Detecting "Small" Mean Shifts in Time Series
Author(s): John U. Farley and Melvin J. Hinich
Source: *Management Science*, Vol. 17, No. 3, Theory Series (Nov., 1970), pp. 189-199
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/2629088
Accessed: 17/05/2010 16:09

DETECTING "SMALL" MEAN SHIFTS IN TIME SERIES*

JOHN U. FARLEY† AND MELVIN J. HINICH‡§

Analyzing time series data for evidence of changes in market conditions is a problem central to development of a marketing information system. Examples are tracking field reports on competitors' prices and analyzing market share estimates computed from retail audits and consumer panel reports. These series may involve substantial random measurement error, and any changes, often small relative to this error, are difficult to detect promptly and accurately. One way to approach this problem is with techniques recently developed for handling time series with small stochastic mean shifts as well as random error. Two special procedures are compared in Monte Carlo analyses with a simpler data filtering and control-chart approach; the latter appears the most promising.

I. Introduction

Developing a marketing information system involves a variety of rather unique statistical problems; partial solutions are available for some of these problems, while no formal solutions are now available for others. For example, such problems face a group which is responsible for keeping management of a major petroleum company apprised of movements in competitors' prices for a wide array of products. Several types of data flow in at different intervals—daily, weekly, and monthly. Some data are reported from within the firm, others come from commercial data collection sources and still others from numerous government agencies. The data inputs are subject to various reporting errors and distortions which might be considered random error or "noise". Prices in these markets may shift discretely rather than drifting up and down, and there may be a substantial time lag between a *de facto* change in price by one competitor and a public announcement that such a change has taken place. Tracking these price series for evidence of changes in competitive conditions is thus a major function of the information system.

The process is complicated by the fact that there may be several thousand separate series to track, and only a handful of the array can possibly receive explicit attention at any one time. To function smoothly and promptly, the computer-based information system requires built-in statistical procedures which seek evidence of shifts in these many series. Prompt, accurate evaluation of current conditions is vital. It is important to identify *bona fide* changes as soon as possible to initiate tactical response to changed conditions, but it is likewise important to avoid needlessly initiating major shifts in price or advertising which might be unprofitable or which might lead to degeneration of a stable market equilibrium. The need for speed dictates response as soon as possible, while the need for accuracy suggests waiting for as many observations as possible to be sure that a change has actually occurred. Besides tracking prices, similar problems occur in connection with detecting significant shifts in company sales series or in changes in market shares reported by panels or audits [5]. Such shifts are also often infrequent, discrete and subtle. It is tempting to develop and implement *ad hoc* statistical procedures in an attempt to balance off these two types of risks, but there are a variety of

such procedures to choose among and there are also a number of statistical techniques with known properties which might be especially well suited for these purposes.

The most direct statistical analogy to the problem just described is to consider that each series is subject at any point of time to random parameter change which is small relative to the variance of measurement errors. It is important to distinguish this problem from other parameter shift problems which should be handled quite differently. For example, if a parameter shift is very large relative to the variance of the measurement error, simple visual examination of the series will in all likelihood find a shift immediately. Similarly, if the position of any potential shift is pinpointed and only its existence at that point uncertain, regression techniques provide tests of whether a shift in fact occurred with very good properties [2].

However, the problem is quite different if a shift is likely to be small relative to the measurement error variance and if the analyst, unsure whether a shift occurs at all, thinks that one might occur at any point in the series.

II. The Mean-Shift Problem

Formally, the tracking problem is modeled as a stochastic shift in the mean of a time series. A stochastic process, $X(t) = \theta(t) + N(t)$, generates sequence data, where where $\theta(t)$ is a mean parameter (possibly shifting, possibly stable) and $N(t)$ is stationary Gaussian noise with zero mean and a known covariance function, Σ. Any discrete parameter change can be modeled as a step function. For $i = 1, 2, \cdots$,

$$(1) \qquad \theta(t) = \theta_i \quad \text{when} \quad t_{i-1}^* < t \leqq t_i^*.$$

It is useful to assume that the number of points when θ changes (the t_i^*) per unit time are Poisson distributed with a shift rate parameter, γ, which is assumed known [6]. (The change rate, as we shall see, is very important and will cause considerable difficulty in dealing with this problem. In particular, γ is seldom known in applications like those discussed in §I.) $\theta(t)$ is thus the expected value of $X(t)$, and it changes from one constant value to another at random times t_i^*.

This paper uses artificial data to examine the behavior of three alternative estimation procedures—two especially developed to deal with this particular problem and one based on more general filter techniques. The paper also examines two processing procedures—one in which a finite number of discrete observations of the series are formed into a finite record which is processed as a whole, and the other in which the series is considered as a running record and processed with filter techniques. The filter technique is based on procedures resembling control charts on filtered data produced by smoothing raw data.

III. The Discrete Observation, Fixed Record Case

The time series is viewed as if it were drawn during nonoverlapping periods nS time units long. Each record consists of n successive observations of $X(t)$ where the observations are taken S time units apart. If nS, the record length, is smaller than γ^{-1}, the mean time between shifts, the possibility of two or more shifts in a record can be disregarded. The sample points constitute the random vector $\{X_1, \cdots, X_n\}$, where each X_i is the value of the random process $X(t) = \theta(t) + N(t)$ at time t_i. The mean is a step function,

$$(2) \qquad \begin{aligned} \theta(t) &= \theta &&\text{if } t < t^*, \\ &= \theta + \mu &&\text{if } t \geqq t^*, \end{aligned}$$

where t^* is the time of a random shift in the mean. Since we are interested in shifts small relative to the measurement error, we assume that μ is smaller than σ, the standard deviation of $N(t)$. Two methods for estimating the post-shift mean are compared—one developed by the authors and one by Chernoff and Zacks.

(a) *Hinich-Farley estimators* [6].

Estimators of θ and μ which are unbiased and efficient in a special sense turn out to be linear combinations of the vector of observations, the error covariance matrix and the shift rate. The estimators are

(3)
$$\hat{\theta} = \frac{(a\Sigma^{-1}a')(X\Sigma^{-1}1') - (1\Sigma^{-1}a')(X\Sigma^{-1}a')}{(1\Sigma^{-1}1)(a\Sigma^{-1}a') - (1\Sigma^{-1}a')^2},$$

$$\hat{\mu} = \gamma^{-1}\frac{(1\Sigma^{-1}1')(X\Sigma^{-1}a') - (1\Sigma^{-1}a')(X\Sigma^{-1}1')}{(1\Sigma^{-1}1')(a\Sigma^{-1}a') - (1\Sigma^{-1}a')^2},$$

where Σ is the covariance matrix, 1 a unit row vector, $a = \{1, 2, \cdots, n\}$, and X the row vector of observations.

If the errors are homoscedastic and serially independent (that is, "white noise"), Σ is the scalar matrix $\sigma^2 I$, so

(4)
$$\hat{\theta} = \frac{2}{n(n-1)}[(2n+1)\Sigma_j X_j - 3\Sigma_j j X_j],$$

$$\hat{\mu} = \frac{6}{n(n-1)\gamma}\left[\frac{2}{n+1}\Sigma_j j X_j - \Sigma_j X_j\right].$$

These estimators are easy to calculate and contain most of the information about the parameters in a range [6].

(b) *The Chernoff-Zacks ad hoc procedure* [1]

Starting with a more general and very complex Bayesian estimator, Chernoff and Zacks developed an *ad hoc* procedure to handle this problem. A generalization of their approach is given by Mustafi [8]. The procedure uses a sequence of tests starting at the end of the series to attempt to detect any shift in a series; then an "at most one shift" estimator is applied to the observations following the point at which the first shift was detected in order to estimate the mean as of the last observation. If no shift is detected, the entire series is used to estimate the mean.

The basic elements of the procedure are, for a sample of n points, the *a-posteriori* probability of a shift having taken place at point $(n-k)$, given what has happened in the last m observations:

(5) $$B_n(m, k) = D_m^{*-1}P_k(m)\psi_k(X^{(m)}), \quad k = 0, 1, \cdots, m-1,$$

where

(6) $$X^{(m)} = \{X_{n-m+1}, \cdots, X_n\},$$

the last m observations in the series,

(7) $$P_0^{(m)} = (1-\gamma)^{m-1},$$

the likelihood that no shift occurred in $m-1$ observations,

(8) $$P_k^{(m)} = \gamma(1-\gamma)^{m-2} \quad k = 1, 2, \cdots, m-1,$$

the likelihood that one shift did occur. (This assumes, as did the Hinich-Farley estimators, an equal likelihood of a shift between any two points.) Further,

$$(9) \quad \psi_k(X^{(m)}) = \frac{1}{[m + \sigma^2 k(m-k)]^{\frac{1}{2}}} \exp \frac{1}{2} \left[\frac{k^2(m-k)^2\sigma^2}{m^2 + \sigma^2 mk(m-k)} (\bar{X}_k^{(m)} - \bar{X}_{m-k}^{*(m)})^2 \right]$$

$$\text{for } k = 0, 1, \cdots, n-1,$$

in which σ is the size of the shift and γ is the shift rate parameter as before. Expression (9) also involves two partial means,

$$(10) \quad \bar{X}_k^{(m)} = \frac{1}{k} \sum_{i=n-m+1}^{n-m+k} X_i$$

and

$$(11) \quad \bar{X}_{m-k}^{*(m)} = \frac{1}{m-k} \sum_{i=n-m+k+1}^{n} X_i,$$

further weighted to form

$$(12) \quad D_m^* = \sum_{k=0}^{m-1} P_k^{(m)} \psi_k(X^{(m)}).$$

Using K_m as the value of k which maximizes $B_n(m, k)$, the *ad hoc* procedure consists of computing $K_2, K_3, \cdots$ until $K_m \neq 0$ is reached, say at K^*. This means it is more likely that a shift occurred at K^* than that no change occurred at all from K^* on. Then an "at most one change" Bayes estimator is applied to the observations X_{n-K^*+1}, $X_{n-K^*+2}, \cdots, X_n$ to estimate the mean. This Bayes estimator for large σ, μ_n^{**}, is closely related to the detector (K_m):

$$(13) \quad \mu_n^{**} = (D^{**})^{-1} \left\{ \frac{P_0}{n^{\frac{1}{2}}} \bar{X}_n + \frac{1}{\sigma} \sum_{m=1}^{n-1} \frac{P_m \bar{X}_{n-m}^*}{\sqrt{m(n-m)}} \right.$$
$$\left. \cdot \exp \left[\frac{1}{2} \frac{m(n-m)}{n} (\bar{X}_m - \bar{X}_{n-m}^*)^2 \right] \right\},$$

where D^{**} is the limit of D^* as $\sigma \to \infty$, i.e.,

$$(14) \quad D^{**} = \frac{P_0}{n^{\frac{1}{2}}} + \frac{1}{\sigma} \sum_{m=1}^{n-1} \frac{P_m}{\sqrt{m(n-m)}} \exp \left[\frac{1}{2} \frac{m(n-m)}{n} (\bar{X}_m - \bar{X}_{n-m}^*)^2 \right].$$

The Chernoff-Zacks procedure is more complex than the Hinich-Farley procedure, but there is a distinct appeal to attempting to detect the shift and to disregarding data before it in estimating the mean.

IV. Monte Carlo Analysis of the Procedures

Several false starts with actual marketing data showed that there was too much specification uncertainty to provide an adequate test of the procedures' performance, and artificial data of known configuration provided a more natural starting point. The performance of the two estimators was compared using series of 100 observations with errors $(N(0, I))$, generated by the method discussed by Kronmal [7]. $\theta_0 = 0$ for all series—that is, the initial mean was zero. Five values of the γ were used, and the size of the shift was .5 (half the standard deviation of the noise). Single shifts occurred in half of the series; half of the shifts were positive and half were negative, with the signs randomized. The shift positions were randomized rectangularly, producing an equal probability of a shift occurring at any point.

Table 1 contains measures summarizing how the two techniques performed under various circumstances. The Hinich-Farley estimators are relatively straightforward to evaluate. The means of $\hat{\theta}$ are centered about the true value (zero), but the standard deviations are large, given the relative luxury of 100 data points. The jump heights, on the other hand, are not estimated well in general. The signs are right, but the estimated values are sensitive to γ (which shows up in the denominator of (4)), and the result is a tendency for the $\hat{\mu}$ to become larger as the likelihood of a change decreases. The standard errors of $\hat{\mu}$ also are related to γ. The basic problem involves sensitivity to the specification of the change rate. For small γ, the estimator is overly sensitive to shifts. For large γ, there is high likelihood of more than one jump leading to an underestimate of the shift size. When the expected number of shifts in the series corresponds to the actual number (i.e., $\gamma = .01$) the procedure works quite well, but is it not robust with respect to errors in specification of γ.

While the Hinich-Farley technique attempts to estimate jointly the level of the mean at the beginning of the series as well as the size of the change (if any), the Chernoff-Zacks technique starts at the end of a series, searches back for a change in mean, and then estimates the mean value of the series forward from the point at which such a change is found. In this case, the performance of both the detector and of the mean level estimator are important. Table 1 indicates that the detector is also sensitive to the probability of change. This is a case of the common statistical problems of two types of error where discrimination power is weak. When the probability of change is high the detector finds changes with approximately equal likelihood whether or not there is a change. When the probability of a change is small, fewer changes are detected in cases where none occurred, but the detector errs by failing to find changes when they do occur. Regardless of the change rate, though, there is large average error in the position where a change is detected and the variance of this error is also large.

Despite the weak performance of the detector, the estimator works reasonably well whether or not the detector is used.

1. *Using the detector.* The estimates when no shift occurred are somewhat inferior to the Hinich-Farley statistics, both in terms of bias and squared error. When the mean shifts, the Chernoff-Zacks technique is superior in terms of bias, although standard errors are similar. It should be noted, however, that much more computer time is required by the Chernoff-Zacks technique—especially when γ is small and the detector searches back over a large part of most series in search of a shift.

2. *Without the detector.* Surprisingly, the estimator performed as well or better in terms of variance as it did when the detector was used. Of course, a bias towards zero is introduced in the cases where change occurs by the fact that a considerable amount of pre-shift data is retained on average.

That the Hinich-Farley and Chernoff-Zacks techniques perform similarly is hardly surprising, because the structure in both cases involves manipulations of $\sum_i i X_i$ and the change rate (γ) as the basic components of the analysis. This structure is intuitively appealing because, given equal likelihood of a shift in any interval, later observations in a series contain more information as it becomes increasingly likely that a shift has occurred somewhere before that observation.

Despite desirable theoretical properties, though, the performance is rather disappointing, largely due to sensitivity to the change rate (γ), which is usually unknown and very difficult to estimate. When the expected time between shifts is equal to the record length (that is, $\gamma = .01$) the Hinich-Farley procedure out-performed the Chernoff-Zacks procedure in terms of bias and estimation error variance. Further,

TABLE 1

Measures of Performance of Alternative Estimating Techniques

Change Rate	.25			.1			.05			.01			.005		
Size of Change	0	.5	-.5	0	.5	-.5	0	.5	-.5	0	.5	-.5	0	.5	-.5
Hinich-Farley, estimators															
Initial mean	-.0268 (.1831)	-.06857 (.2745)	-.0224 (.2581)	-.01546 (.22201)	-.07199 (.31553)	.09673 (.32448)	.0019 (.1981)	-.0707 (.2401)	.0549 (.1714)	.02680 (.1837)	-.06578 (.27455)	-.0224 (.2583)	-.0135 (.2208)	-.0719 (.3012)	-.0967 (.3245)
mean change	.0022 (.0011)	.0193 (.0469)	.0193 (.0529)	.00212 (.0123)	.05465 (.00234)	-.0344 (.0021)	.0024 (.0682)	.1104 (.0674)	-.1018 (.0559)	.05527 (.09169)	.5366 (.37147)	-.4755 (.4174)	.0483 (.7853)	1.0929 (.8649)	-.6953 (.9071)
Chernoff-Zacks estimators															
Detector															
Change found (no. times)	48	21	27	25	16	19	12	11	10	7	20	23	0	5	2
Average error in change location	NA	17.0 (17.61)	22.92 (18.862)	NA	13.941 (14.149)	20 (19.656)	NA	14.272 (12.174)	15.70 (14.846)	NA	6.333 (7.5053)	6.250 (3.862)	NA	7.4 (6.58)	9.0 (4.242)
Average distance from change when detector tripped	NA	16.75 (15.52)	24.571 (15.99)	NA	23.353 (16.992)	22.00 (14.438)	NA	25.181 (14.655)	30.10 (21.957)	NA	17.667 (7.63)	22.00 (7.11)	NA	31.2 (14.2724)	33.5 (9.19)
Average position where change was found given no change	65.33 (23.32)	NA	NA	54.28 (22.8006)	NA	NA	50.916 (19.865)	NA	NA	50.2 (20.92)	NA	NA	0	NA	NA
Estimator															
Using detector	.0470 (.5196)	.4252 (.6251)	-.6759 (.6531)	.08425 (.39499)	.6467 (.49186)	-.4047 (.4171)	.0357 (.4213)	.5791 (.2427)	-.4677 (.2783)	.14090 (.1779)	.8629 (.4346)	-.9574 (.4688)	.0212 (.1881)	.6233 (.1806)	-.9149 (.3748)
Using all 100 observations	-.0469 (.2683)	-.4646 (.2467)	-.5205 (.2881)	.03327 (.2544)	.4872 (.4210)	-.4144 (.2950)	-.0104 (.2159)	.4656 (.3062)	-.3632 (.1930)	.0365 (.1359)	.4205 (.2271)	-.4726 (.2648)	.0212 (.1881)	.4496 (.3828)	-.3906 (.2518)
Number of series	50	22	28	50	22	28	50	29	21	50	22	28	50	22	28

NA—Not applicable

Standard deviations, where appropriate, are in the parentheses

techniques similar to the Hinich-Farley procedure have been shown to perform well when the error variance is unknown and must also be estimated [3]. Both procedures, however, are very sensitive to the specification of γ. Further, there is reason to believe that both perform poorly when multiple mean shifts in a record are likely [1]. This means that it is worthwhile to examine radically different approaches to the problem, like a filter theory approach discussed in the next section.

V. The Discrete Observation, Continuous Record Case

Concepts from digital filter theory, useful in continuous tracking of a mean when new observations constantly become available (real-time estimation), can be modified heuristically to deal with a fixed number of observations. Besides the real-time advantage, these procedures are computationally simple and are a promising approach for dealing with records which might involve several mean shifts.

The basic digital processing procedure involves some type of filter [5]. A filter is a time origin invariant linear operator which weights a time series $\{X_t\}$ by a sequence of constants $\{h_t\}$,

$$(15) \qquad Y_t = \sum_{t'} h_{t'} X_{t-t'} .$$

A filter is called low pass if it smooths the input, i.e., it attenuates the high frequencies and passes the mean and low frequencies. There are many types of filters, but the simplest low-pass filter is a simple moving average with equal coefficients, i.e.,

$$(16) \qquad Y_t = \frac{1}{2K+1} \sum_{s=-K}^{K} X_{t-s} .$$

By the linearity of the filter,

$$(17) \qquad \{Y_t\} = \{\Phi_t\} + \{\bar{N}_t\},$$

where the filtered mean has values

$$(18) \qquad \Phi_t = \frac{1}{2K+1} \sum_{s=-K}^{K} \theta_{t-s}$$

and the filtered noise has values

$$(19) \qquad \bar{N}_t = \frac{1}{2K+1} \sum_{s=-K}^{K} N_{t-s} .$$

The variance of the filtered noise $\{\bar{N}_t\}$ is

$$(20) \qquad E\bar{N}_t^2 = \frac{1}{2K+1} .$$

The filtered mean, from (16) and (18), is

$$
\begin{aligned}
(21) \qquad \Phi_t &= \theta & &\text{if } t < t^* - K, \\
&= \theta + \frac{t - t^* + K + 1}{2K+1} \mu & &\text{if } t^* - K \leq t < t^* + K, \\
&= \theta + \mu & &\text{if } t \geq t^* + K.
\end{aligned}
$$

Thus the filtered mean $\{\Phi_t\}$ is an upward ramp if the shift is positive and a downward ramp if the shift is negative. A detectability measure, the "signal-to-noise ratio", is

the size of shift divided by the standard deviation of the noise. If D_I denotes the detectability of the mean in $\{X_t\}$ and D_0 the detectability of the mean in $\{Y_t\}$,

(22) $$D_I = \mu$$

and

(23) $$D_0 = \mu(2K + 1)^{\frac{1}{2}}$$

from (20). The "gain" in filtering is defined as the ratio of detectabilities D_0/D_I. In this case, the gain is

$$D_0/D_I = (2K + 1)^{\frac{1}{2}}.$$

By choosing K sufficiently large, the variance of $\{\bar{N}_t\}$ can be made much less than the amount of shift μ, and a shift can be detected with probability approaching one. The trade-off is quite apparent; if K is made larger in order to reduce the variance of the estimated mean to improve the gain, detection is delayed.

The simple low pass filter in (16) is one of a variety of $2K + 1$ coefficient filters which attenuate the high frequencies; some produce an even greater filter gain for the shift detection. For example, gain is increased by filtering the process twice with the filter in (16), which reduces the noise variance from $\frac{2}{3}(2K + 1)^{-1}$. Thus, the gain is $D_0/D_I = (1.5)^{\frac{1}{2}}(2K + 1)^{\frac{1}{2}}$, about 23 per cent more than in the simple filter. However, it also requires more data to produce a filtered record of length n than does the simple rectangular filter. The choice of filters available is endless, although most common are rectangular, triangular and exponential. The rectangular filter is chosen here for purposes of convenience.

VI. Monte Carlo Analysis of the Rectangular Filter

Heuristic decision rules were used for detection by starting at the end of the series and searching backward for a shift point. Such rules, useful in developing and testing control-chart procedures [7], avoid the problems in specification of the shift rate (γ) mentioned earlier. If a shift is detected at t_0, observations of X_t preceding t_0 (for $t < t_0$) are discarded in estimating the current mean, since they should not yield any information about its new value. Artificial data also were used to test the filtering procedure. Eight sets of 100 series $\{Y_t\}$, each containing 90 observations smoothed rectangularly with $K = 5$, were constructed using methods similar to those described in §IV. The decision rules all involve identifying a shift at t_0 if p out of q observations $Y_{t_0}, Y_{t_0+1}, \cdots Y_{t_0+q-1}$ are greater than T or less than $-T$ (p and q are integers such that $1 \leqq p \leqq q$). Four rules were used for each of 2 levels of T ($T = .5$ and $T = .65$); $p = 3, q = 4; p = 4, q = 5; p = 5, q = 6;$ and $p = 6, q = 7$. If a shift is detected at t_0, the new mean is then estimated by a simple average of the input observations $\{X_t\}$ taken from t_0 to the end of the series. If the decision rule does not detect a shift as it goes backwards over the series, the estimate of the mean is given by an average of all values of the input series. This detection-estimation procedure is similar in spirit to control chart procedures common in quality control [8], and will be robust with respect to the shift rate γ provided it is not too high and if the variance σ^2 is stable.

Summary measures describing this analysis are shown in Table 2. With the looser decision rules (smaller p and q, and lower threshold value) the filter "detects" shifts in a high proportion of the series, regardless of whether one occurred. As p and q become larger and as the threshold is raised, fewer shifts are found in any series. Varying p, q, and T in any application will thus allow explicit trade-off between both

TABLE 2

Summary Measures for Filter-Estimation Procedure

DECISION RULE	Change found	Position where change was found	Absolute difference between actual change position and where filter detected change	Estimated value of series after filter found a change	Number of observations
3 of 4 > T or < −T					
T = .5 Δ = 0	34	40.7 (28.483)	NA	.0354 (.1674)	49
Δ = +.5	30	NA	20.8 (22.0819)	.3613 (.2539)	31
Δ = −.5	19	NA	19.526 (26.435)	−.39021 (.1939)	20
T = .65 Δ = 0	17	51.82 (28.76)	NA	−.0147 (.0736)	51
Δ = +.5	23	NA	15.52 (21.69)	.4810 (.2830)	27
Δ = −.5	18	NA	9.83 (11.92)	−.4735 (.2393)	22
4 of 5 > T or < −T					
T = .5 Δ = 0	34	44.76 (29.09)	NA	.0392 (.1544)	49
Δ = +.5	27	NA	17.11 (18.55)	.4547 (.1661)	29
Δ = −.5	20	NA	16.6 (24.60)	−.4022 (.1798)	22
T = .65 Δ = 0	9	52.78 (23.11)	NA	−.00342 (.1066)	51
Δ = +.5	23	NA	15.000 (14.67)	.4945 (.2995)	27
Δ = −.5	17	NA	18.42 (15.81)	−.3945 (.1829)	22
5 of 6 > T or < −T					
T = .5 Δ = 0	26	44.88 (26.03)	NA	.0384 (.2459)	51
Δ = +.5	26	NA	20.38 (22.81)	.4101 (.1945)	27
Δ = −.5	20	NA	18.80 (20.98)	−.4030 (.2131)	22
T = .65 Δ = 0	5	61.4 (25.67)	NA	.00676 (.1174)	51
Δ = +.5	22	NA	15.54 (16.33)	.4802 (.2375)	27
Δ = −.5	16	NA	13.06 (20.47)	−.3944 (.1929)	22
6 of 7 > T or < −T					
T = .5 Δ = 0	21	49.76 (25.44)	NA	.0325 (.08011)	51
Δ = +.5	25	NA	16.24 (17.98)	.4803 (.2403)	27
Δ = −.5	17	NA	18.00 (25.09)	−.4204 (.2142)	22
T = .65 Δ = 0	3	47.66 (23.03)	NA	.0156 (.1195)	51
Δ = +.5	15	NA	22.6 (16.57)	.4496 (.2505)	27
Δ = −.5	15	NA	18.87 (12.96)	−.4601 (.2327)	22

types of errors as well as the time required for detecting a change. In contrast to the results in §IV, though, there is increasing discriminatory power (especially for the higher threshold), indicating that the filter techniques are preferable to those described in the last section. Errors in locating the position of shifts are comparable over thresholds and decision rules, indicating again that precise estimation of the point of shift is not possible when the error variance is large and the time uncertainty of the shift is great. The estimated mean values using the filter are less biased than those presented in §IV and their standard errors are smaller. As with the earlier results, however, there is an estimation bias towards zero when a shift does occur caused by asymmetric error in the filter, which tends to pass by a change a considerable distance before finding it.

VII. Discussion

Some ways have been examined to systematically track time series which potentially contain small stochastic mean shifts as well as random measurement errors. A "small shift" is one which is small relative to measurement error. Such problems may arise in an information system which monitors large numbers of data series, such as competitors' prices and advertising behavior, reported by representatives in the field or by other data collecting sources such as store audits or consumer panels. Three approaches were tested with artificial data using mean shifts which were rather small—that is, mean shifts which were half the magnitude of random measurement error variance. Two techniques (one developed by the authors and the other by Chernoff and Zacks) involved formal estimation under the assumption that there was at most one discrete jump in a data record of fixed length of the type often stored in an information system. Both techniques performed reasonably when the rate of shift occurrence was known, but both techniques are very sensitive to prior specification of the rate at which shifts occur in terms of both classes of errors—that is, missing shifts which occur and identifying "shifts" which do not occur. Knowing the shift rate precisely and knowing that more than one shift in a record is extremely unlikely are two very severe restrictions for many applications.

A simpler filter technique was tested similarly with more promising results in terms of avoiding both classes of errors. The filter approach involved first smoothing the series and then implementing *ad hoc* decision rules based on consecutive occurrences of smoothed values falling outside a predetermined range about the running mean. These *ad hoc* procedures are unsatisfactory from the point of view of certain theoretical considerations, although it is important to note that the Chernoff-Zacks procedure had *ad hoc* elements, despite its over-all complexity and sophistication. This data filtering route appears to be the most promising way now available to approach the development of tracking procedures for information systems when the change rate is unknown. The superiority is due to flexibility, insensitivity to rates at which shifts occur over time, ability to be developed towards the possibility of multiple shifts in a record and clear statement of trade-offs of various types of errors.

References

1. CHERNOFF, H. AND ZACKS, S., "Estimating the Current Mean of a Normal Distribution which is Subjected to Changes in Time," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 999–1018.
2. CHOW, G., "Tests of Equality Between Two Sets of Coefficients in Two Linear Regressions," *Econometrica*, Vol. 28 (1960), pp. 591–605.
3. FARLEY, J. AND HINICH, M., "A Test for a Shifting Slope Coefficient in a Linear Model," to appear in *The Journal of the American Statistical Association*, forthcoming.
4. —— AND ——, "Tracking Marketing Parameters in Random Noise," *Marketing and Economic Development*, P. D. Bennet, ed., American Marketing Association, 1965, pp. 365–369.

5. GRENANDER, U. AND ROSENBLATT, M., *Statistical Analysis of Stationary Time Series*, John Wiley, New York, 1957.

6. HINICH, M. AND FARLEY, J. "Theory and Application of an Estimation Model for Time Series with Nonstationary Means," *Management Science*, Vol. 12 (1966), pp. 648–658.

7. KRONMAL, R., "The Evaluation of a Pseudo-Random Normal Generation," *Journal of the Association for Computing Machines*, Vol. 11 (1964), pp. 357–363.

8. MUSTAFI, C. K., "Inference Problems about Parameters which are Subjected to Changes over Time," *Ann. Math. Statist.*, Vol. 39 (1968), pp. 840–854.

9. ROBERTS, S. W., "A Comparison of Some Control Chart Procedures," *Technometrics*, Vol. 8 (1966), pp. 411–430.