

# Surveillance of a Simple Linear Regression

Abba M. KRIEGER, Moshe POLLAK, and Benjamin YAKIR

---

This article considers an important aspect of the general sequential analysis problem where a process is in control up to some unknown point  $i = \nu - 1$ , after which the distribution from which the observations are generated changes. An extensive sequential analytic literature assumes that the change in distribution is abrupt, for example, from  $N(0, 1)$  to  $N(\mu, 1)$ . There is also an extensive literature that deals with a gradual change in the case where the decision (whether or not a change has occurred) is based on a fixed set of observations, rather than an ongoing process of decision making every time a new observation is obtained. However, there is virtually no literature on the practical case of sequentially detecting a gradual change in distribution (visualize a machine deteriorating gradually). This article considers solutions to this problem. As a first approximation, the gradual change problem can be modeled as a change from a fixed distribution to a model of simple linear regression with respect to time (i.e., there is an abrupt change of slope, from a 0 to a nonzero slope). We study an extension of this case to a general context of sequential detection of a change in the slope of a simple linear regression. The residuals are assumed to be normally distributed. We consider both the case in which the baseline parameters are known and the case in which they are not. Finally, as an application, we monitor for an increase in the rate of global warming.

KEY WORDS: Average run lengths; Control charts; Cusum; Monte Carlo; Statistical process control; Shiriyayev–Roberts.

---

## 1. INTRODUCTION

Imagine that the behavior of a machine is being monitored over time. The concern is that after some point in time, the machine will start to operate incorrectly. The classical sequential analysis problem is to raise an alarm as soon as possible after the machine begins to perform unsatisfactorily, subject to a constraint on the rate of false alarms. Among various policies calibrated to have the same expected number of observations until raising a false alarm, the optimal choice is the one that minimizes the expected number of observations to raising an alarm after the machine begins to perform inadequately.

The classical changepoint problem—that of an abrupt change from one distribution to another in a sequence of independent random variables—is well understood. Efficient detection methods are Cusum and Shiriyayev–Roberts, both of which are known to have optimality properties when both pre-change and post-change distributions are known (Moustakides 1986; Ritov, 1990; Pollak, 1985; Yakir, 1997). These techniques can be adapted to cases where the post-change and/or pre-change distributions are partially or fully unknown to obtain reasonably efficient procedures (Pollak, 1987; Gordon and Pollak, 1997; Yakir, 1998; Lorden and Pollak 2003; and references therein).

However, in many applications it is more natural to assume that the change is gradual. The decline in a company's market share due to a new competitor's arrival is typically gradual; the impact of a social movement is seldom abrupt, even if it is ultimately limited, and the deterioration in the performance of a machine might not happen all at once (Sigal, 1998). In all of these cases, even if the change ultimately leads to a new equilibrium, it may be possible to detect the change in its early stages, where a reasonable description of the process is a change of the slope of the linear regression with respect to time, from 0 to a

nonzero value. These considerations lead us to study the problem of sequential detection of a change in linear regression. In addition, of course, this problem is of interest on its own merits.

Often, simple Cusum procedures have been used (on regression residuals) to detect a change in regression. These procedures are designed for a different problem: detecting an abrupt change from mean 0 to another constant mean. Evaluations of the average run lengths of these procedures that take into account the fact that the post-change mean is not constant have appeared in the literature (Aerne, Champ, and Rigdon 1991; Gan, 1992, 1996). However, procedures designed specifically for detecting a change from one regression to another have not yet been developed. It is this void that we try to fill in this article.

Formally, the problem that we study is expressed as follows. Initially, when the process is in control, it yields independent observations,  $Y_i$ , which are normally distributed with variance  $\sigma^2$  and mean  $(\alpha + \beta x_i)\sigma$ , where  $x_i$  is a fixed scalar regressor. (We choose to parameterize the mean in this form to exploit subsequently an invariance structure.) Should the process go out of control, the distribution of  $Y_i$  will be  $N((\alpha + \beta x_i + \gamma_{i,\nu})\sigma, \sigma^2)$  for  $i \geq \nu$ , where  $\nu$  is the changepoint. We concentrate on the case where the regression is against time ( $x_i = i$ ) and the change that we are concerned with detecting is a change of slope; that is,  $\gamma_{i,\nu}$  is of a simple form, e.g.  $\theta(i - (\nu - 1))$ . We consider both the case where the baseline parameters  $\alpha$ ,  $\beta$ , and  $\sigma$  are known (in which case, without loss of generality, the data can be transformed to have  $\sigma = 1$  and  $\alpha = \beta = 0$ , also pertinent to detecting a gradual change from a fixed distribution, as described earlier) and the case where they are not. The latter case is also of applied interest; imagine a product, recently introduced, whose share of the market is increasing at a steady rate, and one is hopeful—and on the lookout—for an increase of this rate. Because the product is new, the baseline parameters are not known.

Previous studies have shown that pretending that simple estimates of the unknown baseline parameters are their true values often results in large discrepancies between true and nominal values of operating characteristics (Wheeler 2000). The

---

Abba M. Krieger is Professor, Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: [krieger@wharton.upenn.edu](mailto:krieger@wharton.upenn.edu)). Moshe Pollak is Professor, and Benjamin Yakir is Senior Lecturer, Department of Statistics, Hebrew University, Jerusalem, Israel (E-mail: [mmp@mscc.huji.ac.il](mailto:mmp@mscc.huji.ac.il) and [msby@mscc.huji.ac.il](mailto:msby@mscc.huji.ac.il)). This work was supported by grants from the University of Pennsylvania Research Foundation, Israel Science Foundation, and United States–Israel Binational Science Foundation. The authors thank the referees for their careful reading of the manuscript and their insightful comments.

procedures that we propose circumvent this difficulty. Our approach is likelihood ratio based, and we consider and compare appropriate Cusum and Shiriyayev–Roberts schemes. Although we label our schemes by the names “Cusum” and “Shiriyayev–Roberts,” the actual surveillance statistics used differ considerably from the classical Cusum and Shiriyayev–Roberts statistics, which were developed for the abrupt change-of-distribution problem. They are related only in their theoretical approaches; Cusum is a maximum likelihood approach, and Shiriyayev–Roberts is a quasi-Bayesian approach. The difference is due to the different nature of the two problems; for example, when monitoring for a change in the slope of a regression against time, the post-change observations are not identically distributed (and neither are the pre-change observations if the pre-change slope differs from 0).

There is a large literature on the retrospective problem of detecting a change in regression. This has little to do with the sequential version of the problem (which is the content of this article), for two reasons. For one thing, in the retrospective problem, where one looks for a change in a set of (past) observations, the setting is one of hypothesis testing, and the operating characteristics are probabilities of error, whereas in the sequential context the decision to be made is whether or not to continue sampling, and the operating characteristics are average run lengths. More important, in the retrospective case, the analysis usually assumes that there are many post-change observations, whereas in the sequential (prospective!) context, the goal is to see to it that such observations are few.

The literature for applying sequential methods for detecting a change in regression is scant. No exact optimality results are known for this case. Yao (1993) studied the case where the regressors are bounded and obtained first-order asymptotic optimality results. Yakir, Krieger, and Pollak (1999) considered a case of unbounded regressors (which includes linear regression over time) and obtained first-order asymptotic optimality results for this case. Practical guidelines for constructing and applying such schemes (in finite sample size contexts) have not been developed. It is this gap that we try to narrow in this article. We concentrate on the case where the regressor is time and the errors are normally distributed. We construct and compare a number of procedures. We consider both the case of known baseline parameters and the case of unknown baseline parameters. We supply operating characteristics, cutoff values as functions of the average run length (ARL) to false alarm, as well as a simple inequality for the ARL to false alarm. Using Monte Carlo, we compare a number of schemes whose ARL to false alarm is 750 and suggest one scheme that seems to be the most appropriate for a given set of circumstances.

We apply our results to data on the global warming phenomenon. It is known that global temperatures have been rising during the last century. Is the rate of global warming rising as well? Pretending that we were monitoring on-line during the last half century for an increase in the rate, we apply a detection scheme to the sequence of global yearly average temperatures since World War II. We find that in 1983 a rise in the rate of global warming has occurred, which we estimate (in 1983) to have occurred in 1976. We also find that the rate has not risen again since 1976.

The article is organized as follows. In Section 2 we discuss the known baseline case and present four different types of surveillance schemes for detecting an increase in regression slope. For these procedures, in Section 3 we evaluate (by Monte Carlo) the cutoff value as a function of the ARL to false alarm. We also compare (by Monte Carlo) their maximal expected delay when the ARL to false alarm is 750. In Sections 4 and 5 we report on the analogous construction, evaluation, and comparison for the unknown baseline case. In Section 6 we present the application to the global warming phenomenon. We devote Section 7 to a discussion of methodology. We relegate proofs for formulas appearing in Sections 2 and 4 and discussion of computational issues to the Appendix.

## 2. KNOWN BASELINE PARAMETERS

Without loss of generality, let  $Y_1, Y_2, \dots$  be a sequence of independent random variables, and let  $P_\nu$  and  $E_\nu$  denote probability and expectation when  $Y_1, \dots, Y_{\nu-1} \sim N(0, 1)$  and  $Y_i \sim N(\gamma_{i, \nu}, 1)$  for  $i \geq \nu$ . The value  $\nu = \infty$  means that no change ever takes place. The operating characteristics of a stopping rule  $N$  are taken, as usual, to be the ARL to false alarm,  $E_\infty N$ , and the maximal expected delay to detection,  $\sup_{1 \leq \nu < \infty} E_\nu(N - (\nu - 1) | N \geq \nu)$ .

The basic building block for the procedures that we study is the likelihood ratio of the observations at  $\nu = k$  versus  $\nu = \infty$ , that is

$$\begin{aligned} \Lambda_k^n &= \frac{dP_{\nu=k}(Y_1, \dots, Y_n)}{dP_{\nu=\infty}(Y_1, \dots, Y_n)} \\ &= \exp \left\{ \sum_{i=k}^n (\gamma_{i, k} Y_i - \gamma_{i, k}^2 / 2) \right\}. \end{aligned} \tag{1}$$

In practice, the post-change parameter values,  $\gamma_{i, \nu}$ , are unknown. In the classical changepoint problem, the usual procedure is either to take representative values for the unknown post-change parameters, to take a prior over possible post-change parameter values, or to estimate them.

Procedures that are known to have asymptotic first-order optimality properties are Cusum and Shiriyayev–Roberts (Yao, 1993; Yakir et al., 1999). The Cusum procedure has a stopping rule of the form

$$T_A = \min \left\{ n \mid \max_{1 \leq k \leq n} \Lambda_k^n \geq A \right\}. \tag{2}$$

The Shiriyayev–Roberts procedure has the stopping rule

$$N_A = \min \left\{ n \mid \sum_{k=1}^n \Lambda_k^n \geq A \right\}. \tag{3}$$

To implement these rules, one needs to know the relationship between the cutoff value  $A$  and the ARL to false alarm  $E_\infty T_A$  (or  $E_\infty N_A$ ), to set the cutoff value so as to achieve a prespecified ARL to false alarm.

The Shiriyayev–Roberts rule has here, as elsewhere, the lower bound  $E_\infty N_A \geq A$ . The reason is the usual one:  $\{\Lambda_k^n\}_{n=k}^\infty$  is a  $P_\infty$  martingale with unit mean; hence  $\{\sum_{k=1}^n \Lambda_k^n - n\}_{n=1}^\infty$  is a

$P_\infty$  martingale with mean 0, and so, by the optional sampling theorem,

$$\begin{aligned} 0 &= E_\infty \left( \sum_{k=1}^{N_A} \Lambda_k^{N_A} - N_A \right) \\ &= E_\infty \sum_{k=1}^{N_A} \Lambda_k^{N_A} - E_\infty N_A \geq A - E_\infty N_A. \end{aligned}$$

Hence  $E_\infty N_A \geq A$ . This means that one can construct a conservative procedure that will attain a prespecified ARL to false alarm by setting the cutoff level  $A$  to be equal to the prespecified value. Because  $T_A \geq N_A$ , the Cusum procedure has the same lower bound. However, one gains efficiency by evaluating exactly the relationship between the cutoff level and the ARL to false alarm. We obtain this relationship by Monte Carlo in the next section.

Because of the martingale property of the Shirayayev–Roberts statistic, it is natural to require that any variation of the scheme due to estimation of unknown parameters should preserve this martingale structure. In this vein, we define the following procedures, based on the assumption that the surveillance is being conducted for a change of slope.

### 2.1 Choosing a Representative

Choose  $\theta$  and define  $\gamma_{i,v} = \theta(i - (v - 1))$ . Here

$$\begin{aligned} \Lambda_k^n &= \Lambda_k^n(\theta) \\ &= \exp \left\{ \theta \sum_{i=k}^n (i - k + 1) Y_i - \theta^2 (n - k + 1)(n - k + 2) \right. \\ &\quad \left. \cdot (2n - 2k + 3)/12 \right\}, \end{aligned} \tag{4}$$

and the Cusum and the Shirayayev–Roberts rules are

$$T_A^{(\theta)} = \min \left\{ n \mid \max_{1 \leq k \leq n} \Lambda_k^n(\theta) \geq A \right\} \tag{5}$$

and

$$N_A^{(\theta)} = \min \left\{ n \mid \sum_{k=1}^n \Lambda_k^n(\theta) \geq A \right\}. \tag{6}$$

### 2.2 Choosing a Prior and Creating a Mixture

The putative change is a change of slope, so that  $\gamma_{i,v} = \theta(i - (v - 1))$ . But now we must choose a prior  $F$  and pretend that  $\theta \sim F$ . Here

$$\Lambda_k^n = \Lambda_k^n(F) = \int \Lambda_k^n(\theta) dF(\theta). \tag{7}$$

If the surveillance is geared to detect an increase of slope, then a choice

$$F(\theta) = \frac{\left( \Phi\left(\frac{\theta - \mu}{\tau}\right) - \Phi\left(-\frac{\mu}{\tau}\right) \right)^+}{\Phi\left(\frac{\mu}{\tau}\right)}$$

yields a tractable form for  $\Lambda_k^n$ ,

$$\begin{aligned} \Lambda_k^n(F) &= \frac{\exp\left\{-\frac{1}{2} \frac{\mu^2}{\tau^2}\right\}}{\tau \Phi\left(\frac{\mu}{\tau}\right) \sqrt{\frac{1}{\tau^2} + \frac{(n-k+1)(n-k+2)(2n-2k+3)}{6}}} \\ &\quad \cdot \exp \left\{ \frac{1}{2} \frac{\left( \sum_{i=k}^n (i - k + 1) Y_i + \frac{\mu}{\tau} \right)^2}{\frac{1}{\tau^2} + \frac{(n-k+1)(n-k+2)(2n-2k+3)}{6}} \right\} \\ &\quad \cdot \Phi \left( \frac{\sum_{i=k}^n (i - k + 1) Y_i + \frac{\mu}{\tau}}{\sqrt{\frac{1}{\tau^2} + \frac{(n-k+1)(n-k+2)(2n-2k+3)}{6}}} \right). \end{aligned} \tag{8}$$

If the surveillance is geared to detect a change (either increase or decrease) of slope, then a simple scheme prescribes choosing a representative  $\theta$  and a prior  $F = B_{\pm\theta}$  that realizes  $\theta$  and  $-\theta$  each with probability 1/2. Here

$$\Lambda_k^n(B_{\pm\theta}) = \frac{\Lambda_k^n(\theta) + \Lambda_k^n(-\theta)}{2}. \tag{9}$$

A somewhat broader prior is

$$F(\theta) = \frac{1}{2} \Phi\left(\frac{\theta - \mu}{\tau}\right) + \frac{1}{2} \Phi\left(\frac{\theta + \mu}{\tau}\right),$$

for which

$$\begin{aligned} \Lambda_k^n(F) &= \frac{1}{2} \frac{\exp\left\{-\frac{1}{2} \frac{\mu^2}{\tau^2}\right\}}{\tau \sqrt{\frac{1}{\tau^2} + \frac{(n-k+1)(n-k+2)(2n-2k+3)}{6}}} \\ &\quad \cdot \left[ \exp \left\{ \frac{1}{2} \frac{\left( \sum_{i=k}^n (i - k + 1) Y_i + \frac{\mu}{\tau} \right)^2}{\frac{1}{\tau^2} + \frac{(n-k+1)(n-k+2)(2n-2k+3)}{6}} \right\} \right. \\ &\quad \left. + \exp \left\{ \frac{1}{2} \frac{\left( \sum_{i=k}^n (i - k + 1) Y_i - \frac{\mu}{\tau} \right)^2}{\frac{1}{\tau^2} + \frac{(n-k+1)(n-k+2)(2n-2k+3)}{6}} \right\} \right], \end{aligned} \tag{10}$$

which simplifies somewhat if  $\mu$  is chosen to be 0. The Cusum and Shirayayev–Roberts rules are

$$T_A^F = \min \left\{ n \mid \max_{1 \leq k \leq n} \Lambda_k^n(F) \geq A \right\} \tag{11}$$

and

$$N_A^F = \min \left\{ n \mid \sum_{k=1}^n \Lambda_k^n(F) \geq A \right\}. \tag{12}$$

The results of this article can be extended to nonnormal priors. Normal priors are chosen in part for their computational tractability, with the realization that the literature suggests (Pollak, 1987) that the results in this context are not sensitive to the exact shape of the prior.

### 2.3 Choosing an Estimator

Again, the putative post-change parameter is  $\gamma_{i,v} = \theta(i - (v - 1))$ , but here  $\theta$  is to be estimated. To preserve the martingale structure of the Shirayayev–Roberts statistic, we follow Lorden and Pollak (2003) and use an asymptotically efficient estimator, the maximum likelihood estimator (MLE), based on the first  $n - 1$  observations for the likelihood of  $Y_n$ . (This is not merely a theoretical artifact; adding the  $n$ th observation into the

estimation of parameters may hurt the ARL to false alarm, necessitating a higher cutoff level, which would obviate any gain in detection time due to the slightly better estimate of the unknown parameters.) Here we use

$$\hat{\theta}_k^n = \frac{6 \sum_{i=k}^{n-1} (i-k+1) Y_i}{(n-k)(n-k+1)(2n-2k+1)} \quad (13)$$

if the surveillance is geared to detect a change, and  $(\hat{\theta}_k^n)^+$  if it is geared to detect an increase. So

$$\Lambda_k^n(\hat{\theta}^+) = \exp \left\{ \sum_{i=k}^n [(\hat{\theta}_k^i)^+ (i-k+1) Y_i - ((\hat{\theta}_k^i)^+)^2 (i-k+1)^2 / 2] \right\}, \quad (14)$$

$$\Lambda_k^n(\hat{\theta}) = \exp \left\{ \sum_{i=k}^n [(\hat{\theta}_k^i) (i-k+1) Y_i - ((\hat{\theta}_k^i))^2 (i-k+1)^2 / 2] \right\},$$

and

$$\begin{aligned} T_A(\hat{\theta}^+) &= \min \left\{ n \mid \max_{1 \leq k \leq n} \Lambda_k^n(\hat{\theta}^+) \geq A \right\}, \\ T_A(\hat{\theta}) &= \min \left\{ n \mid \max_{1 \leq k \leq n} \Lambda_k^n(\hat{\theta}) \geq A \right\}, \\ N_A(\hat{\theta}^+) &= \min \left\{ n \mid \sum_{k=1}^n \Lambda_k^n(\hat{\theta}^+) \geq A \right\}, \\ N_A(\hat{\theta}) &= \min \left\{ n \mid \sum_{k=1}^n \Lambda_k^n(\hat{\theta}) \geq A \right\}. \end{aligned} \quad (15)$$

### 2.4 A Semiparametric Estimator

The post-change regression may be nonlinear. The aforementioned methods can be applied to more complicated putative post-change models. For example, if

$$\gamma_{i,v} = \theta_1 + \theta_2(i - (v-1)) + \theta_3(i - (v-1))^2,$$

then one can choose representative values for  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$ ; a prior for  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$ ; or an estimator for  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$ . Following Yakir et al. (1999), here, rather than consider a nonlinear parametric form for  $\gamma_{i,k}$ , we instead consider the case where  $\gamma_{i,k}$  is estimated nonparametrically by  $Y_{i-1}$ . (Although  $Y_{i-1}$  is an underestimate of  $\gamma_{i,k}$ —if the change is an increase in mean—it should be reasonably close no matter what the pre-change mean function.) Then the post-change mean ( $\alpha + \beta x_i + \gamma_{i,k}$ ) is semiparametric, and thus

$$\begin{aligned} \Lambda_k^n(\text{semiparametric}^+) &= \exp \left\{ \sum_{i=k}^n (Y_i Y_{i-1}^+ - (Y_{i-1}^+)^2 / 2) \right\}, \\ \Lambda_k^n(\text{semiparametric}) &= \exp \left\{ \sum_{i=k}^n (Y_i Y_{i-1} - Y_{i-1}^2 / 2) \right\}, \end{aligned} \quad (16)$$

and

$$\begin{aligned} T_A^{SP+} &= \min \left\{ n \mid \max_{1 \leq k \leq n} \Lambda_k^n(\text{semiparametric}^+) \geq A \right\}, \\ T_A^{SP} &= \min \left\{ n \mid \max_{1 \leq k \leq n} \Lambda_k^n(\text{semiparametric}) \geq A \right\}, \\ N_A^{SP+} &= \min \left\{ n \mid \sum_{k=1}^n \Lambda_k^n(\text{semiparametric}^+) \geq A \right\}, \\ N_A^{SP} &= \min \left\{ n \mid \sum_{k=1}^n \Lambda_k^n(\text{semiparametric}) \geq A \right\}, \end{aligned} \quad (17)$$

are the procedures for detecting an increase in slope and detecting a change.

### 2.5 Pretending that the Post-change Mean is Constant

The main thrust of this article is the case where baseline parameters are unknown, which will be dealt with in the next section. Part of the rationale for studying the procedures described earlier in the known baseline case is to get a picture of the regret due to the ignorance of baseline parameter values when they are unknown. However, the case of known baseline parameters is also of interest in its own right. In this context, it is natural to question whether the procedures outlined earlier have an advantage over pretending that the problem is detecting an abrupt change of the mean of the pre-change residuals to a constant post-change mean. The answer to this question seems to depend on the ARL to false alarm, as will be described presently. (This question is not relevant in the unknown baseline case; see remark 4 in Sec. 7.)

Without loss of generality, consider the set of observation to be the sequence of residuals (from the pre-change regression), so that again they are iid  $N(0, 1)$  random variables (before to the change). We now pretend that post-change the mean increases to a value  $\mu > 0$ . If  $\mu$  is unknown, then the likelihood ratio for  $\nu = k$  is

$$\frac{f_{\nu=k}(X_1, X_2, \dots, X_n)}{f_{\nu=\infty}(X_1, X_2, \dots, X_n)} = \exp \left\{ \mu \sum_{i=k}^n X_i - (n-k+1)\mu^2 / 2 \right\}.$$

[The classical Cusum raises an alarm at the first time  $n$  that the maximum (over  $1 \leq k \leq n$ ) of these likelihood ratios exceeds a prespecified level.] When  $\mu$  is unknown, various methods have been proposed, in a vein similar to the method developed earlier for the change-of-slope-of-regression detection problem. One method that has been gaining attention recently is the generalized likelihood ratio (GLR) Cusum. When monitoring for a change in mean, the GLR is obtained by estimating  $\mu$  via

$$\mu_k^n = \frac{1}{n-k+1} \sum_{i=k}^n X_i$$

when monitoring for a change in mean

or

$$(\mu_k^n)^+ \quad \text{when monitoring for an increase in mean,}$$

and inserting this in the formula for the likelihood ratio. The GLR Cusum raises an alarm at the first time that the maximum

(over  $1 \leq k \leq n$ ) of these (pseudo) likelihood ratios exceeds a prespecified level, that is, at

$$T_A^{GLR} = \min \left\{ n \mid \max_{1 \leq k \leq n} \exp \left\{ (n-k+1) (\mu_k^n)^2 / 2 \right\} \geq A \right\}$$

or

$$T_A^{GLR^+} = \min \left\{ n \mid \max_{1 \leq k \leq n} \exp \left\{ (n-k+1) ((\mu_k^n)^+)^2 / 2 \right\} \geq A \right\}.$$

### 3. MONTE CARLO-KNOWN BASELINE PARAMETERS

We conducted a Monte Carlo study of detecting an increase in slope. From the methods described in Sections 2.1–2.4, we chose six procedures. For each of these procedures, we constructed a Cusum and a Shirayev–Roberts scheme. We selected  $\theta = .05, \theta = .10$ , and  $\theta = .20$  as three representative values for the post-change slope, believing these to be values most often encountered in practice. We chose  $\mu = .10$  and  $\tau = .05$  for a mixture

$$F(x) = \left[ \Phi \left( \frac{x-\mu}{\tau} \right) - \Phi \left( -\frac{\mu}{\tau} \right) \right] / \Phi \left( \frac{\mu}{\tau} \right),$$

the rationale being that this prior captures the practical range of values without giving weight to extremes. As an estimation procedure, we simulated the one described in Section 2.3 and simulated the semiparametric scheme described in Section 2.4. In addition, we ran a simulation of the GLR Cusum procedure described in Section 2.5.

We ran 62,500 repetitions for each method when there is no change in slope [ $P_\infty$ ; all observations are  $N(0, 1)$ ] to evaluate the cutoff value  $A$  required to obtain a prespecified ARL to false alarm,  $E_\infty N$ . The prespecified values of  $E_\infty N$  were in the range (100, 1,500). The plots of  $A$  versus  $E_\infty T_A$  and of  $A$  versus  $E_\infty N_A$  were almost perfect straight lines (correlations  $> .999$ ); the results are described in Table 1.

Subsequently, we ran each of the procedures when a change is in effect from the very beginning ( $\nu = 1$ ). [The reason for choosing  $\nu = 1$  is that this is the value of  $\nu$  that maximizes both  $E(T_A - (\nu - 1) \mid T_A \geq \nu)$  and  $E(N_A - (\nu - 1) \mid N_A \geq \nu)$ .] We ran each procedure with a number of different post-change mean functions, some linear, some having only a quadratic term, and some having both. Each combination was run 10,000 times. All of the procedures are engineered to have ARL to false alarm equal to 750. The cutoff values were obtained from Table 1. The means and standard deviations of the 10,000 runs are reported in Table 2.

Table 2 indicates that if one can assume that the regression with respect to time will be linear, then among the methods described in Sections 2.1–2.4, the Cusum mixture rule seems to be the best choice, unless one is concerned about detecting a very small change of slope. If quadratic terms are also considered to be possible, then the Cusum estimation procedure seems preferable. The semiparametric rule does surprisingly well, despite its simplicity. Another surprise is that even if the true post-change slope is  $\theta$ , the method that uses the true  $\theta$  as the representative value is not always the best.

Obviously, at ARL to false alarm equal to 750, the GLR Cusum does (somewhat) better than the other procedures outlined in this section, as is clear from the simulations reported in Table 2. This phenomenon, though surprising, has been noticed in a similar context, that of detecting a gradual change from one mean to another (Sigal, 1998). However, for larger ARLs to false alarm, the picture changes. Asymptotically, when the ARL to false alarm is large (i.e., the cutoff level is large), it takes the GLR Cusum more time on average to detect a change of regression slope from 0 to  $\theta$  than the other procedures outlined in this section. To see this, suppose that  $\nu = 1$ . In this case, note that  $\max_{1 \leq k \leq n} (\mu_k^n)^2 / 2 \approx (\mu_1^n)^2 / 2 \approx \theta^2 n^2 / 8$ , and it can be shown that the average delay to detection is of the order  $\log^{1/2}$ (ARL to false alarm). A similar analysis for the other procedures outlined in this section (the estimated linear procedure is the most transparent) yields an average delay to detection of order  $\log^{1/3}$ (ARL to false alarm).

### 4. UNKNOWN BASELINE PARAMETERS

Again, we assume that  $Y_1, Y_2, \dots$  are independent, but now

$$Y_i \sim N((\alpha + \beta x_i)\sigma, \sigma^2) \quad \text{for } i = 1, \dots, \nu - 1, \tag{18}$$

$$Y_i \sim N((\alpha + \beta x_i + \gamma_{i,\nu})\sigma, \sigma^2) \quad \text{for } i = \nu, \nu + 1, \dots,$$

where the regressors  $x_i$  are known ( $x_i = i$  for regression against time) and none of the parameters  $\alpha, \beta$ , or  $\sigma$  is known. We assume that  $\nu \geq 4$ , so that  $Y_1, Y_2$ , and  $Y_3$  are pre-change observations, and that (by design)  $x_1 \neq x_2$ . The methods described below are based on a reduction by invariance. Define the recursive residuals

$$Z_i = \sqrt{\frac{i-1}{i}} (Y_i - \bar{Y}_{i-1}), \quad i = 2, 3, \dots \tag{19}$$

These are independent, normally distributed random variables (both before and after a change) with variance  $\sigma^2$ , and their

Table 1. Cutoff Values  $A$  as a Function of ARL to False Alarm of Various Procedures for Detecting an Increase of Slope

Method	Cusum	Shirayev–Roberts
Representative $\theta = .05$	$A = .59231 + .04958 \cdot E_\infty T_A$	$A = 19.14466 + .55526 \cdot E_\infty N_A$
Representative $\theta = .10$	$A = .80269 + .06696 \cdot E_\infty T_A$	$A = 17.87843 + .46122 \cdot E_\infty N_A$
Representative $\theta = .20$	$A = 1.37147 + .08563 \cdot E_\infty T_A$	$A = 14.65771 + .37074 \cdot E_\infty N_A$
Mixture $F(x)$	$A = 1.57793 + .04849 \cdot E_\infty T_A$	$A = 19.29980 + .47550 \cdot E_\infty N_A$
$= \frac{\Phi(\frac{x-10}{.05}) - \Phi(-2)}{\Phi(2)}$		
Estimation $(\hat{\theta}_k^n)^+$	$A = 1.30997 + .05870 \cdot E_\infty T_A$	$A = 19.87650 + .43830 \cdot E_\infty N_A$
$= \frac{(\sum_{j=k}^{n-1} (j-k+1) Y_j)^+}{\frac{1}{8}(n-k)(n-k+1)(2n-2k+1)}$		
Semiparametric	$A = 3.18234 + .03592 \cdot E_\infty T_A$	$A = 22.47070 + .30950 \cdot E_\infty N_A$
GLR	$A = -19.44368 + .458654 \cdot E_\infty T_A$	

Table 2. Monte Carlo Estimates of Maximal Average Delay to Detection,  $E_{\nu=1}N$ , Based on 10,000 Repetitions, Where the ARL to False Alarm,  $E_{\infty}N = 750$ , for Various Methods and True Post-change Regressions

Post-change regression		Representative													
Linear slope	Quadratic coefficient	$\theta = .05$		$\theta = .10$		$\theta = .20$		Mixture <sup>a</sup>		Estimator <sup>b</sup>		Semiparametric		GLR	
.01	.00	52.66	17.0	55.37	19.0	59.65	22.0	55.75	19.0	57.65	19.0	65.54	24.0	55.90	17.9
		52.92	17.0	55.16	19.0	59.09	21.0	53.86	17.0	59.39	17.0	66.62	24.0		
.05	.00	20.09	4.0	19.91	5.0	20.54	6.0	19.98	5.0	21.17	6.0	22.25	7.0	20.68	5.5
		21.31	4.0	20.22	5.0	20.53	6.0	20.54	5.0	22.48	5.0	23.18	6.0		
.10	.00	14.32	2.0	13.13	3.0	12.99	3.0	12.97	3.0	13.60	3.0	14.10	4.0	13.46	3.3
		15.81	2.0	13.80	2.0	13.13	3.0	13.87	3.0	14.65	3.0	14.86	4.0		
.20	.00	10.81	1.0	9.29	1.0	8.53	2.0	8.85	1.0	8.77	2.0	8.89	2.0	8.75	2.0
		12.21	.9	10.12	1.1	8.87	2.0	9.79	1.0	9.57	2.0	9.47	2.0		
.00	.05	8.92	.50	7.87	.70	7.17	.90	7.47	.8	7.10	1.0	7.15	1.1	7.16	1.1
		9.83	.50	8.42	.60	7.43	.70	8.05	.7	7.47	.9	7.42	1.0		
.00	.10	7.42	.50	6.49	.50	5.76	.60	6.07	.5	5.47	.8	5.51	.8	5.47	0.8
		8.15	.40	7.00	.30	6.05	.50	6.60	.5	5.78	.7	5.73	.8		
.00	.20	6.08	.30	5.34	.50	4.77	.40	5.00	.3	4.23	.6	4.26	.6	4.20	0.6
		7.00	.05	5.97	.20	5.00	.30	5.43	.5	4.49	.6	4.45	.6		
.10	.05	8.28	.50	7.19	.60	6.42	.80	6.74	.7	6.23	1.0	6.27	1.1	6.24	1.0
		9.18	.40	7.79	.50	6.73	.70	7.35	.6	6.62	.9	6.55	1.0		
.10	.10	7.07	.30	6.14	.40	5.40	.50	5.72	.5	5.01	.8	5.04	.8	5.02	0.8
		7.98	.20	6.75	.40	5.73	.50	6.22	.5	5.34	.7	5.28	.7		
.10	.20	6.00	.10	5.14	.30	4.58	.50	4.90	.3	4.00	.6	4.05	.6	3.96	0.6
		6.98	.20	5.86	.30	4.91	.30	5.17	.4	4.26	.6	4.24	.6		
1.00	.20	5.00	.02	4.05	.20	3.62	.50	3.97	.2	2.73	.5	2.77	.5	2.61	0.5
		5.99	.09	4.94	.20	3.98	.10	4.07	.3	2.95	.4	2.95	.4		
1.00	1.00	4.00	.00	3.00	.06	3.00	.02	3.00	.0	2.07	.3	2.08	.3	1.92	0.3
		4.02	.10	3.99	.10	3.00	.00	3.01	.1	2.11	.3	2.14	.4		

NOTE: The four numbers in each cell refer to mean (Cusum) s.d. (Cusum) mean (Shiryayev–Roberts) s.d. (Shiryayev–Roberts).

<sup>a</sup>Mixture  $(N(.10, .05^2) | N(.10, .05^2) > 0)$ .

<sup>b</sup>Estimation  $\hat{\theta}_A^{j+} = \frac{(\sum_{i=k}^{n-1} (i-k+1) Y_i)^+}{(n-k)(n-k+1)(2n-2k+1)/6}$ .

means do not involve  $\alpha$ . Next, note that

$$Z_i - Z_2 \sqrt{2} \sqrt{\frac{i-1}{i} \frac{x_i - \bar{x}_{i-1}}{x_2 - x_1}}$$

are (dependent) jointly normally distributed random variables whose means and standard deviations are multiples of  $\sigma$  and do not involve  $\alpha$  and  $\beta$ , and whose correlations are independent of  $\alpha$ ,  $\beta$ , and  $\sigma$ . Hence

$$W_i = \frac{Z_i - Z_2 \sqrt{2} \sqrt{\frac{i-1}{i} \frac{x_i - \bar{x}_{i-1}}{x_2 - x_1}}}{|Z_3 - Z_2 \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1}|}, \quad i = 4, 5, \dots \quad (20)$$

is a sequence of (dependent) random variables whose (joint) distribution does not depend on  $\alpha$ ,  $\beta$ , and  $\sigma$  and whose distribution when  $\nu = \infty$  is well defined. The same applies to

$$W_i^* = \frac{Z_i - Z_2 \sqrt{2} \sqrt{\frac{i-1}{i} \frac{x_i - \bar{x}_{i-1}}{x_2 - x_1}}}{Z_3 - Z_2 \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1}}, \quad i = 4, 5, \dots \quad (21)$$

When the surveillance is being conducted for an increase in slope, it is appropriate to base the procedure on the sequence  $\{W_n\}_{n=4}^{\infty}$ , and when monitoring for a change of slope (in any direction), a procedure based on  $\{W_n^*\}_{n=4}^{\infty}$  is appropriate. (It may seem that the denominator of  $W_i$  is a poor surrogate for the unknown standard deviation; however, there

is a 1-to-1 correspondence between  $Y_1, Y_2, Y_3, Y_4, \dots, Y_n$  and  $Y_1, Y_2, Y_3, W_4, \dots, W_n$ . Therefore, by basing surveillance on  $W_4, \dots, W_n$ , we are losing only the information contained in the three variables  $Y_1, Y_2$ , and  $Y_3$ .) As in the previous section, the main building blocks will be the likelihood ratios of the sequence being monitored (here  $\{W_n\}$  or  $\{W_n^*\}$ ).

Calculation of the likelihood ratios is formidable, and we relegate it to the Appendix. In this section we consider the special case where regression is against time— $x_i = i$ —and one is on the alert for an increase or a change in slope. (The formulas in the Appendix are implemented here after substituting  $i$  for  $x_i$ .) We develop analogs to the taking of a representative for the unknown difference in slope caused by the change. We also present mixture rules, although there is less technical flexibility in their construction than in the known baseline case. Because of technical difficulties in implementing the estimation approach in the unknown baseline case, we dwell on it only briefly here. Finally, we present an analog of the semiparametric procedure of Section 2.

#### 4.1 Choosing a Representative

As in Section 2.1, choose a representative value for  $\theta$ . Define

$$\mu_{i,k} = \theta \sqrt{\frac{i-1}{i}} \left( i - k + 1 - \frac{(i-k)(i-k+1)}{2(i-1)} \right)^+$$

$$= \theta \sqrt{\frac{i-1}{i}} ((i-k+1)^+)^{\frac{i+k-2}{2(i-1)}} \quad \text{for } i, k \geq 4,$$

$$d_n = \frac{6}{(n-1)n(n+1)},$$

$$a_n = 1 + \sum_{i=4}^n W_i^2 - d_n \left( \sum_{i=4}^n \sqrt{\frac{i(i-1)}{2}} W_i + \sqrt{3} \right),$$

$$a_n^* = 1 + \sum_{i=4}^n (W_i^*)^2 - d_n \left( \sum_{i=4}^n \sqrt{\frac{i(i-1)}{2}} W_i^* + \sqrt{3} \right),$$

$$b_{n,k} = \sum_{i=4}^n W_i \mu_{i,k} - d_n \left( \sum_{i=4}^n \sqrt{\frac{i(i-1)}{2}} W_i + \sqrt{3} \right)$$

$$\cdot \left( \sum_{i=4}^n \sqrt{\frac{i(i-1)}{2}} \mu_{i,k} \right),$$

$$b_{n,k}^* = \sum_{i=4}^n W_i^* \mu_{i,k} - d_n \left( \sum_{i=4}^n \sqrt{\frac{i(i-1)}{2}} W_i^* + \sqrt{3} \right)$$

$$\cdot \left( \sum_{i=4}^n \sqrt{\frac{i(i-1)}{2}} \mu_{i,k} \right),$$

$$c_{n,k} = d_n \left( \sum_{i=4}^n \sqrt{\frac{(i-1)i}{4}} \mu_{i,k} \right)^2 - \frac{1}{2} \sum_{i=4}^n \mu_{i,k}^2,$$

$$g_m(x) = \int_0^\infty z^m e^{-(z-x)^2/2} dz,$$

$$g_m^*(x) = \int_{-\infty}^\infty |z|^m e^{-(z-x)^2/2} dz.$$

(Calculation of  $g_m$  and  $g_m^*$  can be done with a recursion formula; see Remark 3 in Sec. 7.) Then

$$\begin{aligned} \Lambda_k^n &= \frac{dP_k(W_4, \dots, W_n)}{dP_\infty(W_4, \dots, W_n)} \\ &= \frac{g_{n-3} \left( \frac{b_{n,k}}{\sqrt{a_n}} \right)}{g_{n-3}(0)} \cdot \exp\{b_{n,k}/\sqrt{a_n}\} \cdot \exp\{c_{n,k}\}, \end{aligned} \tag{23}$$

$$\begin{aligned} \Lambda_k^{n*} &= \frac{dP_k(W_4^*, \dots, W_n^*)}{dP_\infty(W_4^*, \dots, W_n^*)} \\ &= \frac{g_{n-3}^* \left( \frac{b_{n,k}^*}{\sqrt{a_n^*}} \right)}{g_{n-3}^*(0)} \cdot \exp\{b_{n,k}^*/\sqrt{a_n^*}\} \cdot \exp\{c_{n,k}\}. \end{aligned}$$

To monitor for an increase in slope, the Cusum and Shirayayev–Roberts rules are

$$T_A = \min \left\{ n \mid \max_{4 \leq k \leq n} \Lambda_k^n \geq A \right\} \quad \text{and}$$

$$N_A = \min \left\{ n \mid 3 + \sum_{k=4}^n \Lambda_k^n \geq A \right\}. \tag{24}$$

To monitor for a change in slope, the Cusum and Shirayayev–Roberts rules are

$$T_A^* = \min \left\{ n \mid \max_{4 \leq k \leq n} \Lambda_k^{n*} \geq A \right\} \quad \text{and}$$

$$N_A^* = \min \left\{ n \mid 3 + \sum_{k=4}^n \Lambda_k^{n*} \geq A \right\}. \tag{25}$$

(The number 3 is added to the sum of the likelihood ratios in the Shirayayev–Roberts statistic for aesthetic reasons only, to return it to its usual form where its expectation under  $P_\infty$  after  $n$  observations is  $n$ .)

### 4.2 Choosing a Prior and Creating a Mixture

Denote

$$q_{n,k} = \frac{b_{n,k}/\sqrt{a_n}}{\theta}, \quad s_{n,k} = -\frac{c_{n,k}}{\theta}.$$

(Note that  $q_{n,k}$  and  $s_{n,k}$  do not depend on  $\theta$ .) Thus

$$\begin{aligned} \Lambda_k^n(\theta) &= \frac{dP_k(W_4, \dots, W_n)}{dP_\infty(W_4, \dots, W_n)} \\ &= \frac{g_{n-3}(\theta q_{n,k})}{g_{n-3}(0)} \exp\{\theta^2(q_{n,k}^2 + s_{n,k})\}. \end{aligned} \tag{26}$$

Let  $F(\theta) = \Phi((\theta - \mu)/\tau)$ . We get

$$\begin{aligned} \Lambda_k^n(F) &= \int_{-\infty}^\infty \Lambda_k^n(\theta) \left[ \frac{\phi((\theta - \mu)/\tau)}{\tau} \right] d\theta \\ &= \frac{g_{n-3} \left( -\frac{q_{n,k}\mu/\tau^2}{\sqrt{(q_{n,k}^2 + s_{n,k} + \frac{1}{\tau^2})(s_{n,k} + \frac{1}{\tau^2})}} \right)}{g_{n-3}(0)} \\ &\quad \cdot \frac{(q_{n,k}^2 + s_{n,k} + \frac{1}{\tau^2})^{\frac{m}{2}}}{\tau (s_{n,k} + \frac{1}{\tau^2})^{\frac{m+1}{2}}} \\ &\quad \cdot \exp \left\{ \frac{1}{2} \frac{\mu^2}{\tau^4} \frac{q_{n,k}^2 + s_{n,k} + \frac{1}{\tau^2}}{(q_{n,k}^2 + s_{n,k} + \frac{1}{\tau^2})(s_{n,k} + \frac{1}{\tau^2})} - \frac{1}{2} \frac{\mu^2}{\tau^2} \right\}. \end{aligned} \tag{27}$$

For the  $\{W_n^*\}$  sequence,

$$\begin{aligned} \Lambda_k^{n*}(F) &\stackrel{\text{def}}{=} \int \Lambda_k^{n*}(\theta) \left[ \frac{\phi(\frac{\theta - \mu}{\tau})}{\tau} \right] d\theta \\ &= \int \left[ \frac{dP_k(W_4^*, \dots, W_n^*)}{dP_\infty(W_4^*, \dots, W_n^*)} \right] \left[ \phi\left(\frac{\theta - \mu}{\tau}\right) \right] d\theta \end{aligned}$$

will be the same as expression (27), with  $W_n$ ,  $q_{n,k}$ , and  $s_{n,k}$  replaced by  $W_n^*$ ,

$$q_{n,k}^* = \frac{b_{n,k}^*/\sqrt{a_n^*}}{\theta}, \quad \text{and} \quad s_{n,k}^* = -\frac{c_{n,k}^*}{\theta}.$$

For detecting an increase in slope, one would want a prior to be concentrated on  $(0, \infty)$ . Unfortunately, there does not seem to be such a prior that is computationally tractable. A possible compromise is to take a  $N(\mu, \tau^2)$  prior that gives most of its mass to positive values, such as the  $N(.10, .05^2)$  prior considered earlier for the known baseline case, without conditioning on its being positive. As usual, the Cusum stopping rule is

$$T_A(F) = \min \left\{ n \mid \max_{1 \leq k \leq n} \Lambda_k^n(F) \geq A \right\},$$

and the Shiriyayev–Roberts stopping rule is

$$N_A(F) = \min \left\{ n \mid \sum_{k=1}^n \Lambda_k^n(F) \geq A \right\}.$$

For detecting a change in slope, because of the symmetry of  $\Lambda_k^{n*}(\theta)$ , the prior  $N(\mu, \tau^2)$  is equivalent to a mixture of  $N(\mu, \tau^2)$  and  $N(-\mu, \tau^2)$  with equal probability, and thus is a reasonable prior. The Cusum stopping rule is

$$T_A^*(F) = \min \left\{ n \mid \max_{1 \leq k \leq n} \Lambda_k^{n*}(F) \geq A \right\},$$

and the Shiriyayev–Roberts stopping rule is

$$N_A^*(F) = \min \left\{ n \mid \sum_{k=1}^n \Lambda_k^{n*}(F) \geq A \right\}.$$

Taking  $\mu = 0$  will simplify the calculation of  $\Lambda_k^{n*}(F)$  tremendously (because the  $g_{n-3}$  terms cancel), but a  $N(0, \tau^2)$  prior gives a relatively high probability to parameter values near 0, which is not desirable unless detecting very small changes in slope is important.

### 4.3 An Estimation Approach

In a manner similar to the known baseline case, one needs to estimate  $\hat{\theta}_{i,k}$  to substitute for  $\theta$  (in  $\mu_{i,k}$ ) in the expression of the likelihood ratio. The estimate  $\hat{\theta}_{i,k}$  should be a function of  $W_4, \dots, W_{i-1}$  (or  $W_4^*, \dots, W_{i-1}^*$ ). A MLE requires considerable numerical integration, and, considering the fact that each  $\Lambda_k^n$  entails  $n - k$  such calculations, the resulting algorithm will be very unwieldy, rendering the approach impractical even with present-day state-of-the-art computation. (Although when the baseline is known, the estimation approach has a slight advantage over the other methods, the inefficiency of a non-MLE will offset any gain.)

### 4.4 A Semiparametric Procedure

For an analog of the semiparametric procedure of Section 2, define for,  $n \geq 4$ ,

$$\begin{aligned} \hat{\beta}_n &= \frac{\sum_{i=1}^n iY_i - \frac{n+1}{2} \sum_{i=1}^n Y_i}{n(n^2 - 1)/12}, \\ \hat{\alpha}_n &= \bar{Y}_n - \frac{\hat{\beta}_n(n+1)}{2}, \\ \hat{\sigma}_n^2 &= \frac{\sum_{i=1}^n (Y_i - \hat{\alpha}_n - \hat{\beta}_n i)^2}{n-2}, \\ \hat{Y}_{i,n} &= \frac{Y_i - \hat{\alpha}_n - \hat{\beta}_n i}{\hat{\sigma}_n}, \\ \hat{\Lambda}_k^n &= \exp \left\{ \sum_{i=k}^n \left( \hat{Y}_{i-1,n}^+ \hat{Y}_{i,n} - (\hat{Y}_{i-1,n}^+)^2 / 2 \right) \right\}, \\ \hat{\Lambda}_k^{n*} &= \exp \left\{ \sum_{i=k}^n \left( \hat{Y}_{i-1,n} \hat{Y}_{i,n} - (\hat{Y}_{i-1,n})^2 / 2 \right) \right\}. \end{aligned} \tag{28}$$

The Cusum and the Shiriyayev–Roberts rules are

$$\hat{T}_A = \min \left\{ n \mid n \geq 4, \max_{4 \leq k \leq n} \hat{\Lambda}_k^n \geq A \right\}$$

and

$$\hat{N}_A = \min \left\{ n \mid n \geq 4, 3 + \sum_{k=4}^n \hat{\Lambda}_k^n \geq A \right\}, \tag{29}$$

for detecting an increase in slope, and are

$$\hat{T}_A^* = \min \left\{ n \mid n \geq 4, \max_{4 \leq k \leq n} \hat{\Lambda}_k^{n*} \geq A \right\}$$

and

$$\hat{N}_A^* = \min \left\{ n \mid n \geq 4, 3 + \sum_{k=4}^n \hat{\Lambda}_k^{n*} \geq A \right\} \tag{30}$$

for detecting a change in slope.

## 5. MONTE CARLO–UNKNOWN BASELINE PARAMETERS

We studied four procedures, each by Cusum and Shiriyayev–Roberts, all for detecting an increase in slope. In the known baseline case,  $E_\nu(T_A - (\nu - 1) \mid T_A \geq \nu)$  and  $E_\nu(N_A - (\nu - 1) \mid N_A \geq \nu)$  have a maximum at  $\nu = 1$  (and for this reason we regarded only  $\nu = 1$  in Sec. 3). However, in the unknown baseline case, if the post-change regression is only a change of slope (with no quadratic term or higher), then  $\nu = 1$  is indistinguishable from  $\nu = \infty$ . Therefore, rather than consider  $\sup_{1 \leq \nu < \infty} E_\nu(T_A - (\nu - 1) \mid T_A \geq \nu)$  and  $\sup_{1 \leq \nu < \infty} E_\nu(N_A - (\nu - 1) \mid N_A \geq \nu)$ , we ran the procedures for various values of  $\nu$ .

Heuristically, a method that does not take a representative for a change in slope intrinsically estimates it. When the baseline is unknown, such an estimate is very noisy, because it is a difference of the two estimates of the two slopes (before and after change), unless  $\nu$  is large. The noisiness of the estimate will make a change of slope less discernible. Therefore, it is to be expected that for early changepoints (small  $\nu$ ), the mixture, estimation, and semiparametric procedures will do poorly relative to those relying on a representative value. Because of this, we studied only the three representative procedures ( $\theta = .05, .10, .20$ ). We added the semiparametric procedure to show the effect of the baseline being unknown.

We ran 10,000 repetitions for each method when there is no change in slope [ $P_\infty$ ; with no loss of generality, all observations are  $N(0, 1)$ ] to evaluate the cutoff value  $A$  required to obtain a prespecified ARL to false alarm  $E_\infty N$ . The prespecified values of  $E_\infty N$  were in the range (100, 1,000). As in the known baseline case, the plots of  $A$  versus  $E_\infty T_A$  and of  $A$  versus  $E_\infty N_A$  were almost perfect straight lines (correlations  $> .999$ ). The results are described in Table 3.

Subsequently, we ran each of the procedures when the changepoint is  $\nu$ , for various values of  $\nu$ . (The specific values were chosen because they give a picture of the speed of detection of an early change where the ARL to detection is large, and they also illustrate how long it takes for the procedure to act as if the baseline were known.) We ran each procedure with a number of different post-change slope differences, and ran each combination 1,000 times. All of the procedures had ARL to false alarm equal to 750. The cutoff values were obtained from Table 3. The means, the standard deviations, and number of times the procedure did not stop before  $\nu$  are recorded in Table 4.

Judging by Table 4, it seems that the Shiriyayev–Roberts procedure is generally preferable to Cusum, and that the representative  $\theta = .10$  is slightly preferable to the representative  $\theta = .20$

Table 3. Cutoff Values A as a Function of ARL to False Alarm of a Few Procedures for Detecting an Increase of Slope

Method	Cusum	Shiryayev-Roberts
Representative $\theta = .05$	$A = -1.7043 + .04580 \cdot E_{\infty} N$	$A = 34.1262 + .55189 \cdot E_{\infty} N$
Representative $\theta = .10$	$A = -1.7289 + .06254 \cdot E_{\infty} N$	$A = 13.2372 + .47620 \cdot E_{\infty} N$
Representative $\theta = .20$	$A = -2.0601 + .08366 \cdot E_{\infty} N$	$A = 10.5111 + .38115 \cdot E_{\infty} N$
Semiparametric	$A = -.3179 + .03155 \cdot E_{\infty} T_A$	$A = 1.7865 + .27020 \cdot E_{\infty} N_A$

Table 4. Monte Carlo Estimates of Maximal Average Delay to Detection,  $E_{\nu}(N - \nu + 1 | N \geq \nu)$ , Based on 1,000 Repetitions, Where the ARL to False Alarm  $E_{\infty} N = 750$  for the Representative

Change	Slope	Representative										Semiparametric	
		$\theta = .05$		$\theta = .10$		$\theta = .20$							
11	.01	722.606	628.11	1,000	678.925	634.19	1,000	712.750	681.403	1,000	736.641	580	(1,000)
		708.218	674.59	1,000	680.848	689.01	1,000	698.242	676.17	1,000	743.919	596	(1,000)
	.05	607.060	586.59	1,000	611.828	643.27	1,000	678.398	687.91	1,000	714.463	612	(1,000)
		575.440	644.43	1,000	580.482	691.92	1,000	641.452	709.14	1,000	711.234	594	(1,000)
	.10	482.249	538.82	1,000	469.219	561.72	1,000	559.743	672.33	1,000	664.079	559	(1,000)
		401.222	506.17	1,000	407.191	576.79	1,000	484.182	652.83	1,000	675.152	614	(1,000)
.20	255.233	436.29	1,000	213.676	426.14	1,000	184.862	385.60	1,000	661.484	571	(1,000)	
	160.295	270.15	1,000	152.960	326.20	1,000	139.037	318.08	1,000	617.992	564	(1,000)	
21	.01	660.927	587.57	1,000	633.645	630.36	1,000	710.167	693.60	1,000	659.445	546	(1,000)
		598.634	608.76	1,000	616.925	663.47	1,000	679.790	712.20	1,000	684.967	595	(1,000)
	.05	290.267	414.79	1,000	299.281	431.47	1,000	303.062	551.85	1,000	583.471	561	(1,000)
		207.392	332.31	1,000	235.890	380.45	1,000	318.009	484.18	1,000	566.157	582	(1,000)
	.10	60.724	57.88	1,000	58.018	91.66	1,000	87.096	180.18	1,000	477.613	514	(1,000)
		49.943	44.38	1,000	44.409	70.46	1,000	66.875	127.74	1,000	461.652	516	(1,000)
.20	27.560	5.44	1,000	19.658	5.70	1,000	17.293	8.71	1,000	364.095	492	(1,000)	
	26.846	4.77	1,000	18.267	4.98	1,000	15.219	7.82	1,000	271.852	487	(1,000)	
51	.01	361.813	491.05	1,000	429.816	499.45	990	468.820	551.68	973	548.712	525.0	(997)
		296.921	444.04	1,000	415.805	522.43	976	447.705	532.42	946	572.950	546.0	(994)
	.05	33.837	14.44	1,000	35.525	21.71	989	48.179	39.69	976	141.408	186.0	(994)
		29.838	12.76	1,000	31.218	18.28	975	43.455	35.13	954	144.954	223.0	(945)
	.10	19.304	5.18	999	17.594	6.13	993	19.349	8.31	920	34.153	20.0	(1,000)
		17.837	5.11	997	16.378	5.80	979	18.232	7.68	945	33.733	18.0	(993)
.20	12.583	2.51	1,000	10.833	2.83	987	10.405	3.03	969	14.375	5.0	(1,000)	
	11.812	2.63	1,000	10.236	2.87	968	10.051	3.00	941	14.484	5.1	(994)	
101	.01	131.990	134.10	958	173.570	221.61	921	232.686	294.69	916	288.116	312.0	(948)
		120.065	123.72	919	158.128	193.51	889	222.891	285.48	883	301.146	357.0	(948)
	.05	24.274	8.33	959	24.998	9.39	927	26.999	10.91	909	36.724	16.2	(971)
		23.284	8.16	925	24.136	8.99	891	26.116	10.28	871	38.117	16.8	(972)
	.10	15.267	3.91	963	14.680	4.46	923	15.416	5.08	908	19.009	6.3	(972)
		14.850	4.02	922	14.421	4.39	890	15.146	4.97	884	19.819	6.8	(960)
.20	10.166	2.33	944	9.308	2.34	932	9.087	2.50	906	10.637	3.2	(968)	
	9.857	2.46	898	9.198	2.34	894	8.997	2.48	866	10.874	3.3	(963)	
201	.01	70.290	35.71	832	79.964	45.96	811	94.916	58.99	808	120.279	84.0	(882)
		69.077	34.25	791	79.290	43.96	772	94.133	60.06	759	125.460	87.0	(883)
	.05	21.255	6.18	839	21.805	7.09	807	23.525	8.36	789	26.670	9.7	(894)
		21.163	6.20	791	21.670	6.91	770	23.192	8.13	756	27.753	10.1	(872)
	.10	13.859	3.55	842	13.439	3.73	807	13.849	4.06	808	15.647	4.8	(864)
		13.746	3.61	772	13.437	3.68	772	13.733	4.02	763	16.116	5.1	(856)
.20	9.169	2.15	842	8.789	2.00	814	8.744	2.15	773	9.449	2.6	(885)	
	9.133	2.19	804	8.847	1.98	773	8.809	2.17	752	9.657	2.7	(854)	
501	.01	56.124	23.23	548	61.851	25.94	538	68.812	30.04	537	74.879	34.0	(577)
		56.261	23.25	522	61.558	25.53	520	68.409	29.94	509	77.084	35.0	(573)
	.05	19.382	5.58	524	20.017	6.39	530	21.137	6.89	539	24.074	7.7	(598)
		19.628	5.41	514	20.123	6.37	520	21.025	6.85	520	24.856	8.2	(582)
	.10	13.118	3.26	550	12.892	3.40	527	13.383	3.63	548	14.156	4.3	(563)
		13.229	3.31	515	12.964	3.40	507	13.463	3.58	529	14.550	4.4	(582)
.20	8.941	2.23	527	8.587	1.97	560	8.443	2.10	530	8.990	2.5	(571)	
	9.064	2.16	514	8.667	1.97	541	8.505	2.10	517	9.115	2.5	(564)	

NOTE: The six numbers in each cell refer to mean (Cusum) s.d. (Cusum) no. of repetitions  $\geq \nu$  (Cusum) mean (Shiryayev-Roberts) s.d. (Shiryayev-Roberts) no. of repetitions  $\geq \nu$  (Shiryayev-Roberts).

and both are superior to the representative  $\theta = .05$ . When the possibility that the post-change regression may also have non-linear terms is relevant, Table 2 (coupled with Table 4) suggests that representative  $\theta = .20$  (Shiryayev–Roberts) may be the preferable method.

### 6. AN APPLICATION

It is generally recognized that we live in a period of global warming. Yearly average global temperatures have increased by  $1^\circ\text{F}$  over the last century. The phenomenon is described in the May 1998 issue of *National Geographic* (Suplee 1998). Figure 1 describes the global yearly average air temperature at Earth’s surface since World War II. (This plot is part of the diagram on p. 45 of Suplee 1998, from which we obtained the data on which Fig. 1 and the following analyses are based.)

In this section we address the more subtle question of the rate of growth of global warming. Is the rate of growth of global warming constant, or has it been increasing? Clearly, this is a question of change of slope of regression. Even the onset of the phenomenon is more of a question of change of slope than an abrupt change of level; after all, the causes to which the global warming phenomenon is commonly attributed have arisen gradually, so it stands to reason that the phenomenon itself will emerge gradually. Here we apply the methods developed in this article to analyze the post–World War II data. As we show, the assumption of iid normally distributed errors seems to be valid in stretches of data where the slope is constant and the variance of the errors does not change when the slope changes, and thus the conditions of Section 4 are satisfied.

We apply a Shiryayev–Roberts scheme for detecting an increase of slope. We analyze the data as if surveillance were started at 1945, deciding on line (at the end of each year) whether to raise an alarm. Thus we have no learning sample, and we have no prior knowledge of baseline nuisance parameter values. Consistent with our findings in the preceding section, we apply a Shiryayev–Roberts scheme with  $\theta = .2$  as the representative post-change parameter value. The ARL to

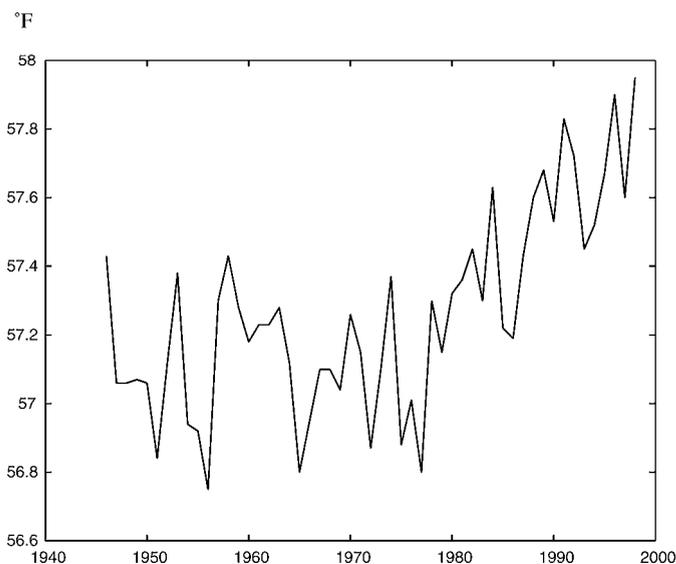


Figure 1. Post-World War II Yearly Average Global Temperatures at Earth’s Surface.

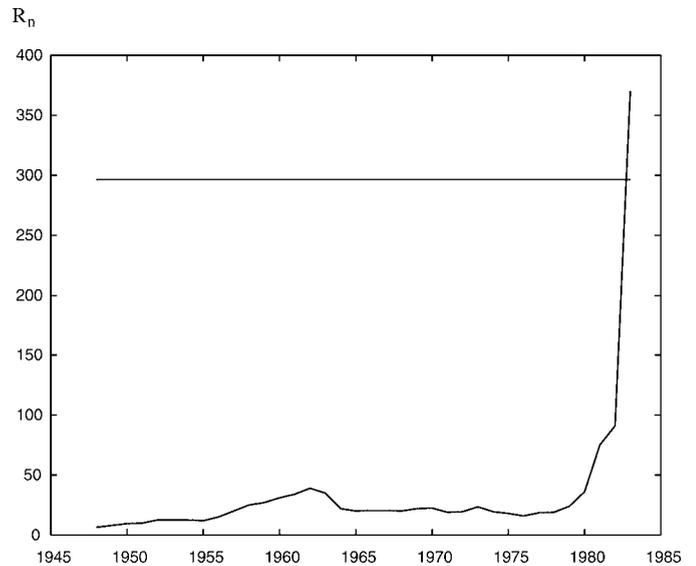


Figure 2. The Representative  $\theta = .2$ . Shiryayev–Roberts Control Chart for an Increase in the Regression Slope of Post-World War II Global Warming Data. The cutoff value is  $A = 296.4$ , for which the corresponding ARL to false alarm is 750.

false alarm is 750, which by Table 3 implies a cutoff value of  $A = 296.4$ . We emphasize that the following analysis is meant to be a demonstration of an application of our methodology, not a full-scale analysis of the global warming phenomenon.

The Shiryayev–Roberts control chart is plotted in Figure 2. The scheme raises an alarm at 1983. Figure 3 is a plot of the likelihood ratios  $\Lambda_k^{n=1983}, k \leq 1983$ . The likelihood is maximal at  $k = 1976$ ; therefore, we estimate that the slope increased beginning in 1976. In other words, starting from 1975–1976, there is a new slope.

At 1983 we start surveillance anew, retroactively using data beginning with 1975. An analysis of post-1975 data (not shown here) does not indicate a further increase in slope.

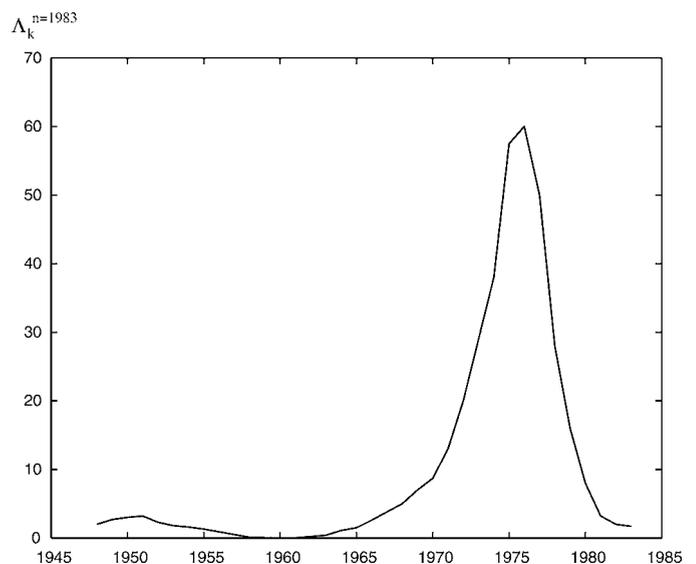


Figure 3. The Likelihood Ratios  $\Lambda_k^{n=1983}$  at the Point at Which an Alarm is Raised.

Simple estimates of the parameter values are as follows. For 1945–1975,

$$\text{temperature } (^\circ\text{F}) = 60.710645 - .001839 \times \text{year}$$

(the slope is not significantly different from 0; the  $p$  value is .6278); root mean squared error (RMSE) equals .186852. For 1975–1997,

$$\begin{aligned} \text{temperature } (^\circ\text{F}) &= -9.929249 + .0339328 \cdot \text{year}, \\ \text{RMSE} &= .167419. \end{aligned}$$

In standard deviation units, the change in slope is approximately .2. The change occurred 30 observations after surveillance commenced. Referring back to Table 4, the average delay is somewhere between 7.82 and 3.00. The delay in this example is 8, which is consistent with the expected value. For the two periods, the Durbin–Watson statistics are 2.03 and 1.59, and the  $p$  values of the Shapiro–Wilks test for normality (of the residuals) are .2892 and .5833.

It may be of interest to compare our results with a fully retrospective analysis (done taking all of the observations into consideration). Wu, Woodroffe, and Mentz (2001) analyzed these data by isotonic regression and also found that the slope was flat from the end of World War II until 1976 (see their fig. 2), at which time an increase occurs. Note that our analysis reaches this conclusion in 1983 (instead of in 1999).

### 7. DISCUSSION

In the previous sections we have described in great detail some of the cases that one may face in practice. In addition, emphasis on certain theoretical nuisances might lead one to use slightly different versions of the control strategies used. In this section we discuss some of these considerations.

1. There are a number of other contexts that can be approached by the same methods; examples include the following:
  - a. The initial intercept and slope are known to be 0 (which is equivalent to their being known, although different from 0), but the variance is unknown.
  - b. The initial slope is known to be 0, but the intercept (the mean level) and the variance are not known.
  - c. The change may be in any or all of intercept, slope, and variance, where the initial parameters may be known, partially known, or completely unknown.

In this article, apart from their intrinsic interest, we chose to consider the known baseline case because it serves as a natural benchmark and the completely unknown baseline case because it is the hardest. We chose to concentrate only on a change of slope because it is natural to many contexts.

2. In principle, one can choose arbitrary  $m \geq 1$  and base the surveillance only on  $\Lambda_k^n$  for  $k \geq m$ . In this article, we choose  $m = 4$ . Choosing  $m > 4$  will have the effect of smaller expected delay when  $\nu \geq m$  at the price of longer expected delay when  $\nu < m$ .
3. In view of the apparent linearity of the cutoff values as a function of the ARL to false alarm, one would suspect that a theorem to that effect would be true, as is the case in the classical changepoint context. We conjecture that the

context studied in this article differs essentially from the classical changepoint problem, as the following reasoning indicates.

Consider the known baseline representative  $\theta$  Shiriyayev–Roberts scheme. Under  $P_\infty$ , when  $A$  is large, argue by analogy with the classical changepoint problem (Pollak 1987) that the expected value of the overshoot of  $\sum_{k=1}^n \Lambda_k^n(\theta)$  over  $A$  has a behavior similar to that of the expected value of the overshoot of the sequential probability ratio test statistic  $\Lambda_1^n(\theta)$  over  $A$ . In other words, we are looking for the  $P_\infty$ -expected value of  $\Lambda_1^\tau(\theta)/A$  conditional on  $\{\tau < \infty\}$ , where  $\tau = \min\{n \mid \Lambda_1^n(\theta) \geq A\}$ , when  $A$  is large. Note that  $E_\infty(\Lambda_1^\tau(\theta) \mid \tau < \infty) = 1/P_\infty(\tau < \infty)$ . Using the probability transformation characteristic of sequential analysis, note that

$$\begin{aligned} P_\infty(\tau < \infty) &= E_{\nu=1} \frac{1}{\Lambda_1^\tau(\theta)} \\ &= E_{\nu=1} \frac{\exp\{-(\log \Lambda_1^\tau(\theta) - \log A)\}}{A}. \end{aligned} \tag{31}$$

Also note that

$$\log \Lambda_1^n = \theta \sum_{i=1}^n iY_i - \frac{1}{2} \theta \frac{n(n+1)(2n+1)}{6}, \tag{32}$$

and so for large  $n$ ,

$$\begin{aligned} E_{\nu=1} \log \Lambda_1^n &= (\theta^2/12)n(n+1)(2n+1) \\ &= (1 + o(1))(\theta^2/6)n^3. \end{aligned} \tag{33}$$

The standard deviation of  $\log \Lambda_1^n$  is

$$\begin{aligned} \sqrt{\text{var}_{\nu=1} \log \Lambda_1^n} &= \sqrt{(\theta^2/6)n(n+1)(2n+1)} \\ &= (1 + o(1))(\theta/\sqrt{3})n^{3/2}. \end{aligned} \tag{34}$$

Let  $n_A = \min\{n \mid (\theta^2/12)n(n+1)(2n+1) \geq \log A\}$ . Note that

$$\begin{aligned} E_{\nu=1} \log \Lambda_1^{n_A} - E_{\nu=1} \log \Lambda_1^{n_A-1} \\ &= E_{\nu=1} \left[ \log \left( \frac{f_{\nu=1}(Y_{n_A})}{f_{\nu=\infty}(Y_{n_A})} \right) \right] \\ &= (\theta^2/2)n_A^2. \end{aligned} \tag{35}$$

This implies that for most values of  $A$ , the expressions  $E_{\nu=1} \log \Lambda_1^{n_A} - \log A$  and  $\log A - E_{\nu=1} \log \Lambda_1^{n_A-1}$  are both of the order of magnitude  $n_A^2$ . This, together with (34), implies that for most values of  $A$ ,  $\tau$  is almost degenerate;  $P_{\nu=1}(\tau = n_A) \approx 1$ . This implies that  $E_{\nu=1} \exp\{-(\log \Lambda_1^\tau - \log A)\}$  is not constant in  $A$ , which in turn would mean that the ARL to false alarm of the Shiriyayev–Roberts scheme is not asymptotically linear in  $A$  (although for short stretches of  $A$  it may appear to be linear). For contiguous alternatives, there does seem to be a theory that will yield an asymptotically linear expression in  $A$  (B. Yakir, private communication), and perhaps the combinations of  $A$  and  $\theta$  studied in this article can be said to fall in this category.

4. In the known baseline case, the GLR Cusum designed for detecting an abrupt change of mean from 0 to a different fixed value is a viable option. However, in the unknown baseline case considered in Section 4, this avenue is closed. The reason is that it is not possible to make a simple reduction of the problem to one based on residuals of the pre-change regression, because the regression parameters are unknown. The ARL to false alarm of the GLR procedure appropriate for the unknown baseline case of detecting an abrupt increase of mean from one unknown value to another is calculated under the assumption that all observations have the same mean. If this procedure were to be applied to the original data considered in the change-of-slope-of-regression detection problem (where pre-change mean is a function not constant in time), then the true ARL to false alarm would be (very) different from the nominal one.
5. The crux of our article deals with an abrupt change of regression slope. One can imagine a more refined version, where the change of slope is gradual. Although one can construct models for a continuous change of slope that enable computational tractability, our experience (based on Monte Carlo) indicates that nothing is to be gained by such a refinement. Intuitively, the reason for this is that such a refinement is basically a modification of the very early departures from the in-control state, and that a change is in effect typically will be hard to distinguish at such an early stage. By the time a change becomes apparent, the slope will be close to being constant, and the time of onset of the change will be in the background. Because the most recent observations carry the greater weight in the surveillance statistic, modeling the start of the change as an abrupt change of slope or as a continuous one will matter little.

## APPENDIX A: TECHNICAL DETAILS

Here we present technical details that lead to the formulas given in Sections 2 and 4, and describe computational issues that arise in implementing the various schemes.

### A.1 Known Baseline Parameters

*A.1.1 Choosing a Representative.* If  $\theta$  is the chosen representative, then, under  $P_k$ ,  $Y_i \sim N(0, 1)$  for  $i < k$  and  $Y_i \sim N(\theta(i - k + 1), 1)$  for  $i \geq k$ . Because the  $Y_i$ 's are independent, the  $P_k$  versus  $P_\infty$  likelihood ratio of  $Y_1, \dots, Y_n$  is

$$\begin{aligned} \Lambda_k^n(\theta) &= \frac{f_{P_k}(Y_1, \dots, Y_n)}{f_{P_\infty}(Y_1, \dots, Y_n)} = \prod_{i=1}^n \frac{f_{P_k}(Y_i)}{f_{P_\infty}(Y_i)} = \prod_{i=k}^n \frac{f_{P_k}(Y_i)}{f_{P_\infty}(Y_i)} \\ &= \prod_{i=k}^n \exp\{\theta(i - k + 1)Y_i - \theta^2(i - k + 1)^2/2\} \\ &= \exp\left\{\theta \sum_{i=k}^n (i - k + 1)Y_i - \theta^2(n - k + 1) \right. \\ &\quad \left. \cdot (n - k + 2)(2n - 2k + 3)/12\right\}. \end{aligned} \tag{A.1}$$

*A.1.2 Choosing a Prior and Creating a Mixture.* For

$$F(x) = \frac{\Phi\left(\frac{x-\mu}{\tau}\right) - \Phi\left(-\frac{\mu}{\tau}\right) + \Phi\left(\frac{\mu}{\tau}\right)}{\Phi\left(\frac{\mu}{\tau}\right)},$$

$$\begin{aligned} \Lambda_k^n(F) &= \int_{-\infty}^{\infty} \Lambda_k^n(\theta) dF(\theta) \\ &= \frac{1}{\Phi\left(\frac{\mu}{\tau}\right)} \int_0^{\infty} \exp\left(\theta \sum_{i=k}^n (i - k + 1)Y_i - \frac{\theta^2(n - k + 1)(n - k + 2)(2n - 2k + 3)}{12}\right) \\ &\quad \cdot \frac{1}{\sqrt{2\pi} \tau^2} \exp\left\{-\frac{1}{2} \frac{(\theta - \mu)^2}{\tau^2}\right\} d\theta \\ &= \frac{\exp\left\{-\frac{1}{2} \frac{\mu^2}{\tau^2}\right\} \exp\left\{\frac{1}{2} \frac{(\sum_{i=k}^n (i - k + 1)Y_i + \frac{\mu}{\tau})^2}{\frac{1}{\tau^2} + \frac{(n - k + 1)(n - k + 2)(2n - 2k + 3)}{6}}}\right\}}{\tau \sqrt{\frac{1}{\tau^2} + \frac{(n - k + 1)(n - k + 2)(2n - 2k + 3)}{6}}} \\ &\quad \cdot \frac{\Phi\left(\frac{\sum_{i=k}^n (i - k + 1)Y_i + \frac{\mu}{\tau}}{\sqrt{\frac{1}{\tau^2} + \frac{(n - k + 1)(n - k + 2)(2n - 2k + 3)}{6}}}\right)}{\Phi\left(\frac{\mu}{\tau}\right)}. \end{aligned} \tag{A.2}$$

A similar calculation for  $F(x) = \Phi((x - \mu)/\tau)$  gives the same result, except that the  $\Phi(\cdot)/\Phi(\cdot)$  part of (A.2) is cancelled. This accounts for (10).

*A.1.3 Choosing an Estimator.* From (A.1), we obtain that the  $P_k$  versus  $P_\infty$  log-likelihood ratio of  $Y_1, \dots, Y_{n-1}$  is

$$\theta \sum_{i=k}^{n-1} (i - k + 1)Y_i - \frac{\theta^2(n - k)(n - k + 1)(2n - 2k + 1)}{12}.$$

Therefore, the maximum  $P_k$ -likelihood estimator of  $\theta$  based on  $Y_1, \dots, Y_{n-1}$  is

$$\hat{\theta}_k^n = \frac{\sum_{i=k}^{n-1} (i - k + 1)Y_i}{\frac{1}{6}(n - k)(n - k + 1)(2n - 2k + 1)},$$

which accounts for (13).

*A.1.4 A Semiparametric Estimator.* For  $i \geq k$ ,

$$\frac{f_{P_k}(Y_i)}{f_{P_\infty}(Y_i)} = \exp\{Y_i E_k Y_i - (E_k Y_i)^2/2\}. \tag{A.3}$$

Estimating  $E_k Y_i$  by  $Y_{i-1}^+$  or by  $Y_{i-1}$  accounts for (16). Although this is a very simplistic estimate (and generally is an underestimate/overestimate for an increase/decrease in slope), it is obviously robust against departures from a linear regression, and could be expected to do well for changes of large magnitude. This intuition is corroborated by Table 2.

### A.2 Unknown Baseline Parameters

*A.2.1 Choosing a Representative.* Due to the invariance property of  $W_i$  and  $W_i^*$ , calculation of  $\Lambda_k^n$  and  $\Lambda_k^{n*}$  can be done without loss of generality for  $\alpha = \beta = 0$ ,  $\sigma = 1$ . We sketch a derivation of  $\Lambda_k^{n*}$ ; the derivation of  $\Lambda_k^n$  is completely analogous. A full derivation can be obtained from the authors on request.

Note that, conditional on  $Z_2$  and the denominator of  $W_i^*$ , the variables  $W_4^*, W_5^*, \dots$  are (conditionally) independent. Therefore, we evaluate  $\Lambda_k^{n*}$  by first conditioning on  $Z_2$  and the denominator.

For  $k \geq 4$ , the variables  $Z_2$  and  $Z_3$  have the same distribution under  $P_k$  as under  $P_\infty$ . By conditioning first on  $Z_2$ , note that

$$f(x, y)_{Z_2, Z_3 - Z_2 \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1}} = \phi(x) \phi\left(y + x \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1}\right), \tag{A.4}$$

so that

$$f(x)_{Z_2 | Z_3 - Z_2 \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1} = y} = \frac{1}{\sqrt{2\pi \frac{1}{1 + \frac{4}{3} [b] \left(\frac{x_3 - \bar{x}_2}{x_2 - x_1}\right)^2}}} \cdot \exp\left\{-\frac{1}{2} \frac{\left(x + y \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1} / \left(1 + \frac{4}{3} \left(\frac{x_3 - \bar{x}_2}{x_2 - x_1}\right)^2\right)\right)^2}{\frac{1}{1 + \frac{4}{3} \left(\frac{x_3 - \bar{x}_2}{x_2 - x_1}\right)^2}}\right\}. \tag{A.5}$$

Denote

$$\begin{aligned} \gamma_{i,v} &= 0 \quad \text{for } i \leq v - 1, \\ \mu_{i,v} &= E_\nu Z_i = 0 \quad \text{for } i \leq v - 1, \\ \mu_{i,v} &= E_\nu Z_i = \sqrt{[b] \frac{i-1}{i}} (\gamma_{i,v} - \bar{\gamma}_{i-1,v}) \quad \text{for } i \geq v, \end{aligned} \tag{A.6}$$

where

$$\bar{\gamma}_{j,v} = \sum_{i=1}^j \frac{\gamma_{i,v}}{j}.$$

For  $k \geq 4$ ,

$$\begin{aligned} f_{v=k; W_4^*, \dots, W_n^*}(w_4, \dots, w_n) &= \frac{\partial^{n-3} E_k P_k(W_4^* \leq w_4, \dots, W_n^* \leq w_n | Z_2, Z_3 - Z_2 \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1})}{\partial w_4 \dots \partial w_n} \\ &= E_\infty \left\{ \left( \frac{|Z_3 - Z_2 \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1}|}{\sqrt{2\pi}} \right)^{n-3} \cdot \exp\left\{-\frac{1}{2} \sum_{i=4}^n \left( Z_2 \sqrt{2} \sqrt{\frac{i-1}{i}} \frac{x_i - \bar{x}_{i-1}}{x_2 - x_1} + \left( Z_3 - Z_2 \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1} \right) w_i - \mu_{i,k} \right)^2 \right\} \right\}. \end{aligned} \tag{A.7}$$

Calculate this by conditioning on

$$Z_3 - Z_2 \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1} = y,$$

obtaining

$$\begin{aligned} f_{v=k; W_4^*, \dots, W_n^*}(w_4, \dots, w_n) &= \int_{-\infty}^{\infty} |z|^{n-3} e^{-\frac{1}{2}(z - \frac{b_{n,k}}{\sqrt{a_n}})^2} dz e^{\frac{1}{2} \frac{(b_{n,k})^2}{a_n} + c_{n,k}} d_n^{-\frac{1}{2}} (2\pi a_n)^{-\frac{n-2}{2}}, \end{aligned} \tag{A.8}$$

where

$$d_n = 2 \sum_{i=4}^n \frac{i-1}{i} \left( \frac{x_i - \bar{x}_{i-1}}{x_2 - x_1} \right)^2 + 1 + \frac{4}{3} \left( \frac{x_3 - \bar{x}_2}{x_2 - x_1} \right)^2,$$

$$a_n = 1 + \sum_{i=4}^n w_i^2 - d_n^{-1} \left( \sqrt{2} \sum_{i=4}^n \sqrt{\frac{i-1}{i}} w_i + \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1} \right)^2,$$

$$\begin{aligned} b_{n,k} &= \sum_{i=4}^n w_i \mu_{i,k} \\ &- d_n^{-1} \left( \sum_{i=4}^n \sqrt{2} \sqrt{\frac{i-1}{i}} \frac{x_i - \bar{x}_{i-1}}{x_2 - x_1} w_i + \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1} \right) \\ &\cdot \left( \sum_{i=4}^n \sqrt{2} \sqrt{\frac{i-1}{i}} \frac{x_i - \bar{x}_{i-1}}{x_2 - x_1} \mu_{i,k} \right), \end{aligned} \tag{A.9}$$

$$c_{n,k} = d_n^{-1} \left( \sum_{i=4}^n \sqrt{\frac{i-1}{i}} \frac{x_i - \bar{x}_{i-1}}{x_2 - x_1} \mu_{i,k} \right)^2 - \frac{1}{2} \sum_{i=4}^n \mu_{i,k}^2.$$

Let

$$g_m^*(\mu) = \int_{-\infty}^{\infty} |z|^m e^{-(z-\mu)^2/2} dz.$$

It follows that

$$\frac{f_{v=k}(w_4, \dots, w_n)}{f_{v=\infty}(w_4, \dots, w_n)} = \frac{g_{n-3}^*\left(\frac{b_{n,k}}{\sqrt{a_n}}\right)}{g_{n-3}^*(0)} \exp\left\{\frac{1}{2} \frac{(b_{n,k})^2}{a_n} + c_{n,k}\right\}, \tag{A.10}$$

from which the expression for  $\Lambda_k^{n*}$  in (23) follows. The expression for  $\Lambda_k^n$  in (23) is obtained in similar fashion. The only difference is that

$$Z_3 - Z_2 \frac{2}{\sqrt{3}} \frac{x_3 - \bar{x}_2}{x_2 - x_1}$$

is replaced in (A.7) by its absolute value, which causes the boundaries of the integral in (A.8) to be  $(0, \infty)$  and the integral to be multiplied by 2, which cancels in the ratio (A.10).

A.2.2 Choosing a Prior and Creating a Mixture. Note that

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_0^{\infty} x^m e^{-\frac{(x-\theta a)^2}{2}} dx e^{-\frac{\theta^2 b}{2}} \frac{1}{\sqrt{2\pi \tau^2}} e^{-\frac{(\theta-\mu)^2}{2\tau^2}} d\theta \\ &= \int_0^{\infty} x^m e^{-\frac{x^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \{\theta^2 [a^2 + b + \frac{1}{\tau^2}] - 2\theta [ax - \frac{\mu}{\tau^2}]\}} \\ &\quad \cdot \frac{1}{\sqrt{2\pi \tau^2}} d\theta e^{-\frac{\mu^2}{2\tau^2}} dx \\ &= \int_0^{\infty} x^m e^{-\frac{1}{2} [x^2 [1 - \frac{a^2}{a^2 + b + 1/\tau^2}] + 2x [\frac{a\mu/\tau^2}{a^2 + b + 1/\tau^2}]]} dx \\ &\quad \cdot \frac{e^{-\frac{1}{2} \frac{\mu^2}{\tau^2}} e^{\frac{1}{2} \frac{\mu^2}{\tau^4} \frac{1}{a^2 + b + 1/\tau^2}}}{\tau \sqrt{a^2 + b + 1/\tau^2}} \\ &= \int_0^{\infty} y^m e^{-\frac{1}{2} (y^2 + 2y \frac{a\mu/\tau^2}{\sqrt{a^2 + b + 1/\tau^2} \sqrt{b + 1/\tau^2}})} dy \\ &\quad \cdot \left( \frac{a^2 + b + 1/\tau^2}{b + 1/\tau^2} \right)^{\frac{m+1}{2}} \frac{e^{-\frac{1}{2} \frac{\mu^2}{\tau^2}} e^{\frac{1}{2} \frac{\mu^2}{\tau^4} \frac{1}{a^2 + b + 1/\tau^2}}}{\tau \sqrt{a^2 + b + 1/\tau^2}} \\ &= \int_0^{\infty} y^m e^{-\frac{1}{2} \left( y + \frac{a\mu/\tau^2}{\sqrt{a^2 + b + 1/\tau^2} \sqrt{b + 1/\tau^2}} \right)^2} dy \frac{(a^2 + b + 1/\tau^2)^{\frac{m}{2}}}{\tau (b + 1/\tau^2)^{\frac{m+1}{2}}} \\ &\quad \cdot e^{-\frac{1}{2} \frac{\mu^2}{\tau^2}} e^{\frac{1}{2} \frac{\mu^2}{\tau^4} \frac{1}{a^2 + b + 1/\tau^2} \left[ \frac{a^2}{b + 1/\tau^2} + 1 \right]}, \end{aligned} \tag{A.11}$$

from which (27) follows. Calculation of  $\Lambda_k^{n*}(F)$  is similar, with  $\int_0^\infty$  in (A.11) replaced by  $\int_{-\infty}^\infty$  of the absolute value of the integrand.

### A.3 Computational Issues

The major computational issue is the calculation of

$$\zeta(x) = \frac{g_m(x)}{g_m(0)}, \tag{A.12}$$

the difficulty being that both numerator and denominator become very large as  $m$  grows. Note that

$$\begin{aligned} \int_0^\infty z^m e^{-\frac{1}{2}(z-x)^2} dz &= e^{-\frac{1}{2}x^2} \int_0^\infty z^m e^{-\frac{1}{2}z^2} e^{xz} dz \\ &= e^{-\frac{1}{2}x^2} \sum_{j=0}^\infty \frac{x^j}{j!} \int_0^\infty z^{m+j} e^{-\frac{1}{2}z^2} dz \end{aligned} \tag{A.13}$$

and that

$$\int_0^\infty z^\ell e^{-\frac{1}{2}z^2} dz = \frac{1}{2} \int_0^\infty y^{\frac{\ell-1}{2}} e^{-\frac{1}{2}y} dy = 2^{\frac{\ell-1}{2}} \Gamma\left(\frac{\ell+1}{2}\right). \tag{A.14}$$

Hence

$$\begin{aligned} \zeta(x) &= e^{-\frac{1}{2}x^2} \sum_{j=0}^\infty \frac{x^j}{j!} 2^{\frac{m+j-1}{2}} \frac{\Gamma\left(\frac{m+j+1}{2}\right)}{\left[2^{\frac{m-1}{2}} \Gamma\left(\frac{m+1}{2}\right)\right]} \\ &= e^{-\frac{1}{2}x^2} \sum_{j=0}^\infty \frac{(\sqrt{2}x)^j}{j!} \frac{\Gamma\left(\frac{m+j+1}{2}\right)}{\Gamma\left(\frac{m+1}{2}\right)}. \end{aligned} \tag{A.15}$$

Operationally, we find an integer  $J$  such that

$$\sum_{j=J+1}^\infty \frac{(\sqrt{2}x)^j}{j!} \frac{\Gamma\left(\frac{m+j+1}{2}\right)}{\Gamma\left(\frac{m+1}{2}\right)}$$

is negligible and define

$$\begin{aligned} \mu_j &= j \log(\sqrt{2}x) + \log \Gamma\left(\frac{m+j+1}{2}\right) - \log \Gamma\left(\frac{m+1}{2}\right) \\ &\quad - \log j!, \end{aligned} \tag{A.16}$$

$$\mu_{\max} = \max_{1 \leq j \leq J} \mu_j,$$

where  $J$  and  $\mu_{\max}$  depend on  $x$ , so that

$$\zeta(x) \cong e^{-\frac{1}{2}x^2} \left( \sum_{j=0}^J e^{\mu_j - \mu_{\max}} \right) e^{\mu_{\max}}. \tag{A.17}$$

When programming the procedure, we first calculate  $\mu_1, \dots, \mu_J$ , from which we obtain  $\mu_{\max}$ . For all values of  $x$  that we encountered in extensive simulations (under conditions where there is no change as well as those where there is a change, either large or small, the value of  $J$  was easily manageable (a few thousand at most). Thus for practical purposes, (A.17) provides an accurate evaluation of  $\zeta(x)$ . There exists a recursion for calculating  $\zeta(x)$ , but using it gives far less accurate results than (A.17).

The second concern is computation time. It is clear that in the unknown baseline case, the computational complexity is considerable, and computation time increases rapidly with  $n$ . Although for a single application the operation time is not long enough to be a major issue, when simulating the procedure—especially when there is no change, and the procedure is run to estimate the ARL to false alarm—the simulation time is so large that a shortcut is needed. We solved the problem

by creating a grid of values  $\zeta(x)$  and interpolating from the grid for  $x$ 's arising in the simulation instead of calculating  $\zeta(x)$  every time anew. The interpolation is easy; we found that for values of  $x$  not very close to 0,  $\ln(\zeta(x))$  is almost perfectly linear in  $x$  when  $m$  is large. The time saved by calling the grid and interpolating instead of calculating  $\zeta(x)$  wherever it appears is tremendous, and is what makes the simulations feasible. If the sequence being monitored can become very large before a change occurs, then the grid approach should be used in practice.

The same method applies to the case where the post-change regression may have either higher slope or smaller slope. Simply break up the integral

$$\begin{aligned} &\int_{-\infty}^\infty |z|^m e^{-\frac{1}{2}(z-x)^2} dz \\ &= \int_0^\infty z^m e^{-\frac{1}{2}(z-x)^2} dz + (-1)^m \int_{-\infty}^0 z^m e^{-\frac{1}{2}(z-x)^2} dz \\ &= \int_0^\infty z^m e^{-\frac{1}{2}(z-x)^2} dz - \int_0^\infty t^m e^{-\frac{1}{2}(t+x)^2} dt, \end{aligned} \tag{A.18}$$

to which (48) and its sequel can be applied.

[Received August 2001. Revised June 2002.]

### REFERENCES

Aerne, L., Champ, C., and Rigdon, S. (1991), "Evaluation of Control Charts Under Linear Trend," *Communications in Statistics, Part A—Theory and Methods*, 20, 3341–3349.

Gan, F. F. (1992), "CUSUM Control Charts Under Linear Drift," *The Statistician*, 41, 71–84.

— (1996), "ARL for Cumulative Sum Control Charts Under Linear Trends," *Applied Statistics*, 45, 505–512.

Gordon, L., and Pollak, M. (1997), "Average Run Length to False Alarm for Surveillance Schemes Designed with Partially Specified Pre-Change Distribution," *The Annals of Statistics*, 25, 1284–1310.

Lorden, G., and Pollak, M. (2003), "Non-anticipating Estimation Applied to Sequential Analysis and Change Point Detection," California Institute of Technology Technical Report.

Moustakides, G. V. (1986), "Optimal Stopping Times for Detecting Changes in Distributions," *The Annals of Statistics*, 14, 1379–1387.

Pollak, M. (1985), "Optimal Detection of a Change in Distribution," *The Annals of Statistics*, 13, 206–227.

— (1987), "Average Run Lengths of an Optimal Method of Detecting a Change in Distribution," *The Annals of Statistics*, 15, 749–779.

Ritov, Y. (1990), "Decision Theoretic Optimality of the CUSUM Procedure," *The Annals of Statistics*, 18, 1464–1469.

Sigal, B. M. (1998), "Detection of a Smooth Signal in Fixed Sample and Sequential Problems," doctoral dissertation, Stanford University, Dept. of Statistics.

Suplee, C. (1998), "Unlocking the Climate Puzzle," *National Geographic*, 193, 38–71.

Wheeler, D. J. (2000), Discussion of "Controversies and Contradictions in Statistical Process Control," by W. H. Woodall, *Journal of Quality Technology*, 32, 361–363.

Wu, W. B., Woodroffe, H., and Mentz, G. (2001), "Isotonic Regression: Another Look at the Changepoint Problem," *Biometrika*, 88, 793–804.

Yakir, B. (1997), "A Note on Optimal Detection of a Change in Distribution, When the Observations Are Independent," *The Annals of Statistics*, 25, 2117–2126.

— (1998), "On the Average Run Length to False Alarm in Surveillance Problems Which Possess an Invariance Structure," *The Annals of Statistics*, 26, 1198–1214.

Yakir, B., Krieger, A. M., and Pollak, M. (1999), "Detecting a Change in Regression: First-Order Optimality," *The Annals of Statistics*, 27, 1896–1913.

Yao, Q. (1993), "Asymptotically Optimal Detection of a Change in a Linear Model," *Sequential Analysis*, 12, 201–210.