The Canadian Journal of Statistics
Vol. 24, No. 1, 1996, Pages 37–53
La Revue Canadienne de Statistique

Estimation in the multipath change point problem for correlated data*

Lawrence JOSEPH, Alain C. VANDAL and David B. WOLFSON

Montreal General Hospital, McGill University and *McGill University*

Key words and phrases: Autoregressive process, change point; EM algorithm, mixture distribution, panel data, sedimentology.
AMS 1991 subject classifications: 62P99, 62M10, 62F99.

ABSTRACT

In many experiments, several measurements on the same variable are taken over time, a geographic region, or some other index set. It is often of interest to know if there has been a change over the index set in the parameters of the distribution of the variable. Frequently, the data consist of a sequence of correlated random variables, and there may also be several experimental units under observation, each providing a sequence of data. A problem in ascertaining the boundaries between the layers in geological sedimentary beds is used to introduce the model and then to illustrate the proposed methodology. It is assumed that, conditional on the change point, the data from each sequence arise from an autoregressive process that undergoes a change in one or more of its parameters. Unconditionally, the model then becomes a mixture of nonstationary autoregressive processes. Maximum-likelihood methods are used, and results of simulations to evaluate the performance of these estimators under practical conditions are given.

RÉSUMÉ

Nombreuses sont les expériences dans lesquelles on est appelé à mesurer la même variable à plusieurs reprises dans le temps, dans l'espace ou selon un autre schème de référence. Dans de telles circonstances, il est naturel de chercher à savoir si les paramètres de la loi de la variable sont partout les mêmes. Bien souvent, les données sont corrélées, et lorsqu'il y a plusieurs unités expérimentales sous observation, on dispose alors de suites de données corrélées. Ce problème se présente en géologie, par exemple, lorsqu'on souhaite identifier les strates d'une couche sédimentaire. Cette illustration sert de point de départ à la présentation d'un modèle et d'une stratégie d'analyse de ce type de données. On suppose que conditionnellement à un point de rupture, les données de chaque suite sont issues d'un processus autorégressif dont l'un ou plusieurs des paramètres subissent un changement. Marginalement, le modèle s'exprime alors comme un mélange de processus autorégressifs non stationnaires. On montre comment il est possible d'appliquer le principe du maximum de vraisemblance dans un pareil contexte, et on évalue le mérite de cette approche sous différentes conditions expérimentales, au moyen de simulations.

1. INTRODUCTION

The analysis of panel data has received considerable attention, particularly by econometricians (Gong and Sickles 1992, Baltagi, Chang and Li 1992) and sociologists (Palmquist and Green 1992). Kim and Basawa (1992) discuss an empirical Bayes approach to panel

*This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, and the Fonds pour la formation de chercheurs et l'aide à la Recherche Gouvernement du Québec. Lawrence Joseph is a research scholar, supported by the Fonds de la Recherche en Santé du Québec.

data arising from AR(1) processes. Although work in this area has concentrated on both sequences of independent and dependent observations, none has sought to investigate *independent sequences* of *dependent* observations which undergo a change at possibly different points from sequence to sequence, by including these change points as model parameters. Tang and MacNeill (1993) give a thorough discussion of the effect of serial correlation on changes in the parameters of a regression model. They are concerned with testing for a change rather than with estimating the change point. Also, unlike the model presented below, each sequence is assumed to have the same change point. Allowing for different change points from sequence to sequence leads to a model with mixtures of autoregressive processes in each row, complicating the problem and preventing the use of standard methodology for vector-valued autoregressive processes. The results of Tang and MacNeill do, however, show that not taking into account the underlying correlational structure (if it exists) of the problem can compromise the power of tests or lead to false Type 1 error probabilities. It is therefore plausible that similar observations would apply to mixtures of autoregressive processes were their correlational structure ignored. More recently, Müller and Rosner (1994) have modelled the decline and recovery of white-blood-cell count (WBC) in a group of patients undergoing chemotherapy, using a Bayesian hierarchical approach; the within-patient WBC readings, are, however, assumed to be independent.

In Section 2 a model is introduced in which it is assumed that the data consist of independent sequences of correlated observations with each sequence undergoing a change at a possibly different location. Specifically, this paper presents maximum-likelihood methodology for segments of autoregressive processes, each of which is subject to a change in one or more of its parameters. The common distribution of the times of change is considered to be of importance in the analysis, as the soil-profile example below illustrates. The methodology presented, however, is applicable in a variety of other settings, for example when there are anticipated changes in blood pressure readings after treatment or in weight gain after diet modification. This extends the previous work of Joseph and Wolfson (1993) to include sequences of correlated observations.

Soil profiles change with increasing depth. Frequently these changes occur quite abruptly because of the way sedimentary beds were laid down. Within a sedimentary bed there may also be trends in specific soil characteristics. These trends may themselves undergo sudden changes (Preston and Davis 1972). Soil profiles are not only important in providing a geological history of a region but also directly affect the rate at which water is absorbed and retained, and are a crucial component of petrology and mineral exploration. Attempts to ascertain estimates of the distribution of the depth at which a change occurs have depended on stochastic-process models, difficult to verify, of the actual sedimentation process that led to the sedimentary beds (Schwarzacher 1972a, b). It is possible, however, to avoid modelling this sedimentation mechanism and let only the data collected at different depths per site be used directly for inference.

In the application that follows, such data are analyzed by introducing a multipath change-point model. Andrews and Herzberg (1985) provide a set of observations collected at the West Side Field Station of the University of California, located in Fresno County, 40 miles southwest of Fresno. Twenty soil profile samples were obtained by randomly selecting twenty sites over a 150-hectare region. The analysis of Section 3 was based on 11 sites, located in geologically similar regions. At each site three readings were taken (on the percentages of sand, clay and silt) at each of 12 successive depths. Because the depths are separated by only 15 cm, the 12 observations within each type (sand, clay and

silt) are ostensibly dependent. Here the component silt is considered, since a preliminary analysis on clay and sand showed no change. At several of the sites the first differences of the percentages of silt appear to change abruptly, but the depth of the change is not necessarily constant from site to site. By imposing an autoregressive model which undergoes a change at a random depth, the distribution of the depth of change, as well as the remaining parameters of the autoregressive process, is estimated.

In Section 2, the model is presented and the method of maximum likelihood used to estimate the parameters. The sedimentology problem is analyzed in Section 3. The results of simulations of autoregressive processes of order 1, presented in Section 4, show the method to work well, even in the important case when there are few sequences.

2. THE MODEL AND ESTIMATION

A sequence of random variables $X_1, X_2, \ldots, X_N$ is said to have a change point at τ if $X_1, X_2, \ldots, X_\tau$ follow a distribution F_1, while $X_{\tau+1}, X_{\tau+2}, \ldots, X_N$ follow another distribution F_2. Statistical inference about an unknown change point in a single sequence of observations, known as the *single-path change-point problem*, has been extensively studied. Most of the single-path literature is devoted to sequences of independent random variables (Hinkley 1970, Smith 1975, Pettitt 1979, Worsley 1986). Picard (1985) investigated the asymptotic (number of observations per sequence $\to \infty$) behaviour of the maximum-likelihood estimator of the change point in single-path time-series models. In the context of a problem in material accountancy, Henderson (1986) examined correlated change-point data with arbitrary but known correlation matrix. Smith (1977) considered the same setting with unknown correlation matrix from a Bayesian viewpoint.

In the multipath setting the data consist of several sequences, each one containing a possibly different change point. Recent work on the multipath problem has focused on independent sample path segments of independent random variables. Joseph and Wolfson (1992) introduced the problem and reviewed a range of possible estimation approaches. The consistency of maximum-likelihood estimators of the change point was addressed by Joseph and Wolfson (1993), and Joseph *et al.* (1996) investigated Bayesian methods.

Many applications of the multipath change-point problem, however, arise from data which are dependent. In this paper, a multipath change-point model is introduced by extending the earlier work of Joseph and Wolfson (1993) to correlated observations arising from panel data.

Conditional on $\tau_1, \tau_2, \ldots, \tau_M$, consider the array of random variables

$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1\tau_1} & X_{1(\tau_1+1)} & \cdots & X_{1N} \\ X_{21} & X_{22} & \cdots & X_{2\tau_2} & X_{2(\tau_2+1)} & \cdots & X_{2N} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ X_{M1} & X_{M2} & \cdots & X_{M\tau_M} & X_{M(\tau_M+1)} & \cdots & X_{MN} \end{pmatrix}, \tag{1}$$

where the observations from the ith row, $i = 1, 2, \ldots, M$, arise from an autoregressive process of order p, AR(p), which possibly undergoes a change in its mean and/or its autocovariance function. To keep notation simple, row subscripts are suppressed whenever there is no risk of ambiguity.

Specifically, the ith row in the array (1) is a sample path segment from an AR(p)

process, $\{X_t, t = 1, 2, \ldots\}$, defined by the usual stochastic difference equations:

$$\begin{aligned} X_t - \mu &= \sum_{u=1}^{p} \phi_u (X_{t-u} - \mu) + \epsilon_t \qquad \text{for } p+1 \le \tau,\ p+1 \le t \le \tau \\ &\qquad\qquad \text{or } \tau \ge N - p + 1,\ t \ge p+1, \\ X_t - \mu' &= \sum_{u=1}^{p} \phi'_u (X_{t-u} - \mu') + \epsilon_t \qquad \text{for } p+1 \le \tau,\ \tau+1 \le t \le N, \\ &\qquad\qquad \text{or } \tau \le p,\ t \ge p+1, \end{aligned} \tag{2}$$

where $\{\epsilon_t\}$ are independent and identically distributed (i.i.d.) random variables, independent of $X_{t-1}, X_{t-2}, \ldots$ for all $t = 2, 3, \ldots$. Suppose also, as is customary, that $X_1, X_2, \ldots, X_p$ have a multivariate normal distribution.

The time point τ is called the change point even through it is assumed that no change occurs if $\tau \le p$ or $\tau \ge N - p + 1$. Different change points τ_i are allowed from row to row. The τ_i's are assumed to be i.i.d. with common probability function $\alpha_k = P(\tau_i = k)$, $k = 1, 2, \ldots, N$. Joseph and Wolfson (1993) discuss this assumption in the independent-observations setting.

The inference problem are:

(i) to estimate α_k, $k = 1, 2, \ldots, N$, and
(ii) to estimate μ, μ', ϕ_1, $\phi_2, \ldots, \phi_p$, ϕ'_1, $\phi'_2, \ldots, \phi'_p$, and σ_ϵ^2.

2.1. Maximum-Likelihood Estimation: The Likelihood Function.

As above, row subscripts are suppressed here for ease of notation. For each row, conditional on $\tau = k$, the sequence $X_1, X_2, \ldots, X_N$ defines a sample path segment from a Gaussian process. Hence, conditional on $\tau = k$, the variables $X_1, X_2, \ldots, X_k$ have a multivariate normal distribution with k-dimensional mean vector $\boldsymbol{\mu} = (\mu, \mu, \ldots, \mu)$ and covariance matrix $\boldsymbol{\Sigma}$, say, while $X_{k+1}, X_{k+2}, \ldots, X_N$ have a multivariate normal distribution with $N - k$-dimensional mean vector $\boldsymbol{\mu}' = (\mu', \mu', \ldots, \mu')$ and covariance matrix $\boldsymbol{\Sigma}'$, say.

Define $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)$ and $\boldsymbol{\phi}' = (\phi'_1, \phi'_2, \ldots, \phi'_p)$. Let

$$l_k(\mathbf{x}) = l_k(x_1, x_2, \ldots, x_N;\ \mu, \mu', \boldsymbol{\phi}, \boldsymbol{\phi}', \sigma_\epsilon^2)$$

be the full observed likelihood conditional on $\tau = k$, for a particular sample path. Let

$$l_k(\mathbf{x}_{l,k}) = l_k(x_1, x_2, \ldots, x_k;\ \mu, \boldsymbol{\phi}, \sigma_\epsilon^2)$$

be the observed likelihood of the first k observations, conditional on $\tau = k$. Let

$$l_k(\mathbf{x}_{k+1,N} | \mathbf{x}_{1,k}) = l_k(x_{k+1}, x_{k+2}, \ldots, x_N | x_1, x_2, \ldots, x_k;\ \mu, \mu', \boldsymbol{\phi}, \boldsymbol{\phi}', \sigma_\epsilon^2)$$

be the conditional observed likelihood of the last $N - k$ observations, given $x_1, x_2, \ldots, x_k$ and $\tau = k$.

By assumption, $l_k(\mathbf{x})$ is a function of μ', $\boldsymbol{\phi}'$, σ_ϵ^2 only, for $1 \le k \le p$, and is a function of μ, $\boldsymbol{\phi}$, σ_ϵ^2 only, for $N - p + 1 \le k \le N$.

The likelihood for the ith row is given by

$$l({}_i\mathbf{x}) = \sum_{k=1}^{N} \alpha_k l_k({}_i\mathbf{x}), \tag{3}$$

and the full likelihood for all M sequences is then given by

$$l(\mathbf{X}) = \prod_{i=1}^{M} l(_i\mathbf{x}), \tag{4}$$

where $\mathbf{X}$ is the observed array arising from (1). Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)$.

Equation (3) allows us to interpret the problem of estimation of μ, μ', $\boldsymbol{\phi}$, $\boldsymbol{\phi}'$, σ_ϵ^2 and $\boldsymbol{\alpha}$ as a mixture problem with unknown mixing parameter $\boldsymbol{\alpha}$. This leads to an iterative procedure based on that suggested by Peters and Walker (1978), who obtained maximum-likelihood estimators in a multivariate normal setting.

However, unmodified exploitation of the multivariate normality of all joint and conditional distributions leads to intractable expressions; in typical time-series settings the covariance matrices are of high dimension. It is therefore essential to make use of the autoregressive nature of the observations within each row, conditional on the change point in that row. Unconditionally, therefore, Equation (3) describes a mixture of multivariate normal distributions rendering adaption of the methods of Srivastava and Worsley (1986) inappropriate.

Following Box and Jenkins (1976, Section A7.5), let $M_N^{(p)}$ be the product of the error variance and the inverse of the autocovariance matrix of a sequence $X_1, X_2, \ldots, X_N$, generated by a stationary AR(p) process. In particular,

$$M_p^{(p)} = \{m_{rs}^{(p\times p)}\} = \sigma_\epsilon^2 \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{pmatrix}^{-1},$$

where $\gamma_{|r-s|} = Cov(X_r, X_s)$, and $m_{rs}^{(p\times p)}$ is the (r,s) element of $M_p^{(p)}$.

Let

$$S_{p+1,N}(\mu, \boldsymbol{\phi}) = \sum_{r=1}^{p}\sum_{s=1}^{p} m_{rs}^{(p\times p)}(x_r - \mu)(x_s - \mu) + \sum_{t=p+1}^{N} \{(x_t - \mu) - \phi_1(x_{t-1} - \mu) - \cdots - \phi_p(x_{t-p} - \mu)\}^2.$$

Let $|M_N^{(p)}|$ denote the determinant of $M_N^{(p)}$. The likelihood of $(x_1, x_2, \ldots, x_N)$ is then

$$l(\mathbf{x}) = \left(\frac{1}{2\pi\sigma_\epsilon^2}\right)^{N/2} |M_p^{(p)}|^{\frac{1}{2}} \exp\left(-\frac{S_{p+1,N}(\mu, \boldsymbol{\phi})}{2\sigma_\epsilon^2}\right). \tag{5}$$

The expression (5) depends on the awkward $M_N^{(p)}$ only through $M_p^{(p)}$, which may be computed by exploiting its double symmetry.

For each segment of an AR(p) process with a change at k, $p + 1 \le k \le N - p$, write

$$l(\mathbf{x}) = l_k(\mathbf{x}_{k+1,N}|\mathbf{x}_{l,k}) l_k(\mathbf{x}_{l,k}). \tag{6}$$

The second term on the r.h.s. of (6) is, because of (5), equal to

$$\left(\frac{1}{2\pi\sigma_\epsilon^2}\right)^{k/2} |M_p^{(p)}|^{\frac{1}{2}} \exp\left(-\frac{S_{p+1,k}(\mu, \boldsymbol{\phi})}{2\sigma_\epsilon^2}\right). \tag{7}$$

Next, because the process is Markov of order p, the first term on the r.h.s. of (6) can be written as

$$\left(\frac{1}{2\pi\sigma_\epsilon^2}\right)^{(N-k)/2} \exp\left(\frac{1}{2\sigma_\epsilon^2}\sum_{t=k+1}^{N}\{x_t-\mu'-\phi_1'(x_{t-1}-\mu')-\cdots-\phi_p'(x_{t-p}-\mu')\}^2\right). \quad (8)$$

For $k \leq p$, the likelihood for a single row is given by the expression (6) with $\mu = \mu'$ and $\boldsymbol{\phi} = \boldsymbol{\phi}'$. For $k \geq N - p + 1$, the likelihood is (6) with $\mu = \mu$ and autoregressive parameter vector $\boldsymbol{\phi}$, i.e.,

$$l(\mathbf{x}) = \left(\frac{1}{2\pi\sigma_\epsilon^2}\right)^{N/2} |M_p^{(p)}|^{\frac{1}{2}} \exp\left(-\frac{S_{1,N}(\mu', \boldsymbol{\phi}')}{2\sigma_\epsilon^2}\right) \qquad \text{for} \quad k \leq p \qquad (9)$$

and

$$l(\mathbf{x}) = \left(\frac{1}{2\pi\sigma_\epsilon^2}\right)^{N/2} |M_p^{(p)}|^{\frac{1}{2}} \exp\left(-\frac{S_{1,N}(\mu, \boldsymbol{\phi})}{2\sigma_\epsilon^2}\right) \qquad \text{for} \quad k \geq p - p + 1. \qquad (10)$$

The maximum-likelihood estimators of μ, μ', $\boldsymbol{\phi}$, $\boldsymbol{\phi}'$, σ_ϵ^2 and $\boldsymbol{\alpha}$ are obtained from a set of equations which must be solved iteratively. The approach is equivalent to that of the EM algorithm for mixture problems (Dempster, Laird and Rubin 1977). Here, however, the equations are derived by equating the partial derivatives with respect to the parameters of $L(X) = \ln l(X)$ to zero. Tedious calculations, which depend on standard time-series approximations, lead to final iterative equations given in the Appendix. While these equations would have defied solution before the advent of modern high-speed computers, that is no longer the case.

2.2. *Consistency.*

It would seem as if consistency of the maximum-likelihood estimators, $\hat{\mu}$, $\hat{\mu}'$, $\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\phi}}'$, $\hat{\sigma}_\epsilon^2$ and $\hat{\boldsymbol{\alpha}}$ might be established by standard maximum-likelihood theory for problems with finitely many parameters, but because the information matrix in the present setting is intractable, this approach is not feasible. The only alternative seems to be to use a compactness argument based on that proposed by Wald (1948, 1949). Indeed, the proof given below follows from that of Redner (1981), who modified Wald's results on consistency. We have assumed compactness of the parameter space to streamline the exposition by making Redner's Theorem 5 directly applicable. For practical purposes, this assumption is almost always reasonable, as *a priori* one may in practice confine the parameters to some compact set. Dropping this assumption would necessitate the definition of a suitable metric as in Joseph and Wolfson (1993), and the extension of the parameter space to include points at infinity, so that the compactness argument of Wald could be used.

THEOREM 1. *Let* $l(X) = l(X; \theta)$ *be the likelihood defined by Equations* (3) *and* (4), *where* $\boldsymbol{\theta} = (\mu, \mu', \boldsymbol{\phi}, \boldsymbol{\phi}', \sigma_\epsilon^2, \boldsymbol{\alpha}) \in \Omega$, *the parameter space of* θ. *Suppose that* Ω *is a compact subset of* $\mathbb{R}^{3+3p+N}$. *Let* $\boldsymbol{\theta}^*$ *be the true parameter. Then the maximum-likelihood estimator* $\hat{\boldsymbol{\theta}}$ *is strongly consistent for* $\boldsymbol{\theta}^*$ *as* $M \longrightarrow \infty$.

Proof. Redner (1981, Theorem 5) establishes consistency of $\hat{\boldsymbol{\theta}}$ for mixture families when identifiability is violated. Lack of identifiability necessitates the consideration of quotient topological space, with equivalence classes defined by the equivalence relation $\boldsymbol{\theta}_1 \sim \boldsymbol{\theta}_2$

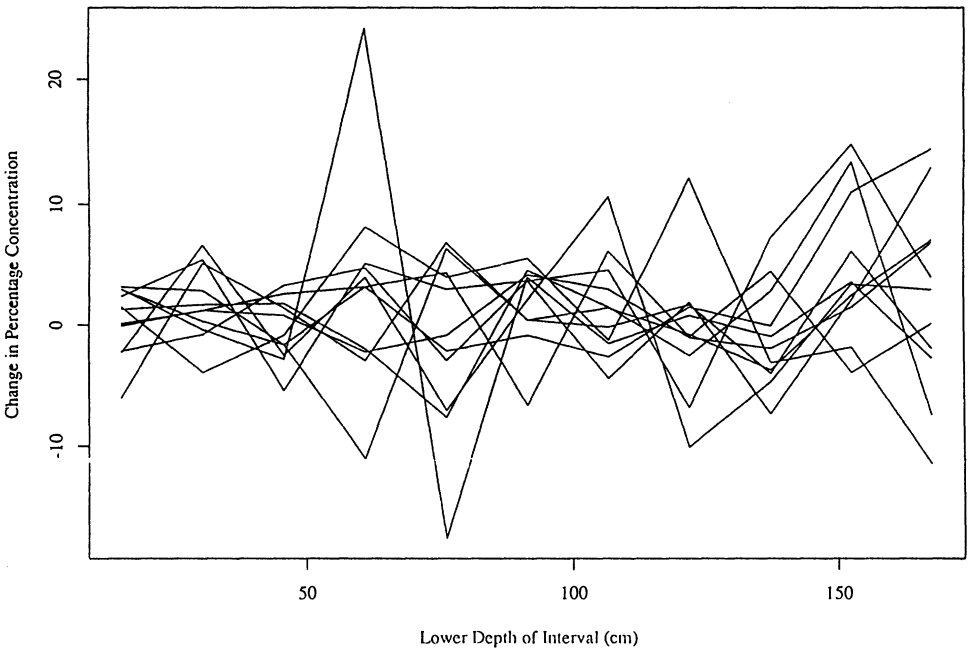

FIGURE 1: West Side Field Station silt concentration — first differences.

if and only if $\mu_{\theta_1} = \mu_{\theta_2}$. Here, μ_θ is the mixture of multivariate normal densities defined by Equation (3) of Section 1. The convergence of Redner's Theorem 5 is in this quotient space. When there is identifiability, however, the quotient space reduces to the space Ω itself and the convergence reduces to classical almost sure convergence.

In the setting of this paper, conditional on $\tau = k$, the joint distribution of the observations in any row is multivariate normal. Yakowitz and Spragins (1968) show that mixtures of multivariate normal densities are identifiable. As is alluded to by Redner, the compactness of Ω ensures that his conditions 1, 2a, 4′ and 5 are satisfied. Hence $\hat{\boldsymbol{\theta}} \longrightarrow \boldsymbol{\theta}^*$ almost surely as $M \longrightarrow \infty$. □

NOTE. The convergence here is ensured by allowing the number of *paths*, M, to tend to infinity, while the number of observations per path remain fixed.

3. SOIL PROFILES

3.1. The Data.

Figure 1 shows plots of the differences in silt concentrations with increasing depth. All 11 sequences from the same geographic region are presented. The data were differenced to remove the increasing trend in the proportion of silt.

3.2. Model and Notation.

The differences were assumed to follow the model of Section 2 with the autoregressive parameter $p = 1$. A rough analysis, carried out by averaging the 11 series, justifies the choice of an AR(1) model with a change point; the estimated autocorrelation function of the first differences decays, and only the first term of the partial autocorrelation function approaches significance. A more careful approach to model selection which takes into

account a putative change point is not possible on such a small data set, and would also entail the examination of residuals arising from a mixture model. Let α_i, $i = 1, 2, \ldots, 10$, denote the probability that the rate of change of silt proportion alters after a depth of $15i$ cm. The probability of no change is given by α_{11}. Let μ and μ' denote the expected differences in silt proportion before and after the change in each sequence. Let σ_ϵ^2 denote the error variance of the differences. Let ϕ and ϕ' denote the autoregressive parameters in the AR(1) models describing the differences before and after the change.

3.3. Test for a Change Point.

In order to first test for the presence of a change point, the statistic $LR = -2 \log \lambda$ was calculated, where λ is the ratio of the maximized likelihood under an AR(1) model with no change to the maximized likelihood under the model of Section 2.

Under the hypothesis of no change, the asymptotic distribution of the LR is not the usual chi-square, the model of Section 2 being a mixture model (Titterington, Smith and Makov 1985, pp. 154–155). The null distribution of the LR was obtained via a parametric bootstrap. See Figure 2. The observed LR was 13.48, yielding a p-value of 0.04, which lends credence the change-point model. It is interesting to note that the bootstrap null LR distribution has a shape similar to a chi-square distributions, but with a point mass at zero. This phenomenon has been conjectured as a general result for the distribution of the LR statistic in order-testing problems for mixtures of exponential family distributions (Böehning 1992).

3.4. Results.

It was found that $\hat{\mu} = 0.69$, $\hat{\mu}' = 7.87$, $\hat{\phi} = -0.42$, $\hat{\phi}' = 0.18$, $\hat{\sigma}_\epsilon = 4.8$, $\hat{\sigma}_9 = 0.41$, $\hat{\alpha}_{11} = 0.59$, and $\hat{\alpha}_i = 0$, $i \neq 9, 11$. The interpretation is that while an estimated 59% of sites in this region would not have a change in mean difference, remaining constant at an increase of 0.69% per 15-cm increase in depth, the other 41% do change. This change is estimated to occur at a depth of 135 cm, and is relatively large, in that the postchange mean difference is 7.87% per 15-cm increase in depth. The estimated standard deviation of about 5% indicates that the change in mean difference is approximately of the order of one standard deviation. The negative value for $\hat{\phi}$ may indicate that a larger than average difference at one depth is followed by a smaller difference at the next depth. The estimated value for ϕ' is more difficult to interpret, since it is based on short segments.

Standard goodness-of-fit criteria for time series such as those based on estimated residuals seem not to apply to models with an unknown change; the unknown random change point in each path precludes the calculation of residuals. The model of Section 2 was used for several reasons: (i) it is plausible and parsimonious, (ii) higher-order ARMA models with a change point require extensive data to fit, and (iii) estimation for ARMA models is based on iterative least-squares analysis of the residual sum of squares. In our change-point model the observations in each row arise as a *mixture* of multivariate normal random variables, so that the least-squares methods are no longer equivalent to maximum likelihood. Our approach is through the likelihood and the EM algorithm.

3.5. Practical Implementation.

Since the likelihood surface may be multimodal, the EM algorithm may not converge to the global maximum. To increase the probability of attaining the true maximum, the EM algorithm was implemented starting from 25 different data-dependent initial sets of parameter values. The final parameter estimate with the highest likelihood value was

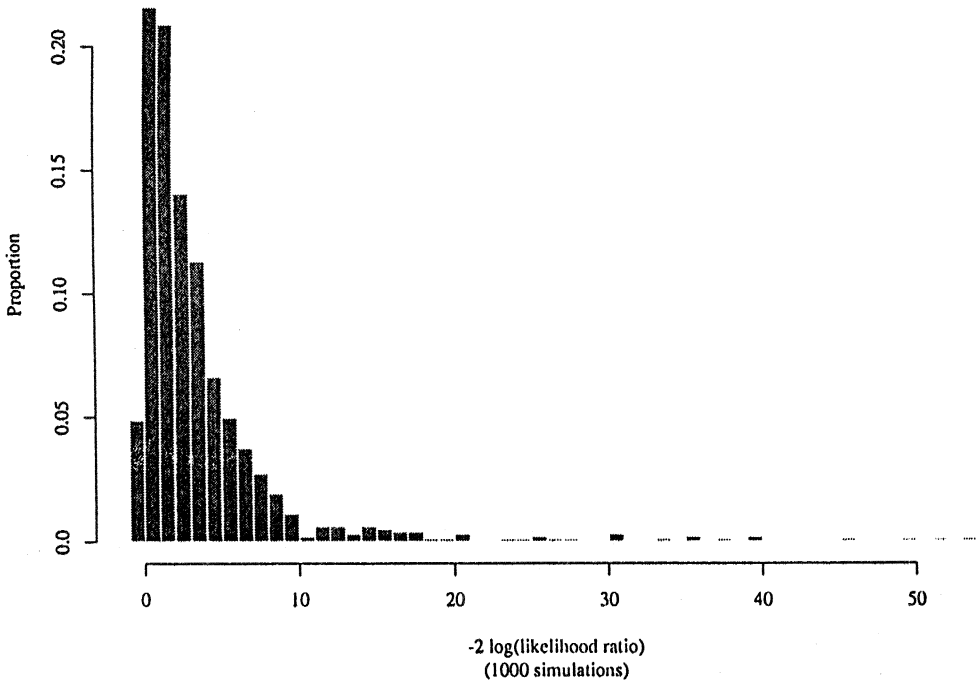

FIGURE 2: West Side Field Station silt concentration — LR bootstrap distribution.

then selected as the best contender for the global maximum. In order to explore the multimodality of the likelihood surface, standard techniques from principal-components analysis were used on the sample correlation matrix of the final estimates. Twenty-five vectors, each connecting an initial guess with the corresponding final estimate, were created. These were plotted in the two-dimensional orthogonal subspace created by the first two principal components. The results, shown in Figure 3, while not constituting a proof that the global maximum has been attained, do suggest that the same peak is reached from most of the starting values, and that this value corresponds to the observed maximum of the likelihood function. Another approach for choosing starting values is given by Lindsay and Brasak (1993).

4. SIMULATIONS

Simulations were performed to evaluate the methods for small and moderate sample sizes. The AR(1) case was chosen to conform with the example of Section 3. Referring to (1), fix M sequences of length $N = 40$, where $M = 10$, 30, 100 or 500. Two sets of choices for the normal parameters were considered: a change from $\mu = 0$ to $\mu' = 1$, and a change from $\mu = 0$ to $\mu' = 2$. Throughout, $\sigma_\epsilon^2 = 1$ was used. Four sets of autoregressive parameters were chosen: $\phi = \phi' = 0.5$, a change from $\phi = 0.5$ to $\phi' = 0.75$, $\phi = \phi' = -0.5$, and a change from $\phi = -0.5$ to $\phi' = -0.75$.

For $N = 40$, two choices for the change-point distribution were considered:

1. $U(15, 24)$, a uniform distribution on the integers from 15 to 24, so that $\text{pr}\{\tau = k\} = 0.1$, $k = 15, \ldots, 24$, and zero elsewhere.

2. $T(16, 24)$, a "tent-shaped" function, with peak at $\tau = 20$, and sloping down linearly to zero at $\tau = 15$ and $\tau = 25$. Hence $\text{pr}\{\tau = 20\} = 0.2$, $\text{pr}\{\tau = 19\} = \text{pr}\{\tau = 21\} =$

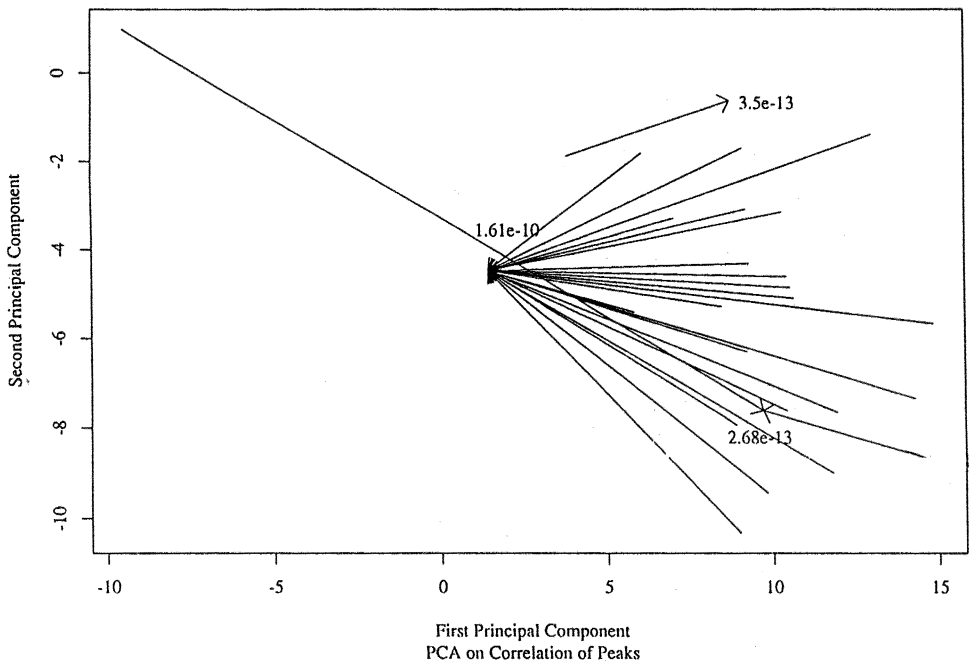

FIGURE 3: West Side Field Station silt concentration — likelihood surface.

0.16, $\text{pr}\{\tau = 18\} = \text{pr}\{\tau = 22\} = 0.12$, $\text{pr}\{\tau = 17\} = \text{pr}\{\tau = 23\} = 0.08$, and $\text{pr}\{\tau = 16\} = \text{pr}\{\tau = 24\} = 0.04$, all other choices for k having zero probability.

These choices cover two likely possible shapes for the distributions of τ. In the first, the change is equally likely to occur anywhere in a specified region, but nothing is known about the relative probabilities within the region. The second covers the case where the most likely location for the change is known, but points near this value are also possible, with decreasing probability further from the centre.

The two choices for the change-point distribution, combined with two sets of normal parameters, four choices of autoregressive parameters c, and four choices for M, give 64 different situations. Each of these 64 combinations was simulated 300 times.

Outcome measures for the simulations include the average error in $\hat{\boldsymbol{\alpha}}$, defined by $|\alpha_k - \hat{\alpha}_k|$ averaged over all N possible locations for k and over all 300 simulations, and various statistics summarizing the largest error, defined by $\sup_{1 \le k \le N} |\alpha_k - \hat{\alpha}_k|$. These included the mean, median and range of the largest error over the 300 simulations. The stopping criterion was

$$\sum_{k=1}^{N} |\hat{\alpha}_k^{l+1} - \hat{\alpha}_k^{l}| + |\mu^{l+1} - \mu^{l}| + |\mu'^{l+1} - \mu'^{l}| + |\phi'^{l+1} - \phi'^{l}| + |\sigma^{2\,l+1} - \sigma^{2\,l}| \le 0.00001,$$

where the superscript l represents the iteration number.

The procedure for obtaining initial estimates was as follows: The estimate for τ was chosen as that value that maximized the difference between the average before-change and after-change sample means. The values for μ and μ' were then initialized to those mean values. The autoregressive coefficients were obtained by averaging the quotients of successive centered observations. Finally, the initial value of σ_ϵ^2 was obtained from the residual mean squares given the other initial estimates.

TABLE 1: AR(1) change-point simulations with positive AR coefficients. 300 simulations per set of parameters. Average and maximum errors over the 300 simulations and 40 locations for the change, and average number of iterations until convergence are given.

No.	M	μ	μ'	ϕ	ϕ'	PDF	Avg. err.	Max. err.	Avg. iters.
1	10	0.0	1.0	0.50	0.50	$U(15, 24)$	0.03592	0.84639	235
2	10	0.0	1.0	0.50	0.50	$T(16, 24)$	0.03384	0.71562	212
3	10	0.0	1.0	0.50	0.75	$U(15, 24)$	0.03648	0.83788	214
4	10	0.0	1.0	0.50	0.75	$T(16, 24)$	0.03502	0.95091	214
5	10	0.0	2.0	0.50	0.50	$U(15, 24)$	0.01994	0.47551	101
6	10	0.0	2.0	0.50	0.50	$T(16, 24)$	0.01889	0.47376	97
7	10	0.0	2.0	0.50	0.75	$U(15, 24)$	0.02165	0.55589	101
8	10	0.0	2.0	0.50	0.75	$T(16, 24)$	0.02003	0.42218	92
9	30	0.0	1.0	0.50	0.50	$U(15, 24)$	0.02835	0.56496	324
10	30	0.0	1.0	0.50	0.50	$T(16, 24)$	0.02655	0.46498	305
11	30	0.0	1.0	0.50	0.75	$U(15, 24)$	0.02864	0.51681	328
12	30	0.0	1.0	0.50	0.75	$T(16, 24)$	0.02655	0.61101	319
13	30	0.0	2.0	0.50	0.50	$U(15, 24)$	0.01385	0.29487	103
14	30	0.0	2.0	0.50	0.50	$T(16, 24)$	0.01276	0.34053	101
15	30	0.0	2.0	0.50	0.75	$U(15, 24)$	0.01515	0.27528	125
16	30	0.0	2.0	0.50	0.75	$T(16, 24)$	0.01360	0.25468	129
17	100	0.0	1.0	0.50	0.50	$U(15, 24)$	0.01964	0.28713	337
18	100	0.0	1.0	0.50	0.50	$T(16, 24)$	0.01806	0.33816	311
19	100	0.0	1.0	0.50	0.75	$U(15, 24)$	0.01992	0.30186	335
20	100	0.0	1.0	0.50	0.75	$T(16, 24)$	0.01690	0.31369	323
21	100	0.0	2.0	0.50	0.50	$U(15, 24)$	0.00792	0.14650	80
22	100	0.0	2.0	0.50	0.50	$T(16, 24)$	0.00720	0.13360	74
23	100	0.0	2.0	0.50	0.75	$U(15, 24)$	0.00825	0.13000	97
24	100	0.0	2.0	0.50	0.75	$T(16, 24)$	0.00754	0.13805	92
25	500	0.0	1.0	0.50	0.50	$U(15, 24)$	0.01025	0.17003	292
26	500	0.0	1.0	0.50	0.50	$T(16, 24)$	0.00927	0.14671	296
27	500	0.0	1.0	0.50	0.75	$U(15, 24)$	0.00976	0.12653	303
28	500	0.0	1.0	0.50	0.75	$T(16, 24)$	0.00865	0.14179	306
29	500	0.0	2.0	0.50	0.50	$U(15, 24)$	0.00340	0.05891	67
30	500	0.0	2.0	0.50	0.50	$T(16, 24)$	0.00312	0.06605	64
31	500	0.0	2.0	0.50	0.75	$U(15, 24)$	0.00350	0.05959	86
32	500	0.0	2.0	0.50	0.75	$T(16, 24)$	0.00324	0.05597	84

In all cases, maximum-likelihood estimates were computed via the equations given in the Appendix, programmed in Fortran, and run on a Sun Microsystems SPARCstation ELC.

The results of the simulations are tabulated in Tables 1 and 2. The results for negative AR coefficients were slightly better. This is because the means of autoregressive processes are more accurately estimated when the autoregressive coefficients are negative than when they are positive. The reason for this is that in any given finite sequence of observations, the series with negative coefficients is more likely to be centered around the mean, since it typically will alternate values above and below the mean. However, series with positive coefficients have a tendency to drift above or below the mean for longer periods of time.

As expected, errors in estimation of the change-point distribution decreased appreciably as M increased. For $M = 500$ even the maximum error was commonly much less than 0.1, while maximum errors of 0.5 were typical when $M = 10$. As Figure 4 shows, however, even with $M = 10$, the estimated probability tended to be in the same neighbourhood as the true probability, but was often moved over by one or two indices along the x-

TABLE 2: AR(1) change-point simulations with negative AR coefficients. 300 simulations per set of parameters. Average and maximum errors over the 300 simulations and 40 locations for the change, and average number of iterations until convergence are given.

No.	M	μ	μ'	ϕ	ϕ'	PDF	Avg. err.	Max. err.	Avg. iters.
33	10	0.0	1.0	−0.50	−0.50	$U(15, 24)$	0.02595	0.81924	217
34	10	0.0	1.0	−0.50	−0.50	$T(16, 24)$	0.02591	0.80507	211
35	10	0.0	1.0	−0.50	−0.75	$U(15, 24)$	0.02529	0.89983	189
36	10	0.0	1.0	−0.50	−0.75	$T(16, 24)$	0.02356	0.89967	179
37	10	0.0	2.0	−0.50	−0.50	$U(15, 24)$	0.01381	0.41123	80
38	10	0.0	2.0	−0.50	−0.50	$T(16, 24)$	0.01313	0.51861	62
39	10	0.0	2.0	−0.50	−0.75	$U(15, 24)$	0.01196	0.34317	51
40	10	0.0	2.0	−0.50	−0.75	$T(16, 24)$	0.01129	0.65759	51
41	30	0.0	1.0	−0.50	−0.50	$U(15, 24)$	0.02084	0.47753	280
42	30	0.0	1.0	−0.50	−0.50	$T(16, 24)$	0.01965	0.47958	268
43	30	0.0	1.0	−0.50	−0.75	$U(15, 24)$	0.01907	0.39582	223
44	30	0.0	1.0	−0.50	−0.75	$T(16, 24)$	0.01813	0.40923	201
45	30	0.0	2.0	−0.50	−0.50	$U(15, 24)$	0.00995	0.22326	68
46	30	0.0	2.0	−0.50	−0.50	$T(16, 24)$	0.00923	0.29518	58
47	30	0.0	2.0	−0.50	−0.75	$U(15, 24)$	0.00835	0.18494	50
48	30	0.0	2.0	−0.50	−0.75	$T(16, 24)$	0.00772	0.23582	46
49	100	0.0	1.0	−0.50	−0.50	$U(15, 24)$	0.01482	0.24080	304
50	100	0.0	1.0	−0.50	−0.50	$T(16, 24)$	0.01355	0.30318	275
51	100	0.0	1.0	−0.50	−0.75	$U(15, 24)$	0.01285	0.28552	212
52	100	0.0	1.0	−0.50	−0.75	$T(16, 24)$	0.01182	0.24505	195
53	100	0.0	2.0	−0.50	−0.50	$U(15, 24)$	0.00536	0.11264	41
54	100	0.0	2.0	−0.50	−0.50	$T(16, 24)$	0.00511	0.13533	37
55	100	0.0	2.0	−0.50	−0.75	$U(15, 24)$	0.00442	0.08787	29
56	100	0.0	2.0	−0.50	−0.75	$T(16, 24)$	0.00422	0.10360	28
57	500	0.0	1.0	−0.50	−0.50	$U(15, 24)$	0.00791	0.15819	228
58	500	0.0	1.0	−0.50	−0.50	$T(16, 24)$	0.00708	0.16822	191
59	500	0.0	1.0	−0.50	−0.75	$U(15, 24)$	0.00597	0.10582	126
60	500	0.0	1.0	−0.50	−0.75	$T(16, 24)$	0.00584	0.12032	116
61	500	0.0	2.0	−0.50	−0.50	$U(15, 24)$	0.00240	0.05064	31
62	500	0.0	2.0	−0.50	−0.50	$T(16, 24)$	0.00230	0.05229	26
63	500	0.0	2.0	−0.50	−0.75	$U(15, 24)$	0.00203	0.04806	25
64	500	0.0	2.0	−0.50	−0.75	$T(16, 24)$	0.00189	0.04380	21

axis. Under most circumstances, these errors should not greatly decrease the value of the analysis. As M increases, the change points as well as μ, μ', ϕ, ϕ' and σ_ϵ^2 are all estimated with improved precision.

Approximately 200 to 300 iterations were required for convergence. This quantity varied with the difference between μ and μ', requiring fewer iterations for larger differences. Also, more iterations were required on average as M decreased.

5. CONCLUDING REMARKS

The multipath change-point design arises in diverse fields such as geology and medicine, including clinical trials. For example, Lyle *et al.* (1987) decribe a randomized controlled trial to examine the effect of calcium supplementation on blood pressure. After a 4-week baseline period of weekly blood-pressure measurements, 75 men were randomly assigned to either a treatment (calcium) or a placebo group for a 12-week period. One of the aims of the trial was to establish if calcium supplementation lowers mean arterial pressure. The time to reaction for each patient may be different, and not all patients may respond. The methods of Section 2 may be used to estimate the distri-

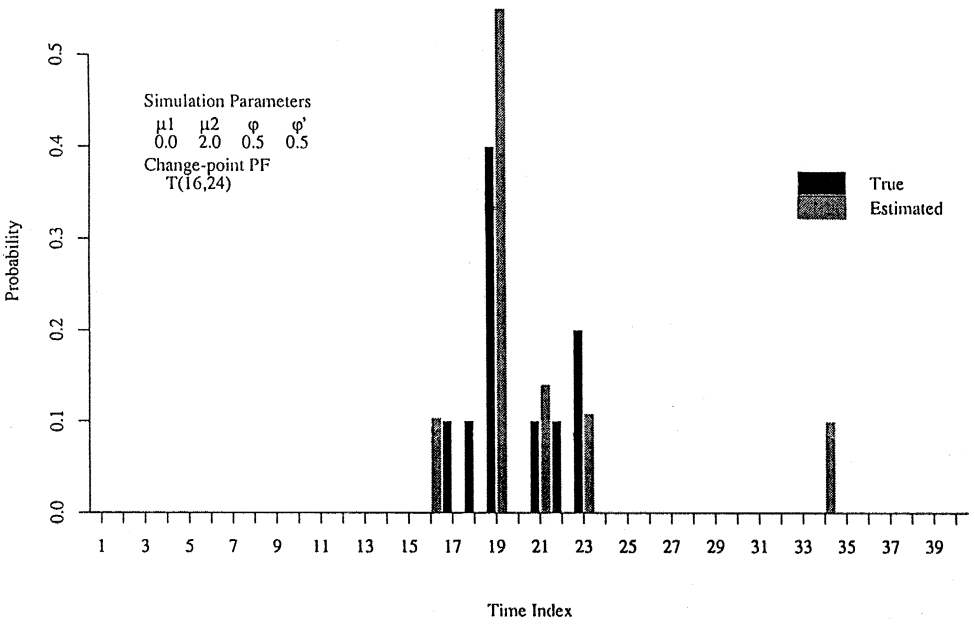

FIGURE 4: Results from one simulation ($M = 10$).

bution of a time to a change after treatment, as well as the before and after mean blood pressures.

Frequently, it is not reasonable to assume that the data within each sequence are independent, but, conditional on the change point in that sequence, can be modelled by an autoregressive process. The model presented here is limited to the case of equally spaced normally distributed data, with unknown but assumed identical means in each sequence. The model could also be extended to include different means across sequences, but identifiability problems occur unless one is willing to impose the condition that there *must* be a change point $p+1 \leq \tau \leq N-p+1$. The case of different means in dependent data with $\Pr\{\text{no change point}\} > 0$, as well as allowance for different pre- and postchange variances, is also important, and could be explored in a Bayesian framework, in a similar fashion to the model of Joseph *et al.* (1996) for independent sequences. The case of unequally spaced autoregressive data could be approached with methods such as those in Jones and Boadi-Boateng (1991).

APPENDIX

The iterative equations below are obtained by setting the partial derivatives of the likelihood function equal to zero [Equation (11) may be obtained simply by using Lagrange multipliers]:

$$\alpha_k = \frac{\alpha_k}{M} \sum_{i=1}^{M} \frac{l_k({}_i\mathbf{x})}{\sum_{j=1}^{N} \alpha_j l_j({}_i\mathbf{x})}, \tag{11}$$

$$\mu = \left[\sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=p+1}^{N-p} \alpha_k l_k({}_i\mathbf{x}_{p+1,k} | {}_i\mathbf{x}_{1,p}) l_k({}_i\mathbf{x}_{k+1,N} | {}_i\mathbf{x}_{1,k}) \right.$$

$$\times \left\{ \left(1 - \sum_{u=1}^{p} \phi_u\right) \sum_{t=p+1}^{k} \left({}_i x_t - \sum_{u=1}^{p} \phi_u \, {}_i x_{t-u} \right) \right.$$

$$\left. + \left(\sum_{u=1}^{p} \phi'_u \right) \sum_{t=k+1}^{k+p} \left({}_i x_t - \mu' - \sum_{u=1}^{p} \phi'_u \, {}_i x_{t-u} \right) \right\}$$

$$+ \sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=N-p+1}^{N} \alpha_k l_k({}_i\mathbf{x}_{p+1,N} | {}_i\mathbf{x}_{1,p})$$

$$\times \sum_{t=p+1}^{N} \left({}_i x_t - \sum_{u=1}^{p} \phi_u \, {}_i x_{t-u} \right) \left(1 - \sum_{u=1}^{p} \phi_u\right) \Bigg]$$

$$\Bigg/ \Bigg[\sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=p+1}^{N-p} \left\{ \left(1 - \sum_{u=1}^{p} \phi_u\right)^2 (k-p) + p \left(\sum_{u=1}^{p} \phi'_u \right)^2 \right\}$$

$$\times \alpha_k l_k({}_i\mathbf{x}_{p+1,k} | {}_i\mathbf{x}_{1,p}) l_k({}_i\mathbf{x}_{k+1,N} | {}_i\mathbf{x}_{1,k})$$

$$+ \sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=N-p+1}^{N} \left(1 - \sum_{u=1}^{p} \phi_u\right)^2 (N-p) \alpha_k l_k({}_i\mathbf{x}_{p+1,N} | {}_i\mathbf{x}_{1,p}) \Bigg] \tag{12}$$

$$\mu' = \Bigg[\sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=p+1}^{N-p} \alpha_k l_k({}_i\mathbf{x}_{p+1,k} | {}_i\mathbf{x}_{1,p}) l_k({}_i\mathbf{x}_{k+1,N} | {}_i\mathbf{x}_{1,k})$$

$$\times \left\{ \left(1 - \sum_{u=1}^{p} \phi'_u\right) \sum_{t=k+p+1}^{N} \left({}_i x_t - \sum_{u=1}^{p} \phi'_u \, {}_i x_{t-u} \right) \right.$$

$$\left. + \sum_{t=k+1}^{k+p} \left({}_i x_t - \sum_{u=1}^{p} \phi'_u ({}_i x_{t-u} - \mu) \right) \right\}$$

$$+ \sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=1}^{p} \alpha_k l_k({}_i\mathbf{x}_{p+1,N} | {}_i\mathbf{x}_{1,p})$$

$$\times \sum_{t=p+1}^{N} \left({}_i x_t - \sum_{u=1}^{p} \phi'_u \, {}_i x_{t-u} \right) \left(1 - \sum_{u=1}^{p} \phi'_u\right) \Bigg]$$

$$\Bigg/ \Bigg[\sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=p+1}^{N-p} \alpha_k l_k({}_i\mathbf{x}_{p+1,k} | {}_i\mathbf{x}_{1,p}) l_k({}_i\mathbf{x}_{k+1,N} | {}_i\mathbf{x}_{1,k})$$

$$\times \left\{ (N-k-p) \left(\sum_{u=1}^{p} \phi'_u \right)^2 + p \right\}$$

$$+ \sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=1}^{p} \alpha_k l_k({}_i\mathbf{x}_{p+1,N} | {}_i\mathbf{x}_{1,p}) \left(1 - \sum_{u=1}^{p} \phi'_u\right)^2 (N-p) \Bigg]. \tag{13}$$

For $r = 1, 2, \ldots, p$,

$$\begin{aligned}\phi_r = \Bigg\{ & \sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=p+1}^{N-p} \alpha_k l_k({}_i\mathbf{x}_{p+1,k} | {}_i\mathbf{x}_{1,p}) l_k({}_i\mathbf{x}_{k+1,N} | {}_i\mathbf{x}_{1,k}) \\ & \times \sum_{t=p+1}^{k} ({}_ix_{t-r} - \mu) \Bigg({}_ix_t - \mu - \sum_{u=1,\, u\neq r}^{p} \phi_u({}_ix)t - u - \mu) \Bigg) \\ & + \sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=N-p+1}^{N} \alpha_k l_k({}_i\mathbf{x}_{p+1,N} | {}_i\mathbf{x}_{1,p}) \\ & \times \sum_{t=p+1}^{N} ({}_ix_{t-r} - \mu) \Bigg({}_ix_t - \mu - \sum_{u=1,\, u\neq r}^{p} \phi_u({}_ix_{t-u} - \mu) \Bigg) \Bigg\} \\ & \Bigg/ \Bigg(\sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=p+1}^{N-p} \alpha_k \sum_{t=p+1}^{k} ({}_ix_{t-r} - \mu)^2 l_k({}_i\mathbf{x}_{p+1,k} | {}_i\mathbf{x}_{1,p}) l_k({}_i\mathbf{x}_{k+1,N} | {}_i\mathbf{x}_{1,k}) \\ & + \sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=N=p+1}^{N} \alpha_k \sum_{t=p+1}^{k} ({}_ix_{t-r} - \mu)^2 l_k({}_i\mathbf{x}_{p+1,N} | {}_i\mathbf{x}_{1,p}) \Bigg). \end{aligned} \tag{14}$$

For $r = 1, 2, \ldots, p$,

$$\begin{aligned}\phi'_r = \Bigg[& \sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=p+1}^{N-p} \alpha_k l_k({}_i\mathbf{x}_{p+1,k} | {}_i\mathbf{x}_{1,p}) l_k({}_i\mathbf{x}_{k+1,N} | {}_i\mathbf{x}_{1,k}) \\ & \times \Bigg\{ \sum_{t=k+p+1}^{N} ({}_ix_{t-r} - \mu') \Bigg({}_ix_t - \mu' - \sum_{u=1,\, u\neq r}^{p} \phi'_u({}_ix_{t-u} - \mu') \Bigg) \\ & + \sum_{t=k+1}^{k+p} ({}_ix_{t-r} - \mu) \Bigg({}_ix_t - \mu' - \sum_{u=1,\, u\neq r}^{p} \phi'_u({}_ix_{t-u} - \mu) \Bigg) \Bigg\} \\ & + \sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=1}^{p} \alpha_k l_k({}_i\mathbf{x}_{p+1,N} | {}_i\mathbf{x}_{1,p}) \\ & \times \sum_{t=p+1}^{N} ({}_ix_{t-r} - \mu') \Bigg({}_ix_t - \mu' - \sum_{u=1,\, u\neq r}^{p} \phi'_u({}_ix_{t-u} - \mu') \Bigg) \Bigg] \\ & \Bigg/ \Bigg(\sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=p+1}^{N} \alpha_k \sum_{t=k+1}^{N} ({}_ix_{t-r} - \mu')^2 l_k({}_i\mathbf{x}_{p+1,k} | {}_i\mathbf{x}_{1,p}) l_k({}_i\mathbf{x}_{k+1,N} | {}_i\mathbf{x}_{1,k}) \\ & + \sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=1}^{p} \alpha_k l_k({}_i\mathbf{x}_{p+1,N} | {}_i\mathbf{x}_{1,p}) \sum_{t=p+1}^{N} ({}_ix_{t-r} - \mu')^2 \Bigg), \end{aligned} \tag{15}$$

$$\sigma_\epsilon^2 = \Bigg[\sum_{i=1}^{M} \frac{1}{l({}_i\mathbf{x})} \sum_{k=1}^{p} \alpha_k l_k({}_i\mathbf{x}) \Bigg\{ ({}_i\mathbf{x}_{1,p} - \boldsymbol{\mu}')^T {}_2M_p^{(p)} ({}_i\mathbf{x}_{1,p} - \boldsymbol{\mu}')$$

$$+\sum_{t=p+1}^{N}\left({}_iX_t-\mu'\right)-\sum_{u=1}^{p}\phi'_u({}_iX_{t-u}-\mu')\Big)^2\Big\}$$

$$+\sum_{i=1}^{M}\frac{1}{l({}_i\mathbf{x})}\sum_{k=p+1}^{n-p}\alpha_k l_k({}_i\mathbf{x})\Big\{({}_i\mathbf{x}_{1,p}-\boldsymbol{\mu})^T{}_1M_p^{(p)}({}_i\mathbf{x}_{1,p}-\boldsymbol{\mu})$$

$$+\sum_{t=p+1}^{k}\left({}_iX_t-\mu\right)-\sum_{u=1}^{p}\phi_u({}_iX_{t-u}-\mu)\Big)^2$$

$$+\sum_{t=k+p+1}^{N}\left({}_iX_t-\mu'\right)-\sum_{u=1}^{p}\phi'_u({}_iX_{t-u}-\mu')\Big)^2$$

$$\times\sum_{t=k+1}^{k+p}\left({}_iX_t-\mu'\right)-\sum_{u=1}^{p}\phi'_u({}_iX_{t-u}-\mu)\Big)^2\Big\}$$

$$+\sum_{i=1}^{M}\frac{1}{l({}_i\mathbf{x})}\sum_{k=N-p+1}^{N}\alpha_k l_k({}_i\mathbf{x})\Big\{({}_i\mathbf{x}_{1,p}-\boldsymbol{\mu})^T{}_1M_p^{(p)}({}_i\mathbf{x}_{1,p}-\boldsymbol{\mu})$$

$$+\sum_{t=p+1}^{N}\left({}_iX_t-\mu\right)-\sum_{u=1}^{p}\phi_u({}_iX_{t-u}-\mu)\Big)^2\Big\}\Big]\Big/NM. \tag{16}$$

ACKNOWLEDGEMENTS

The authors wish to thank Marie-Pierre Aoun for her skills in preparing this document.

REFERENCES

Andrews, D.F., and Herzberg, A.M. (1985). *Data. Particle Size Distribution of Soil Profiles.* Springer-Verlag, New York, 119–122.

Baltagi, B.H., Chang, Y-C., and Li, Q. (1992). Monte Carlo evidence on panel data regressions with AR(1) disturbances and an arbitrary variance on the initial observations. *J. Econometrics*, 52, 371–380.

Böehning, D. (1992). The distribution of the likelihood ratio for mixtures of densities for the one parameter exponential family. *XVIth International Biometrics Conference: Proceedings of Contributed Papers*, Hamilton, New Zealand, 64.

Box, G.E.P., and Jenkins, G.M. (1976). *Time Series Analysis.* Holden Day, Oakland.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39, 1–38.

Gong, B.-H, and Sickles, R.C. (1992). Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. *J. Econometrics*, 51, 259–284.

Henderson, R. (1986). Change-point problem with correlated observations, with an application in material accountancy. *Technometrics*, 28, 381–389.

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

Jones, R.H., and Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics*, 47, 161–175.

Joseph, L., and Wolfson, D.B. (1992). Estimation in multi-path change-point problems. *Comm. Statist. Theory Methods*, 21, 897–913.

Joseph, L., and Wolfson, D.B. (1993). Maximum likelihood estimation in the multi-path change-point problem. *Ann. Inst. Statist. Math.*, 45, 511–530.

Joseph, L., Wolfson, D.B., du Berger, R., and Lyle, R. (1996). A change-point analysis of a randomized trial on the effects of calcium supplementation on blood pressure. *Bayesian Biostatistics* (D.A. Berry and D.K. Stangl, *eds.*), Dekker, New York. 617–649.

Kiefer, J., and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.*, 27, 887–906.

Kim, Y.-W., and Basawa, I.V. (1992). Empirical Bayes estimation for first order autoregressive processes. *Austra. J. Statist.*, 34, 105–114.

Lindsay, B.G., and Brasak, P. (1993). Multivariate normal mixtures: A fast consistent method of moments. *J. Amer. Statist. Assoc.*, 88, 468–476.

Lyle, R.M., Melby, C.L., Hyner, G.C., Edmonson, J.W., Miller, J.Z., and Weinberger, M.H. (1987). Blood pressure and metabolic effects of calcium supplementation in normotensive white and black men. *J. Amer. Med. Assoc.*, 257, 1772–1776.

Müller, P., and Rosner, G.L. (1994). A semiparametric Bayesian model with hierarchical mixture priors. Preprint.

Palmquist, B., and Green, D.P. (1992). Estimation of models with correlated measurement errors from panel data. *Sociol. Methods and Res.*, 22, 119–146.

Peters, B.C., and Walker, H.F. (1978). An iterative procedure for obtaining maximum likelihood estimates of the parameters for a mixture of normal distributions. *SIAM J. Appl. Math.*, 35, 362–378.

Pettitt, A.N. (1979). A nonparametric approach to the change-point problem. *Appl. Statist.*, 28, 126–135.

Picard, D. (1985). Testing and estimating change-points in time series. *Adv. Appl. Probab.*, 17, 841–867.

Preston, F.W., and Davis, J.C. (1972). Sedimentary porous materials as a realization of a stochastic process. *Random Processes in Geology* (D.F. Merriam, *ed.*), Springer-Verlag, New York.

Redner, R. (1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *Ann. Statist.*, 9, 225–228.

Schwarzacher, W. (1972a). Stratigraphic implications of random sedimentation. *Random Processes in Geology* (D.F. Merriam, *ed.*), Springer-Verlag, New York.

Schwarzacher, W. (1972b). The semi-Markov process on a general sedimentation model. *Mathematical Models of Sedimentary Processes.* Plenum, New York.

Smith, A.F.M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, 62, 407–416.

Smith, A.F.M. (1977). A Bayesian analysis of some time varying models. *Recent Developments in Statistics* (J. Barra, F. Brodeau, G. Romier and B. Van Cutsen, *eds.*), North Holland, Amsterdam, 257–267.

Srivastava, M.S., and Worsley, K.J. (1986). Likelihood ratio tests for a change in the multivariate normal mean. *J. Amer. Statist. Assoc.*, 81, 199–204.

Tang, S.M., and MacNeill, I.B. (1993). The effect of serial correlation on tests for parameter change at unknown time. *Ann. Statist.*, 21, 552–575.

Titterington, D.M., Smith, A.F.M., and Makov, V.E. (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley, New York.

Wald, A. (1948). Estimation of a parameter when the number of unknown parameters increases indefinitely with the number of observations, *Ann. Math. Statist.*, 19, 220–227.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.*, 20, 595–601.

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

Yakowitz, S.J., and Spragins, J.D. (1968). On the identifiability of finite mixtures. *Ann. Math. Statist.*, 39, 209–214.

Received 5 October 1994
Revised 1 November 1995
Accepted 1 November 1995

Department of Medicine
Montreal General Hospital
Montreal, Quebec
Canada H3G 1A4

Department of Mathematics and Statistics
McGill University
Montreal, Quebec
Canada H3A 2K6