parameters and developed a procedure to

* Correspondence to: Hyune-Ju Kim, Syracuse University, Department of Mathematics, 215 Carnegie Building, Syracuse, NY 13244-1150, U.S.A. E-mail: hjkim@mailbox.syr.edu

CCC 0277–6715/2000/030335–17$17.50

Received April 1998
Accepted April 1999

construct confidence regions. For a general case where the model allows polynomial segments, Lerman[4] proposed a grid search method to fit segmented regression curves. He assumed the distribution of the likelihood ratio statistic to be an F-distribution, and used it to test hypotheses and to construct an approximate confidence region for the joinpoints. Knowles and Siegmund[2] applied the method suggested by Hotelling[7] to make an inference on a joinpoint in regression. They derived an approximate upper bound for the p-value of the likelihood ratio test and constructed confidence regions based on the likelihood ratio statistic. Their approximation is the first analytic solution to approximate the distribution of the likelihood ratio statistic, but it only provides an upper bound and the accuracy is reasonable only when the regression parameters are assumed to be known. Furthermore it is not known how accurate the approximation would be when the underlying distribution is not normal, such as Poisson.

The joinpoint regression discussed here is different from problems of multi-phase regression where one does not constrain the regression functions to be continuous at the changepoints. The sampling distribution of the likelihood ratio statistics in this case involves quite different mathematical properties.[8–10]

Motivated by problems to identify changes in the recent trend of cancer mortality data, this paper proposes a procedure to determine the number of joinpoints in (1) and to estimate parameters including the joinpoints. For example, to determine up to two joinpoints, we perform the following procedure. We first perform the test of the hypothesis of no change, H_0:$E[y|x] = \beta_0 + \beta_1 x$ against the alternative of two joinpoints, H_1:there exist τ_1 and τ_2 ($\tau_1 < \tau_2$) such that $E[y|x] = \beta_0 + \beta_1 x + \delta_1(x - \tau_1)^+ + \delta_2(x - \tau_2)^+$. If the null hypothesis is rejected, then the similar procedure is applied to test the null hypothesis of one joinpoint against the alternative of two joinpoints. Otherwise, we test for the null hypothesis of no change against the alternative of one joinpoint. A secondary consideration after the joinpoint model is identified is to obtain confidence regions for the parameters including the joinpoints.

This paper is organized as follows. In Section 2, the test statistic is introduced. The test statistic is obtained by the grid search method suggested by Lerman[4] and its p-value is computed using the permutation procedure. In Section 3, we perform simulation studies to assess the performance of the permutation test. Our main concern is on the power and the robustness of the p-value when there are some departures from assumptions such as heteroscedasticity and autocorrelated errors. Finally, in Section 4, we apply our procedure to the U.S. prostate cancer incidence and mortality data to study recent swings in the incidence and mortality rates. Section 5 includes discussion and future research problems.

2. INFERENCE ON JOINPOINTS

In this section we give the details of the approximate permutation test. We wish to test the hypotheses

$$H_0 : \text{there are } k_0 \text{ joinpoints}$$

versus

$$H_1 : \text{there are } k_1 \text{ joinpoints.}$$

For a model with k joinpoints, the ith response is

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \delta_1(x_i - \tau_1)^+ + \cdots + \delta_k(x_i - \tau_k)^+ + \varepsilon_i^{(k)} \\ &= \mu_i^{(k)} + \varepsilon_i^{(k)} \end{aligned}$$

where $\varepsilon_i^{(k)}$ is the error and $\mu_i^{(k)}$ is defined implicitly. For all permutation tests discussed here we assume that under the null model, $E(\varepsilon_i^{(k_0)}) = 0$. We consider the simplest case in Section 2.1, where under the null, $\text{var}(\varepsilon_i^{(k_0)}) = \sigma^2$ for all i and $\text{cov}(\varepsilon_i^{(k_0)}, \varepsilon_j^{(k_0)}) = 0$ for all $i \neq j$. In Section 2.2 we continue to assume that all errors are uncorrelated but we allow the variance of $\varepsilon_i^{(k_0)}$ to depend on i, as is the case with Poisson errors. The problem of autocorrelated errors will be addressed in Section 2.3. Finally, in Section 2.4, we describe our computing algorithm and discuss how to adjust the significance level to account for the fact that several tests are performed to determine the number of joinpoints in the data.

2.1. Constant Variance and Uncorrelated Errors

We describe the approximate permutation test in several steps:

1. We fit the null hypothesis model.
2. We permute the residuals from the null model adding them back onto the means from the null model to obtain N_P 'permutation' data sets.
3. For each of these data sets we fit the alternative model and calculate a scalar goodness-of-fit measure.
4. The p-value is determined from the permutational distribution of the goodness-of-fit statistics.

We now describe the details.

First, we find the least squares estimates of the parameters under the null hypothesis model, that is, we find the values of

$$\beta_0, \beta_1, \delta_1, \ldots, \delta_{k_0}, \tau_1, \ldots, \tau_{k_0}$$

that minimize

$$Q = \sum_{i=1}^{n} (y_i - \mu_i^{(k_0)})^2.$$

Following Lerman,[4] we use the grid search method over the $\tau_1, \ldots, \tau_{k_0}$, where at each step of the grid search the least squares estimates for the other parameters are found by usual linear model methods.

The second step is to permute the residuals and add them back to the null modelled means. Let the $n \times 1$ vector of residuals from this null model be denoted, $\hat{\varepsilon}^{(k_0)}(y) = \hat{\varepsilon}^{(k_0)}$, with ith element, $\hat{\varepsilon}_i^{(k_0)} = y_i - \hat{\mu}_i^{(k_0)}$, where $y' = [y_1, \ldots, y_n]$. Let $\pi_a' = [\pi_{a1}, \ldots, \pi_{an}]$ be an $n \times 1$ vector of permutations of the integers from 1 to n. The permuted data set associated with π_a has the same set of covariates as the original data, and permuted responses are of the form

$$y_{(a)}' = \hat{\mu}^{(k_0)'} + [\hat{\varepsilon}_{\pi_{a1}}^{(k_0)}, \ldots, \hat{\varepsilon}_{\pi_{an}}^{(k_0)}]$$

where $\hat{\mu}^{(k_0)'} = [\hat{\mu}_1^{(k_0)}, \ldots, \hat{\mu}_n^{(k_0)}]$. To perform the complete enumeration of the permutation test, we would create $n!$ permuted data sets. Since this will in general be too large a number, we take Monte Carlo samples from these $n!$ data sets.

The third step is to fit the alternative model to the permuted data sets. These models are also fitted by least squares using the grid search. Let the vector of residuals from the ath permutation data set be denoted by $\hat{\varepsilon}^{(k_1)}(y_{(a)})$. We use the F-statistic as a goodness-of-fit measure. Because we are only interested in the relative values of the F-statistic for different permutations, we compare

the simpler statistic

$$T(y_{(a)}) = \frac{[\hat{\varepsilon}^{(k_0)}(y_{(a)})]'[\hat{\varepsilon}^{(k_0)}(y_{(a)})]}{[\hat{\varepsilon}^{(k_1)}(y_{(a)})]'[\hat{\varepsilon}^{(k_1)}(y_{(a)})]}$$

which is a monotonic transformation of the F-statistic.

Finally, we perform the Monte Carlo calculation.[13] We take a sample of $N_P - 1$ permutation values of $T(y_{(a)})$, $a = 1, \ldots, N_P - 1$, and add the statistic from original data, $T(y) \equiv T(y_{(0)})$. We see how extreme $T(y)$ is in this sample of N_P values. The p-value is

$$p = \frac{\text{number of times that}[T(y_{(a)}) \geqslant T(y)] \text{ for } a \in \{0, 1, \ldots, N_P - 1\}}{N_P}.$$

By choosing N_P large enough, we can essentially achieve as many significant digits as we need for the p-value.

We have referred to this test as an approximate permutation test. That is not only because of the Monte Carlo nature of the calculation, but due to the fact that the condition on the permutation test is only met asymptotically. In order to perform a true permutation test using the residuals, the standard condition is that the residuals be exchangeable under the null hypothesis.[14] In other words, the distribution of $\hat{\varepsilon}^{(k_0)}(y_{(a)})$ is the same as the distribution of $\hat{\varepsilon}^{(k_0)}(y_{(b)})$ for all permutations π_a and π_b. A slightly less restrictive condition is that

$$\Pr[T(y_{(a)}) \leqslant t] = \Pr[T(y_{(b)}) \leqslant t] \text{ under } H_0 \text{ for all } t \text{ and } a, b.$$

These conditions are not met for finite samples because the residuals are correlated due to the constraint that they must sum to zero. To show that these conditions are met as n goes to infinity, it suffices to prove that the least squares estimator of $\mu^{(k_0)}$ is consistent under the null hypothesis. Feder[15] showed in Theorem 3.6 that the least squares estimators of the parameters

$$\beta_0, \beta_1, \delta_1, \ldots, \delta_{k_0}, \tau_1, \ldots, \tau_{k_0}$$

are consistent under suitable identifiability assumptions, which tacitly assume that no two adjacent linear segments are identical. The equally spaced predictor variable, which is our main interest, was used as an example that satisfies the assumptions of Theorem 3.6 in Feder. Since the errors are exchangeable by the assumption, the residuals, which are consistent estimators of the errors, are then asymptotically exchangeable.

2.2. Non-constant Variance

Consider the case where the variance of the errors depends on i, $\text{var}(\varepsilon_i^{(k_0)}) = V_i^{(k_0)} = V_i$. We must make some modifications to the permutation test.

First, instead of using the least squares criterion to find the parameters, we use a weighted least squares criterion, where the ith weight is V_i^{-1}. Second, we try to form more homoscedastic residuals by scaling them. In other words, instead of permuting the residuals, we permute the scaled residuals, where the ith scaled residual is

$$\tilde{\varepsilon}_i^{(k_0)} = \frac{\hat{\varepsilon}_i^{(k_0)}}{\sqrt{\hat{V}_i}}$$

where $\hat{V}_i$ is an appropriate estimate of V_i.

Then the permuted responses are obtained as

$$y'_{(a)} = \hat{\mu}^{(k_0)'} + [\tilde{\varepsilon}^{(k_0)}_{\pi_{a1}}\sqrt{\hat{V}_1}, \ldots, \tilde{\varepsilon}^{(k_0)}_{\pi_{an}}\sqrt{\hat{V}_n}].$$

If the Y_i are assumed to have a Poisson distribution with mean μ_i, then the variance of Y_i is estimated as the observed value of Y_i, y_i, and the weight matrix is the diagonal matrix with $1/y_i$ on the diagonal. When $y_i = 0$ we may estimate the weight with $1/\gamma$ for some positive constant γ and $1/0{\cdot}5 = 2$ was used in our program. To avoid this problem with zeros, we could use the iteratively reweighted least squares algorithm where the final weight for the ith individual is $1/\hat{\mu}_i^{(k_0)}$. However, because of the grid search, that method is too computationally intensive.

For cancer rates analysis, we often use the log of the rates. In this case, we let $Y_i = \log(Z_i/n_i)$ be the log of cancer death rates at time i for a given age group of size n_i, where Z_i are the cancer deaths at time i for the given age group and are assumed to follow a Poisson distribution with mean $n_i\lambda_i$. By a Taylor series expansion, the mean of Y_i is approximately $\log \lambda_i$ and the variance of Y_i is approximately $1/(n_i\lambda_i)$. The variance estimator is not constant in general and can be estimated as $\hat{V}_i = 1/z_i$, where z_i is the observed cancer deaths among the n_i subjects at risk. For age-adjusted rates, we can define $Y_i = \log(\sum_{j=1}^{A} c_j Z_{ij}/n_{ij})$, where Z_{ij} and n_{ij} are the cancer counts and the population size, respectively, at time i for age group j ($j = 1, \ldots, A$), and c_j is the known age standard. If we assume that Z_{ij} follows a Poisson distribution, then the variance of Y_i can be estimated as

$$\hat{V}_i = \frac{\sum_{j=1}^{A} c_j^2 z_{ij}/n_{ij}^2}{(\sum_{j=1}^{A} c_j z_{ij}/n_{ij})^2}$$

where z_{ij} is the observed cancer counts at time i for age group j. For these cases, the weighted least squares with $w_i = 1/\hat{V}_i$ would yield the adjusted p-value of the test.

2.3. Correlated Errors

When the observations are serially correlated, the residuals are not asymptotically exchangeable even for a simple linear regression model. If the correlation matrix of the error is Σ for a model with k joinpoints, then at each step of the grid search, the parameters that are not joinpoints are found by weighted least squares. Let

$$\theta \equiv [\beta_0, \beta_1, \delta_1, \ldots, \delta_k].$$

Then given $\tau_1, \ldots, \tau_k$, $\hat{\theta} = (X'WX)^{-1}X'Wy$, where $W = \hat{\Sigma}^{-1}$ and X is an $n \times (k+2)$ matrix with ith row equal to

$$[1, x_i, (x_i - \tau_1)^+, \ldots, (x_i - \tau_k)^+].$$

After the grid search is completed, the scaled residuals, $\hat{\Sigma}^{-1/2}(y - X\hat{\theta})$, are asymptotically exchangeable, where $\hat{\Sigma}^{-1/2}$ is a symmetric square root of $\hat{\Sigma}^{-1}$. Then the scaled residuals are permuted, are multiplied by $\hat{\Sigma}^{1/2}$, and are finally added back to $\hat{\mu}$ to generate the permuted responses. In practice, any reasonable estimate of Σ would work for the weighted least squares permutation test. For autocorrelated data with lag 1, the ijth element of Σ is $\Sigma_{ij} = \sigma^2\phi^{|i-j|}/(1-\phi)$, where $\sigma^2 = \mathrm{var}(y_i)$ and ϕ is the autocorrelation parameter for a model with k joinpoints. In computing the test statistic, we may assume $\sigma^2 = 1$ without loss of generality

and ϕ can be estimated as

$$\hat{\phi} = \sum_{i=2}^{n} \hat{\varepsilon}_i^{(k)} \hat{\varepsilon}_{i-1}^{(k)} \Big/ \sum_{i=2}^{n} \{\hat{\varepsilon}_{i-1}^{(k)}\}^2.$$

A combined heteroscedastic correlated model can be also fit by using weighted least squares with $W = \hat{\Sigma}^{-1/2} \hat{V}^{-1} \hat{\Sigma}^{-1/2}$, and $\hat{V}$ is the diagonal matrix of estimated heteroscedastic variances.

2.4. Implementation of Algorithm

In this section we discuss details on how the test statistic is calculated. First, we discuss an approximation to the method for Sections 2.2 and 2.3 that greatly reduces computation time. Second, we discuss an adjustment for multiple testing. Finally, we discuss details of our program that is available upon request.

Ideally, for the permutation tests in Section 2.2 and 2.3, we would recalculate the weights for the weighted least squares estimates for each permutation. In practice we calculate the weights only for the original data, and use those same weights for each permuted data set. The simulations have shown that even with this approximation, the tests perform well.

We now briefly discuss how to perform subsequent testings and how to adjust the overall significance level. We begin with testing the null hypothesis of k_0 joinpoints against the alternative of k_1 joinpoints, where $0 \leqslant k_0 < k_1$. If the null hypothesis is rejected at level α_1 and $k_1 - k_0 \geqslant 2$, then we test H_0: there exist $k_0 + 1$ joinpoints against H_1: there exist k_1 joinpoints. If the null hypothesis of k_0 joinpoints is not rejected and $k_1 - k_0 \geqslant 2$, then we test H_0: there exist k_0 joinpoints against H_1: there exist $k_1 - 1$ joinpoints. We then proceed in a similar manner, where we increase the number of joinpoints under the null model by one if the null hypothesis is rejected and decrease the number of joinpoints under the alternative model by one otherwise, until we complete testing the null hypothesis of k joinpoints against the alternative of $k + 1$ joinpoints for some $k_0 \leqslant k < k_1$. The estimated number of joinpoint is $k + 1$ if the final null hypothesis is rejected, and k otherwise. Since to reach a conclusion we run $k_1 - k_0$ tests, we adjust the level of each test as $\alpha/(k_1 - k_0)$ using the Bonferroni correction for the overall significance level of α.

If we begin with testing the null hypothesis of no change against the alternative of N_{JP} joinpoints where N_{JP} is the predetermined maximum number of joinpoints and each test is at the α/N_{JP} level, it can be easily shown that the probability of concluding one or more joinpoints when there is no change is at most α. This overall significance level could be improved by considering the correlation between the consecutive tests. Also, for each $k = 1, \ldots, N_{\mathrm{JP}}$, the probability of concluding exactly k joinpoints when there is no change is at most $k\alpha/N_{\mathrm{JP}}$. Similarly, the probability of selecting more than K joinpoints when the true model has K joinpoints is at most $((N_{\mathrm{JP}} - K)/N_{\mathrm{JP}})\,\alpha$. Also the probability of selecting a model with a fewer number of joinpoints than the true model is small if the individual permutation test is reasonably powerful. Similar arguments can be used to compute probabilities of misclassification when we begin with the null hypothesis of $k_0 (>0)$ joinpoints against the alternative of k_1 joinpoints. The sequential scheme used in our program is efficient since it does not make all of $\binom{N_{\mathrm{JP}}}{2}$ comparisons, while the probability of misclassification is under control in most practical cases.

The program we have developed can analyse normal or Poisson data having independent or autocorrelated errors, and fit a linear or log-linear model having up to four joinpoints. Heteroscedastic data other than Poisson can be analysed by including weights, such as the reciprocal of the variance, in the data file. For the autocorrelated data, the program obtains the residuals fitting

the k joinpoint regression model assuming no autocorrelation and estimates the autocorrelation parameter based on these residuals. Then the k joinpoint regression model is refitted by using the weighted least squares and the grid search will be done to search for the minimum sum of squares of residuals. In addition to the p-value of the permutation test, the program produces the least squares estimates of the parameters and confidence intervals for the joinpoints using Lerman's approximation.[4] The program also calculates the annual percentage change (APC) along with its confidence interval. The APC is calculated when $x_1, \ldots, x_n$ represent years and $y_1, \ldots, y_n$ represent the log of the observed rate. Then the APC between τ_j and τ_{j+1} is $100(e^{\beta_1+\delta_1+\delta_2+\cdots+\delta_j}-1)$. This program can accept rate data from the National Cancer Institute's SEER*STAT program which produces age-specific and age-adjusted cancer incidence and mortality rates and their variances.

The program uses a grid search method to fit the model.[4] The grid search is simple to implement, provides approximate confidence intervals for the joinpoints, and can be computationally efficient when fitting a small number of joinpoints since it allows one to analyse all the permuted data sets simultaneously. The grid search method is not computationally efficient, however, when the number of joinpoints is large, since a k joinpoint model requires a k-dimensional grid search. Since the grid search method is equivalent to choosing the best k covariates amongst a finite number of contenders, an efficient best-linear-regression-model selection procedure such as the leaps and bounds method[1,2] might be utilized when the number of joinpoints is large.

3. SIMULATIONS

To assess and compare the performance of the procedure discussed in Section 2, we simulated samples of data sets for a variety of settings plausible in vital rates data. For all of these simulations, the values of the predictor variable was chosen as 69, 70, … , 95 denoting yearly time points. First we explore the performance of the permutation test when the errors are asymptotically exchangeable. As an example, we simulated independent log-normal observations for the response variable y under various combinations of $(\sigma, \beta_0, \beta_1, \delta_1, \ldots, \delta_{k_0}, \tau_1, \ldots, \tau_{k_0})$ and estimated the size and the power of the permutation test. Then we investigated the performance of the permutation test when the errors are not exchangeable, even asymptotically. We considered the Poisson observations as well as autocorrelated normal errors and examined the performance of the weighted least squares permutation test. For the alternative of one change, τ_1 was assumed as 80 or 90, and for the alternative of two joinpoints, (τ_1, τ_2) is assumed to be (80, 90). The estimated size and the power in the tables are the relative frequencies of the simulation runs where the p-value is less than 0·05 under the null model and under the alternative model, respectively.

Table I shows the estimated size and the power for independent log-normal observations. In Table I(a) for testing the null hypothesis of no change against the alternative of one change, the simulations were run 10,000 times with the grid size 0·1 for $\sigma^2 = 0{\cdot}0001$, 0·0002, 0·0005, 0·001, $\tau = 80$, 90, and for $(\mathrm{APC}_1, \mathrm{APC}_2) = (3, 3)$, (3, 2·4), (3, 2·0), (3, 1·5), where APC_i is the annual percentage change for the ith segment. Note that $(\mathrm{APC}_2\text{–}\mathrm{APC}_1) = 100(e^{\beta_1+\delta_1}-1) - 100(e^{\beta}_1-1) \approx 100\delta_1$. The estimated sizes in Table I(a) are all close to the nominal value of 0·05, indicating that the permutation distribution of the test statistic accurately approximates the true null distribution. Table I(a) also includes the power of the permutation test in detecting a change of a size δ_1 at $\tau_1 = 80$ or 90.

Table I. Size and power for independent log-normal observations

(a) 0 versus 1. Model: $\log y = \beta_0 + \beta_1 x + \delta_1 (x - \tau_1)^+ + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$ and $x = 69, 70, \ldots, 95$

σ^2	(apc_1, apc_2)	$\lvert\delta_1\rvert/\sigma$	$\tau_1 = 80$		$\tau_1 = 90$	
			Power	Mean est. jp $\pm$ SE	Power	Mean est. jp $\pm$ SE
0·0001	(3, 3)	0	0·0525		0·0525	
	(3, 2·4)	0·60	0·9980	80 $\pm$ 0·006	0·7183	89·97 $\pm$ 0·008
	(3, 2·0)	1	1·0000	80 $\pm$ 0·010	0·9914	89·82 $\pm$ 0·017
	(3, 1·5)	1·5	1·0000	80·02 $\pm$ 0·019	1·0000	88·85 $\pm$ 0·040
0·0002	(3, 3)	0	0·0494		0·0494	
	(3, 2·4)	0·42	0·9206	80 $\pm$ 0·009	0·4188	89·97 $\pm$ 0·015
	(3, 2·0)	0·71	0·9998	80·03 $\pm$ 0·015	0·8506	89·30 $\pm$ 0·032
	(3, 1·5)	1·06	1·0000	80·10 $\pm$ 0·031	0·9970	87·26 $\pm$ 0·060
0·0005	(3, 3)	0	0·0506		0·0506	
	(3, 2·4)	0·27	0·5448	80·03 $\pm$ 0·017	0·1901	89·27 $\pm$ 0·033
	(3, 2·0)	0·45	0·9454	80·08 $\pm$ 0·029	0·4514	87·57 $\pm$ 0·057
	(3, 1·5)	0·67	0·9701	80·58 $\pm$ 0·051	0·8163	84·94 $\pm$ 0·076
0·0010	(3, 3)	0	0·0488		0·0488	
	(3, 2·4)	0·19	0·3047	80·07 $\pm$ 0·027	0·1161	87·98 $\pm$ 0·053
	(3, 2·0)	0·32	0·6991	80·29 $\pm$ 0·044	0·2488	85·70 $\pm$ 0·072
	(3, 1·5)	0·47	0·9701	80·99 $\pm$ 0·063	0·4995	83·68 $\pm$ 0·080

(b) 1 versus 2. Model: $\log y = \beta_0 + \beta_1 x + \delta_1 (x - \tau_1)^+ + \delta_2 (x - \tau_2)^+ + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$ and $x = 69, 70, \ldots, 95$

σ^2	(apc_1, apc_2, apc_3)	$\left(\frac{\lvert\delta_1\rvert}{\sigma}, \frac{\lvert\delta_2\rvert}{\sigma}\right)$	$(\tau_1, \tau_2) = (80, 90)$	
			Power	Mean est. jp $\pm$ SE
0·0001	(3, 3, 2)		0·0468	
	(2·4, 3, 2)	(0·6, 1)	0·8976	80·18 $\pm$ 0·024, 89·56 $\pm$ 0·020
0·0002	(3, 3, 2)		0·0467	
	(2·4, 3, 2)	(0·42, 0·71)	0·5984	80·09 $\pm$ 0·035, 88·91 $\pm$ 0·032
0·0005	(3, 3, 2)		0·0522	
	(2·4, 3, 2)	(0·27, 0·45)	0·2438	79·77 $\pm$ 0·048, 87·50 $\pm$ 0·048
0·0010	(3, 3, 2)		0·0476	
	(2·4, 3, 2)	(0·19, 0·32)	0·1331	79·43 $\pm$ 0·054, 86·67 $\pm$ 0·055

We see in the table that the permutation test is reasonably powerful in detecting an effect size greater than 0·4 (that is, $\lvert\delta_1\rvert/\sigma > 0{\cdot}4$) when the change occurs near the middle. If the change occurs near the end of the data, then a reasonable power is expected when $\lvert\delta_1\rvert/\sigma$ is about 0·7 or larger. A similar numerical study was done for testing the null hypothesis of one joinpoint against the alternative of two joinpoints and is summarized in Table I(*b*). For this case, simulations were run 10,000 times with the grid size of 1. A larger grid size was chosen to save the computation

Table II. Size and power for log of Poisson rates

(a) 0 versus 1, $y = \log(z/n)$ where $z \sim$ Poisson $(n\lambda)$ with $\log(\lambda) = \beta_0 + \beta_1 x + \delta_1 (x - 90)^+$

$\sigma^2 = (n\lambda)^{-1}$ at $x = 82$	$n\lambda$ at $x = 82$	$(\text{apc}_1, \text{apc}_2) = (3, 3)$		$(\text{apc}_1, \text{apc}_2) = (3, 2)$		
		Unadjusted size*	Adjusted size†	$\lvert\delta_1\rvert/\sigma$	Unadjusted power*	Adjusted power†
0·0001	10,000	0·0614	0·0514	1·00	0·9934	0·9986
0·0002	5,000	0·0639	0·0513	0·71	0·8486	0·9184
0·0005	2,000	0·0583	0·0498	0·45	0·4320	0·5473
0·0010	1,000	0·0587	0·0518	0·32	0·2301	0·3031

(b) 1 versus 2, $y = \log(z/n)$ where $z \sim$ Poisson $(n\lambda)$ with $\log(\lambda) = \beta_0 + \beta_1 x + \delta_1 (x - 80)^+ + \delta_2 (x - 90)^+$

$\sigma^2 = (n\lambda)^{-1}$ at $x = 82$	$n\lambda$ at $x = 82$	$(\text{apc}_1, \text{apc}_2, \text{apc}_3) = (3, 3, 2)$		$(\text{apc}_1, \text{apc}_2, \text{apc}_3) = (2{\cdot}4, 3, 2)$		
		Unadjusted size*	Adjusted size†	$(\lvert\delta_1\rvert/\sigma, \lvert\delta_2\rvert/\sigma)$	Unadjusted power*	Adjusted power†
0·0001	10,000	0·0643	0·0481	(0·60, 1·00)	0·8943	0·9340
0·0002	5,000	0·0613	0·0481	(0·42, 0·71)	0·5664	0·6508
0·0005	2,000	0·0601	0·0464	(0·27, 0·45)	0·2372	0·2844
0·0010	1,000	0·0547	0·0446	(0·19, 0·32)	0·1215	0·1441

* Relative proportion of the simulation runs whose p-value is less than 0·05

† Relative proportion of the simulation runs whose p-value is less than 0·05 adjusted for heteroscedasticity

time. From additional simulation studies, we have found that the choice of the grid size does not make any significant change in the p-value and power estimation, but a smaller grid size is preferred to find a shorter confidence interval for the joinpoint. Table I also shows the mean estimated joinpoints along with their standard errors. The mean estimated joinpoint is the average of estimated joinpoint through all the simulation runs and the standard error is the standard deviation of these estimates. It is observed that both the standard error and the bias increase as σ increases and also as δ_1 decreases. For the given choices of σ and δ_1, we find that the bias and the standard error increase as the joinpoint approaches the ends.

Table II shows the estimated size and the power for the log of the Poisson rates. Simulations were run 10,000 times and the grid size was 0·1 for 0 versus 1 testing, and 1 for 1 versus 2 testing, respectively. As in the analysis of cancer mortality data, let the response Y_i be the natural logarithm of cancer death rate at time i, that is, $Y_i = \log(Z_i/n_i)$, where n_i is the population size at time i and Z_i is the cancer deaths from a Poisson distribution with mean rate of λ_i or mean count of $n_i\lambda_i$, equivalently. Then, as discussed in Section 2.2, the variance of Y_i at time i is approximately $1/(n_i\lambda_i)$ and is estimated as $\hat{V}_i = 1/z_i$, where z_i is the observed cancer deaths at time i. In the simulation study, $\log(\lambda_i) = \beta_0 + \beta_1 x_i + \delta_1 (x_i - \tau_1)^+ + \delta_2 (x_i - \tau_2)^+$ and various values are chosen for the Poisson parameter λ. The mean counts at $x_i = 82$, $n_i\lambda_i$, are chosen as 10,000, 5000, 2000, 1000 yielding the estimated variance of Y_i given $x_i = 82$ as 0·0001, 0·0002, 0·0005, 0·0010, respectively. Table III shows the 1994 count of cancer deaths in the U.S. and number of cases in the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) population

Table III. Cancer cases (SEER*) and deaths (U.S.†) for selected sites (1994)

	Cases	Deaths
All sites	108,265	534,294
Colon and rectum	15,826	57,407
Pancreas	2,489	26,834
Lung and bronchus	14,981	149,354
Melanomas of the skin	3,491	6,163
Breast	16,199	44,008
Corpus and uterus, NOS	3,118	6,163
Cervix	1,169	4,602
Ovary	2,080	13,500
Prostate	16,818	34,901
Kidney and renal pelvis	2,487	10,749
Brain and other nervous system	1,520	12,313
Lymphomas	5,164	23,248
Leukaemias	2,578	19,833

* Invasive cancers from National Cancer Institute's Surveillance, Epidemiology, and End Results Population-Based registry Program representing about 10 per cent of the U.S. population

† Source National Center for Health Statistics

based registry program which covers approximately 10 per cent of the U.S. population. Although we do not usually analyse trends in crude rates, the variance of the log of the age-adjusted and crude rates are approximately equal, especially when the standard does not differ much from the observed population. Thus, these simulations for crude rates are applicable to age-adjusted rates as well.

Table II shows that the permutation test uncorrected for heteroscedasticity usually underestimates the true *p*-value, therefore rejects more than 5 per cent of the time. The bias decreases as the heteroscedasticity gets less severe. Comparing Tables I and II for $|\delta_1|/\sigma = 0{\cdot}45$ and $\tau_1 = 90$, we observe the power of 0·4514 in Table I and the power of 0·4320 in Table II. These cases are not directly comparable because in the homoscedastic case the variance is constant across all of the years, while in the Poisson case the variance of the log of the rates increases as the rates increase. To see if the weighted least squares method would improve the *p*-value estimation, weighted least squares estimates are obtained in computing the test statistic. First, we consider the weighted least squares permutation test to correct the heteroscedasticity caused by the Poisson observations and summarized the results in Tables II(*a*) and (*b*). When the weighted least squares estimates were used with the weight of $1/\hat{V}_i$ for the *i*th observation, the adjusted sizes in the third columns of Tables II(*a*) and (*b*) show a very good accuracy in the *p*-value estimation. Further, the adjusted tests show an improvement in power over the unadjusted tests.

In Table IV, we explored the performance of the permutation test when normal errors are autocorrelated with autocorrelation coefficients of $-0{\cdot}6$, $-0{\cdot}2$, 0, 0·2 and 0·6. For negative autocorrelation, the unadjusted sizes are less than the 0·05 significance level (severely less for large negative autocorrelation), while for positive autocorrelation the unadjusted size can be much greater than 0·05. Alternatively, the sizes of the tests adjusted for autocorrelation appear close to 0·05 except for when there is large positive autocorrelation. With respect to power, the unadjusted test has higher power than the adjusted test, except for possibly at very high negative correlation.

Table IV. Size and power for autocorrelated errors

(a) 0 versus 1, $y \sim$ normal with $E[y|x] = \beta_0 + \beta_1 x + \delta_1(x - 80)^+$ and $\text{cov}(y_i, y_{i+k}) = 0{\cdot}0001\,\phi^k$

ϕ	$(\text{apc}_1, \text{apc}_2) = (3, 3)$		$(\text{apc}_1, \text{apc}_2) = (3, 2)$		
	Unadjusted size*	Adjusted size†	$\lvert\delta_1\rvert/\sigma$	Unadjusted power*	Adjusted power†
−0·6	0·0011	0·0445	1·00	0·9934	0·9999
−0·2	0·0173	0·0435	1·00	0·9945	0·9741
0	0·0525	0·0492	1·00	0·9914	0·9032
0·2	0·1204	0·0549	1·00	0·9868	0·7687
0·6	0·4033	0·0760	1·00	0·9795	0·5730

(b) 1 versus 2, $y \sim$ normal with $E[y|x] = \beta_0 + \beta_1 x + \delta_1(x - 80)^+ + \delta_2(x - 90)^+$ and $\text{cov}(y_i, y_{i+k}) = 0{\cdot}0001\,\phi^k$

ϕ	$(\text{apc}_1, \text{apc}_2, \text{apc}_3) = (3, 3, 2)$		$(\text{apc}_1, \text{apc}_2, \text{apc}_3) = (2{\cdot}4, 3, 2)$		
	Unadjusted size*	Adjusted size†	$(\lvert\delta_1\rvert/\sigma, \lvert\delta_2\rvert/\sigma)$	Unadjusted power*	Adjusted power†
−0·6	0·0012	0·0429	(0·60, 1·00)	0·8949	0·9973
−0·2	0·0146	0·0418	(0·60, 1·00)	0·8956	0·8473
0	0·0468	0·0523	(0·60, 1·00)	0·8976	0·6799
0·2	0·1210	0·0583	(0·60, 1·00)	0·9085	0·5395
0·6	0·3966	0·1054	(0·60, 1·00)	0·9451	0·4330

* Relative proportion of the simulation runs whose p-value is less than 0·05

† Relative proportion of the simulation runs whose p-value is less than 0·05 adjusted for autocorrelation

Thus, we recommend the unadjusted test if one can assume a non-positive autocorrelation that is unlikely to be very largely negative, since for all non-positive autocorrelation the unadjusted test appears to achieve the nominal significance level and for all except very large negative autocorrelation the unadjusted test appears to have greater power. We recommend the adjusted test over the unadjusted test when one suspects significant positive autocorrelation or when one wants a conservative test with minimal assumptions, since the test with the adjustment achieves the nominal significance level in all cases except very large positive autocorrelation, in which case it performs better than the unadjusted test.

Although not reported in this paper, we have obtained the simulated coverage probabilities based on the confidence interval suggested in Lerman.[4] Simulated coverage probability represents the relative frequency of the simulation runs whose 95 per cent confidence interval includes the true joinpoints. The simulated coverage probabilities usually overestimate the nominal 95 per cent, and is improved by choosing a finer grid. With the grid of size 0·01 year, we have observed the simulated coverage probabilities usually close to 95 per cent.

4. EXAMPLE

As an example, we examine prostate cancer incidence from 1973 to 1995 in approximately 10 per cent of the U.S. population covered by the National Cancer Institute's Surveillance,

Epidemiology, and End Results (SEER) registry program, and U.S. mortality from 1969 to 1995. Prostate cancer incidence has recently seen some of the most dramatic swings ever witnessed for any cancer site. These swings are attributable to the introduction of the prostate specific antigen (PSA) test as a screen for prostate cancer. The introduction of a screening test in a population usually causes a rapid increase in incidence as cases are shifted from the future to the present, followed by a rapid decline. The exact timing of the rise and fall of incidence is a function of both screening rates and lead time, the amount of time diagnosis is advanced due to screening. If screening detects cases that would never have been clinically diagnosed prior to death from other causes (that is, overdiagnosis) then incidence will never return to the background trend before the change. Joinpoint models of prostate cancer incidence can help us determine when incidence started rising, when it peaked, and the degree to which incidence has returned to the background trend before the change. As part of a more extensive modelling exercise, these trends can help us understand unobservable operating characteristics of PSA screening such as lead time and the extent of overdiagnosis. A joinpoint model of mortality can allow us to detect the possible benefits of PSA screening.

The response variables for the analysis of incidence and mortality are the natural logarithms of age-adjusted prostate cancer rates. We fit the heteroscedastic/uncorrelated errors model described in Section 2.2 and the heteroscedastic/autocorrelated errors model described in Section 2.3. The grid search uses a grid size of 0·1 year and the permutation tests are based on 1000 Monte Carlo replicates. We perform three permutation tests with the Bonferroni correction and an overall significance level of 0·05.

Fitting the uncorrelated errors model to prostate cancer incidence data we obtained p-values of 0·001 testing the null hypothesis of 0 joinpoints against the alternative of 3 joinpoints, 0·001 testing 1 joinpoint against 3 joinpoints and 0·009 testing 2 joinpoints against 3 joinpoints. Comparing these to the critical value of 0·05/3, we reject all three null hypotheses and therefore select the three-joinpoint model as our final model. Fitting the autocorrelated errors model we obtained p-values of 0·001 testing 0 joinpoints against 3 joinpoints, 0·002 testing 1 joinpoint against 3 joinpoints and 0·038 testing 2 joinpoints against 3 joinpoints. We reject the first two null hypotheses but not the third and therefore select the two-joinpoint model as our final model. Figure 1(*a*) shows the observed data and the final three-joinpoint model under the uncorrelated model with joinpoints at 1985·4 (95 per cent CI (1980·7, 1986·1)), 1989 (95 per cent CI (1986, 1989·1)) and 1992 (95 per cent CI (1991·9, 1992·3)). Figure 1(*b*) shows the final two-joinpoint model under the autocorrelated model with joinpoints at 1988·4 (95 per cent CI (1987·1, 1989)) and 1992 (95 per cent CI (1991·6, 1992·4)). Comparing the two models we see they are quite similar, except that the two-joinpoint model ignores a slight change in trend in the 1985–1989 interval. From analyses of Medicare claims data,[16] we know that PSA testing rates started rising in 1988 and 1989. While the per cent of men having a PSA test in the past year continued to rise through 1995, the per cent of men having their first test peaked in 1992, and first tests (that is, prevalence screens) have the largest cancer yield.[17] The pre-PSA rise in incidence in the uncorrelated errors model may be associated with other technological changes occurring in this era, such as the change from the core to the fine-needle biopsy and the spring driven ultrasound guided biopsy gun.[18] These easier, less painful, and more accurate biopsy procedures may have lowered the threshold for proceeding with a biopsy in an uncertain situation, and increased the chance of finding cancer if it was present. Incidence is close to returning to the pre-PSA secular trend, although one must be cautious in drawing conclusions about overdiagnosis from this observation. The transurethral resection of the prostate (TURP) is a procedure used to relieve

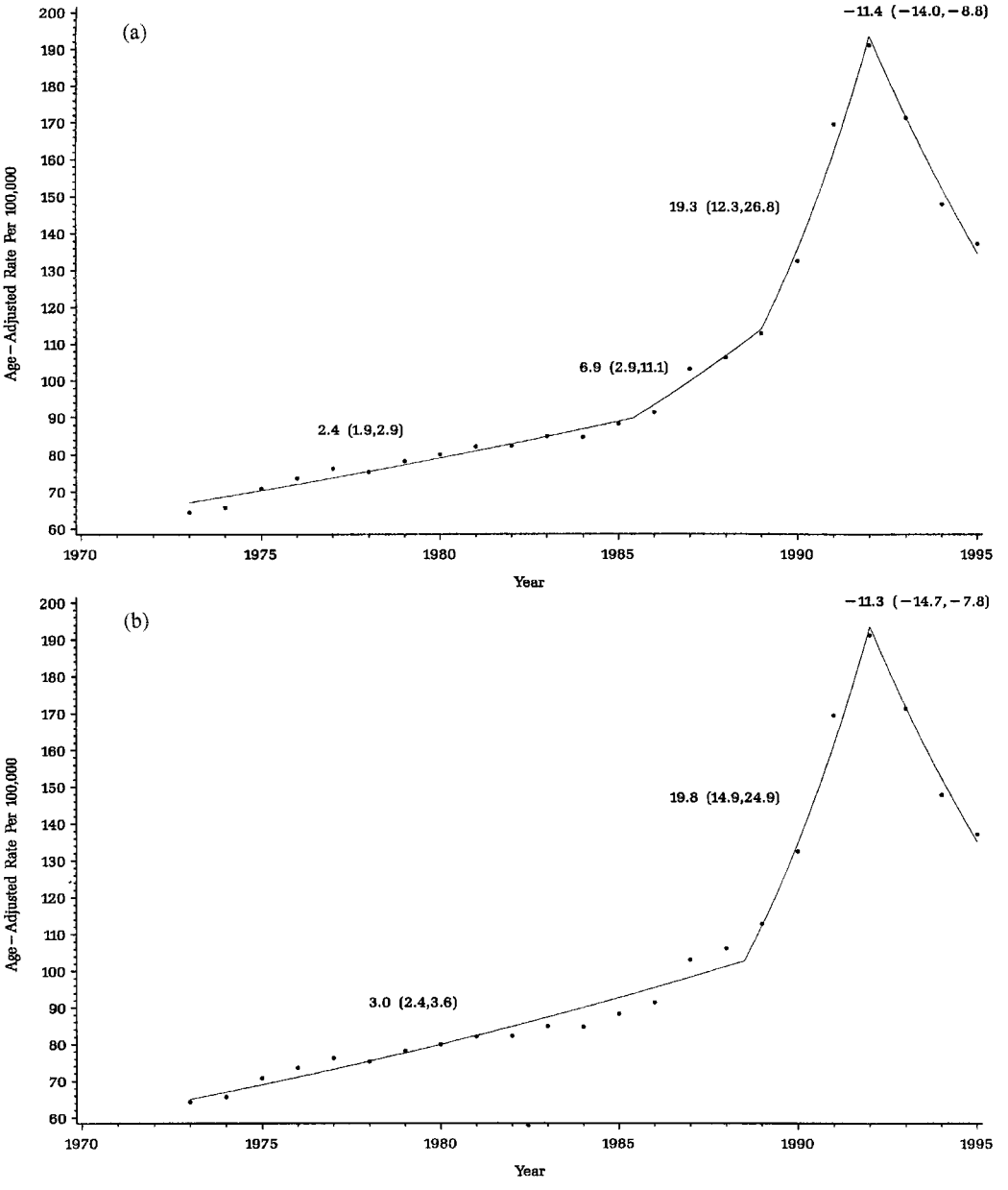

Figure 1. Prostate cancer incidence: (*a*) final uncorrelated errors model (3 joinpoints) with EAPC (95 per cent confidence interval) printed for each line segment; (*b*) final autocorrelated errors model (2 joinpoints) with EAPC (95 per cent confidence interval) printed for each line segment

symptoms of benign prostatic hypertrophy (BPH), although it is not uncommon for an incidental finding of prostate cancer to occur during this procedure.[19] A recent change from the surgical to the medical management of BPH may have lowered the TURP rate, and therefore the background trend in the absence of PSA testing may be lower than projected.[20]

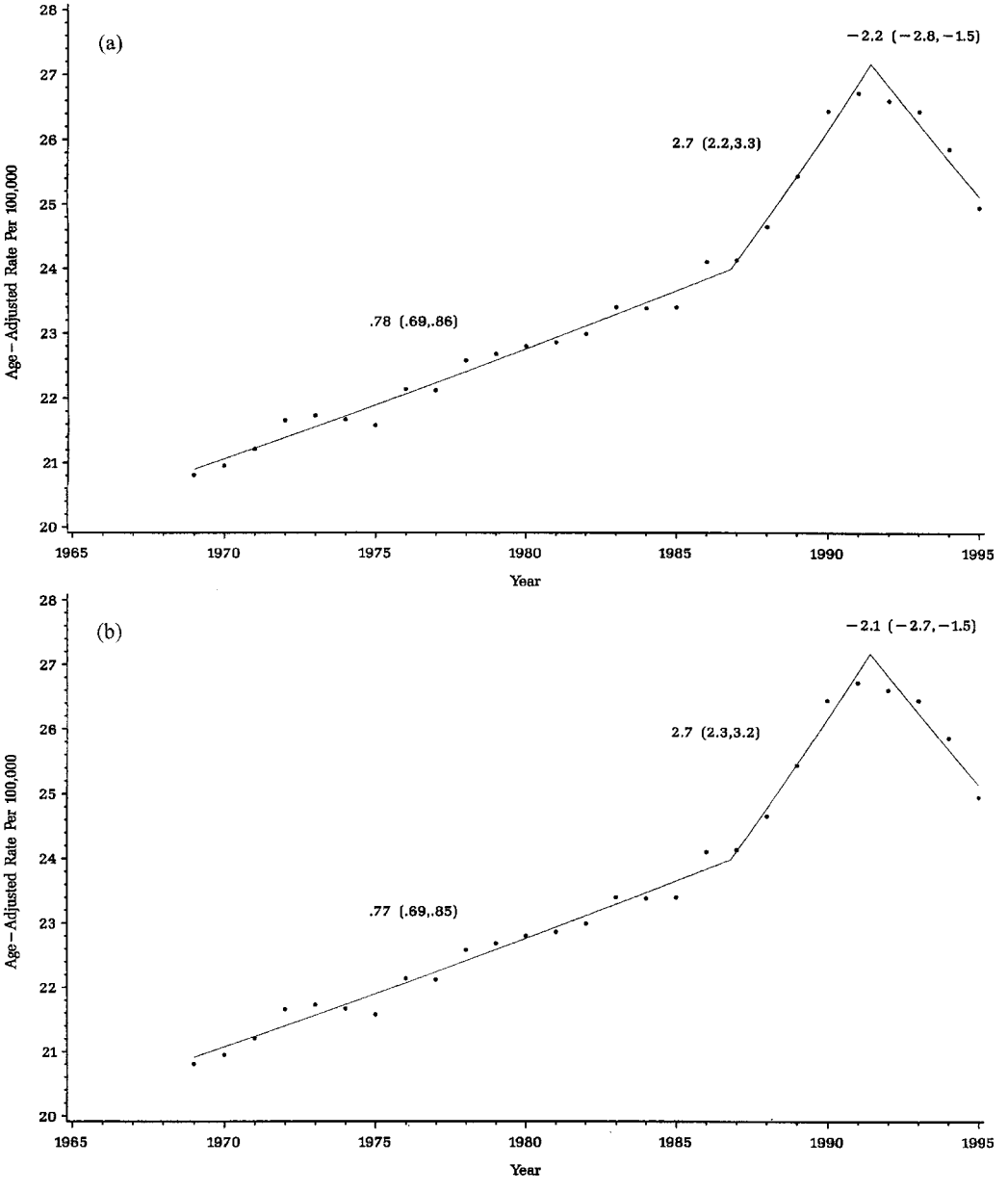

Figure 2. Prostate cancer mortality: (*a*) final uncorrelated errors model (2 joinpoints) with EAPC (95 per cent confidence interval) printed for each line segment; (*b*) final autocorrelated errors model (2 joinpoints) with EAPC (95 per cent confidence interval) printed for each line segment

Fitting the uncorrelated errors model to the prostate cancer mortality data we obtained *p*-values of 0·001 testing 0 joinpoints against 3 joinpoints, 0·001 testing 1 joinpoint against 3 joinpoints and 0·146 testing 2 joinpoints against 3 joinpoints. We reject the first two null hypotheses but not the third and therefore select the two-joinpoint model as our final model.

Fitting the autocorrelated errors model we obtained p-values of 0·001 testing 0 joinpoints against 3 joinpoints, 0·040 testing 1 joinpoint against 3 joinpoints and 0·011 testing 1 joinpoint against 2 joinpoints. We reject the first and third null hypotheses but not the second and therefore select the two-joinpoint model as our final model. Notice that although we selected the same final model assuming uncorrelated and autocorrelated errors, the test results and even the tests performed differed under the two assumptions. Figures 2(*a*) and (*b*) show the observed data and the final two-joinpoint model under the uncorrelated and autocorrelated models, respectively. For the uncorrelated model the joinpoints were estimated as 1986·8 (95 per cent CI 1985·1, 1988) and 1991·4 (95 per cent CI 1990·5, 1991·9) with identical point estimates and only slightly different confidence intervals for the autocorrelated model. The rise and fall in mortality tracks with the PSA induced rise and fall in incidence. The decline in mortality, which may be the first signs of the beneficial effects of PSA screening, has only just returned to the pre-PSA background trend. One might suspect a coding artifact associated with misattribution of cause of death for prostate cancer patients who die of other causes. This misattribution may be consistent with recent mortality trends, especially if one posits that a fixed proportion of the rising and falling pool of newly diagnosed cases mistakenly have their cause of death attributed to prostate cancer merely because they were diagnosed with the disease.

The estimated autocorrelation parameters are 0·18 (95 per cent CI ($-0{\cdot}20$, 0·56)) for the selected incidence model and $-0{\cdot}12$ (95 per cent CI ($-0{\cdot}50$, 0·25)) for the selected mortality model. Deciding between the uncorrelated and autocorrelated errors model is a trade-off between a potential violation of basic assumptions (that is, choosing the uncorrelated model when there is in fact autocorrelation) and possible unnecessary loss of power (that is, using the autocorrelation model when it is really not necessary). The large confidence intervals around the autocorrelation coefficients do not provide much guidance with respect to this choice. For incidence the uncorrelated model has detected a subtle change in the rates prior to the large PSA induced increase, while the autocorrelated model is more conservative. For mortality, the results of the two models agree.

5. DISCUSSION

We proposed a procedure to identify changes in trend data. For testing the null hypothesis of no joinpoint against the alternative of one joinpoint, a classical asymptotic theory, which does not hold in our case, suggests the F-distribution with 2 and $n-4$ degrees of freedom for the null distribution of the test statistic and Hinkley[1] suggested the F-distribution with 3 and $n-4$ degrees of freedom based on empirical evidence. In our numerical study, we have observed strong empirical evidence that the true distribution of the likelihood ratio statistic is between these two F-distributions and that Hinkley's approximation usually yields a conservative test with a p-value larger than the true p-value. Knowles and Siegmund[2] provided an analytic approximations for the p-values and Zhang[21] has modified the procedure in the context of sequential detection. Since the original approximation of Knowles and Siegmund is only for testing no joinpoint against the alternative of one joinpoint and the procedure of Zhang is a sequential version, we cannot make a direct comparison. However, we found in analysing prostate cancer incidence and mortality data that the Knowles and Siegmund procedure and our permutation procedure produced quite comparable joinpoint estimates for the incidence rates, but somewhat different results for the mortality rates.

In the process of comparing our permutation test with the sequential test of Zhang,[21] a fundamental question of modelling arose. In this paper we chose to analyse trends in prostate

cancer mortality over 27 years with special interest in recent trends. If we apply a sequential test starting in 1969, then joinpoints are identified as we include additional years, but once identified they become fixed as we search for additional ones. The test proposed in this paper identifies the best fitting set of joinpoints over the entire range of the data. The two approaches may yield different solutions as more recent joinpoints can have influence on the identification of earlier ones. Since this 27 year sequence of data has already been observed it seems better to analyse the overall best fit instead of treating the data as if it had accumulated in the process of this analysis. In future years, however, one must be concerned with maintaining an overall significance level in the face of repeated analyses, adding one additional point each times.

One of difficulties that hindered the application of likelihood ratio test in joinpoint regression was the intractability of its analytic distribution. A few researchers have studied resampling techniques in the context of change-point detection, but they have been in limited situations and most of them used the bootstrap approach. Romano[22] discussed the asymptotic efficiency of the bootstrap and the permutation tests and concluded the asymptotic equivalence of these tests. We expect that similar results would hold using the bootstrap method.

Bayesian methodologies have also been applied to change-point modelling. When the number of change-points is fixed, Gibbs sampling has been applied to a variety of models.[23,24] However Gibbs sampling may not be appropriate when length of the parameter set is not fixed, as is the case in this paper where the number of change-points is unknown. Green[25] suggests using Hastings algorithms which incorporate jumps between parameter subspaces of different dimensionality for problems where the number of change-points is a random variable. The application of the procedure suggested by Green to the problem of modelling cancer trends is certainly an area for future research. Bayesian approaches incorporate a prior distribution on the number of change-points based on knowledge of the related events and provide an estimate of a posterior distribution for the number of change-points that eliminates the need for multiple tests of the number of change-points.

Non-parametric smoothing can also provide similar results in estimating joinpoints and other regression parameters. As discussed in Section 9.5 of Seber[26] and Eubank,[27] the model we are considering in this paper can be considered as a simple case of spline regression with variable knots. To estimate the number of free knots, Eubank[27] suggested several statistics such as a C_p type statistic, a statistic based on generalized cross-validation, and the one based on Akaike's information criterion. These methods , however, do not provide probabilities of misclassification and the permutation procedure that produces the estimated p-values shows advantages in interpreting the results. If one wants to consider a multi-phase non-linear model, non-parametric smoothing can certainly be an attractive approach.

While other approaches, such as a polynomial fit to the data, could be considered, we have found joinpoint regression a useful way to summarize trends in cancer rates. While we do not really believe that cancer rates change abruptly, connecting linear line segments on a log scale allows us to characterize the trends succinctly (that is, in terms of an annual rate of change for fixed periods). These methods also allow us to test for recent changes in trend.

REFERENCES

1. Hinkley, D. V. 'Inference in two-phase regression', *Journal of the American Statistical Association*, **66**, 736–743 (1971).
2. Knowles, M. and Siegmund, D. 'On Hotelling's approach to testing for a nonlinear parameter in regression', *International Statistical Review*, **57**, 205–220 (1989).

3. Sprent, P. 'Some hypotheses concerning two-phase regression lines', *Biometrics*, **17**, 634–645 (1961).
4. Lerman, P. M. 'Fitting segmented regression models by grid search', *Applied Statistics*, **29**, 77–84 (1980).
5. Smith, A. F. M. and Cook, D. G. 'Straight lines with a change-point: A Baysian analysis of some renal transplant data', *Applied Statistics*, **29**, 180–189 (1980).
6. Kim, H. M. and Lagakos, S. 'Assessing drug compliance using longitudinal marker data, with application to AIDS', *Statistics in Medicine*, **13**, 2141–2153 (1994).
7. Hotelling, H. 'Tubes and spheres in n-spaces and a class of statistical problems', *American Journal of Mathematics*, **61**, 440–460 (1939).
8. Quandt, R. E. 'The estimation of the parameter of a linear regression system obeying two separate regimes', *Journal of the American Statistical Association*, **53**, 873–880 (1958).
9. Worsley, K. J. 'Testing for a two-phase mutliple regression', *Technometrics*, **25**, 34–42 (1983).
10. Kim, H. J. and Siegmund, D. 'The likelihood ratio test for a change-point in simple linear regression', *Biometrika*, **76**, 409–423 (1989).
11. Hudson, D. J. 'Fitting segmented curves whose join points have to be estimated', *Journal of the American Statistical Association*, **61**, 1071–1129 (1966).
12. Furnival, G. M. and Wilson, R. W. 'Regression by leaps and bounds', *Technometrics*, **16**, 499–511 (1974).
13. Edgington, E. S. *Randomization Tests*, 2nd edn, Marcel Dekker, New York, 1987.
14. Good, P. *Permutation Tests*: *A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer-Verlag, New York, 1994.
15. Feder, P. I. 'On asymptotic distribution theory in segmented regression problems-identified case', *Annals of Statistics*, **3**, 49–83 (1975).
16. Potosky, A. L., Miller, B. A., Albertson, P. C. and Kramer, B. S. 'The role of increasing detection in the rising incidence of prostate cancer', *Journal of the American Medical Association*, **273**, 548–552 (1995).
17. Legler, J. M., Feuer, E. J., Potosky, A. L., Merrill, R. M. and Kramer, B. S. 'The role of prostate specific antigen (PSA) testing patterns in the recent prostate cancer incidence decline in the United States', *Cancer Causes and Control*, (1999).
18. Drago, J. R. 'The role of new modalities in the early detection and diagnosis of prostate cancer', *CA-A Cancer Journal for Clinicians*, **39**, 326–336 (1989).
19. Newman, A. J., Graham, M. A., Carleton, C. E. Jr. and Lieman, S. 'Incidental carcinoma of the prostate at the time of transuretuhal resection: importance of evaluating every chip', *Journal of Urology*, **128**, 948–950 (1982).
20. 'Prostate enlargement: benign prostatic hyperplasia', NIH Publication No. 91-3012, Bethesda MD, 1991.
21. Zhang, H. 'Detecting change points and monitoring biomedical data', *Communications in Statistics – Theory and Method*, **24**, 1307–1324 (1995).
22. Romano, J. P. 'Bootstrap and randomization tests of some nonparametric hypotheses', *Annals of Statistics*, **17**, 141–159 (1989).
23. Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. 'Hierarchical Bayesian analysis of change-point problems', *Applied Statistics*, **41**, 389–405 (1992).
24. Stephens, D. A. 'Bayesian retrospective multiple-changepoint identification', *Applied Statistics*, **43**, 159–178 (1994).
25. Green, P. J. 'Reversible jump Markov change Monte Carlo computation and Bayesian model determination', *Biometrika*, **82**, 711–732 (1995).
26. Seber, G. A. F. and Wild, C. J. *Nonlinear Regression*, Wiley, 1989.
27. Eubank, R. L. 'Approximate regression models and splines', *Communications in Statistics – Theory and Method*, **13**, 433–484 (1984).

