MANAGEMENT SCIENCE
Vol. 12, No. 9, May, 1966
Printed in U.S.A.

THEORY AND APPLICATION OF AN ESTIMATION MODEL FOR TIME SERIES WITH NONSTATIONARY MEANS*†

MELVIN HINICH[1] AND JOHN U. FARLEY[2]

Time series models of a complex nature, such as consumer brand switching analyses, have required assumptions of parameter stability because statistical models were not available to deal with parameter change. A model is developed here to estimate a stepwise change in the mean process of a Gaussian time series. Estimators which are small-sample efficient in a special sense are presented, along with examples and suggested applications of the method to brand switching problems.

I. Introduction

The problem of studying a parameter over time where there is random error and systematic stochastic parameter change has posed significant problems in such areas as the analysis of consumer brand switching. Switching behavior has been examined in stochastic frameworks both explicitly and implicitly, and complex models have been developed in Markovian and learning frameworks to analyze sequences of consumer purchases (Farley and Kuehn [4]). These models have used aggregate parameter estimation techniques in static frameworks—that is where underlying process parameters are assumed stable over a period of time. The stability assumptions have been necessary because no satisfactory inference techniques have been available to deal with parameter change.

Researchers have been aware of stability problems, and some work has been done on magnitudes of difficulties posed by instability (Frank [5]). Some more recent models have explicitly incorporated a "mix" approach which attempts to segregate effects of temporal instability from tendencies to form a stable mix of purchases (as with cereals for different members of a family), (Kuehn [7] and Rohloff [10]). Suppose, for example, the brand sequence history for a product reported by three members of a consumer panel in a period of time looked like this:

$$ABABABABAB \qquad (1)$$

$$AAAAABBBBB \qquad (2)$$

$$ABBAABABBAAB \qquad (3)$$

Sequence (1) probably represents a household which divides its purchases evenly between two brands—perhaps between two smokers in a family who favor different brands of cigarettes and who smoke an equal amount. The mix occurs be-

* Received March 1965.

† Much of this work was performed at the Carnegie Institute of Technology as part of the Management Science Research Group, ONR Contract 760(24). Other parts were done at Stanford University and the Hudson Laboratories of Columbia University.

[1] Hudson Laboratories of Columbia University and Carnegic Institute of Technology.

[2] Carnegie Institute of Technology.

cause panel data are gathered by family rather than by user. The second sequence probably shows a *bona fide* switch from brand A to brand B. However, sequence (3) is not so easy to interpret directly, because there may be mixing, switching or both going on. Unfortunately, many actual observed sequences more closely resemble (3) than (1) or (2), and this poses the instability problem discussed earlier. Further, both switching and mixing behavior may involve important random components because of such common factors as stock-outs or occasional purchases in unfamiliar places.

These types of problems have held up model building aimed at testing the effects on brand purchase sequences of various types of changes in merchandising variables—advertising, in-store merchandising, etc. Thus a useful step in the development of brand switching models is to develop statistical techniques to identify basic changes in underlying probability distributions generating discrete panel observations. As is often the case, a lead into this problem came from a seemingly unrelated field—tracking an unknown time-varying parameter in random noise. Part II of this paper lays out an estimation routine for a special class of such problems, while Part III discusses potential applications of the model to problems involving consumer switching behavior.

II. A Model for Tracking a Time-Varying Parameter

A stochastic process, $X(t) = \theta(t) + N(t)$, may generate sequence data where $\theta(t)$ is the mean value process and $N(t)$ is Gaussian noise with zero mean and a known covariance function. The mean process is assumed to behave as a step function, so for $i = 0, 1, 2, \cdots$

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

in an interval where the t_i^* (the times when θ changes) follow a Poisson process with a known rate parameter γ per time unit, where the time unit is defined appropriately for the problem. $\theta(t)$ is thus the expected value of $X(t)$, and it may change from one constant value to another at random times t_i^*. In any T successive time units, the probability of k jumps is $(\gamma T)^k/k!\, e^{-\gamma T}$, and the mean time between jumps is γ^{-1} time units. Figure 1 shows how the step function $\theta(t)$ might behave, and this work will deal with the case in which no more than one change occurs within the total sampling period.

Discrete samples are drawn during non-overlapping periods (called records) n time units long. Each record consists of n successive observations of $X(t)$ where the observations are taken one time unit apart. If n, the record length, is smaller than γ^{-1}, the mean time between jumps, the possibility of two or more jumps in a record can be disregarded in the manner normal for dealing with Poisson variables. Given these conditions, a small-sample efficient estimator of $\theta(t)$ can be developed as a function of the sequence of observations.[3]

[3] This work was motivated by an investigation in adaptive ocean bottom ranging by a type of sonar. Let $\theta(t)$ denote the ping distance from the sonar to the bottom. Due to rough bottom scattering and random perturbations of the medium, $\theta(t)$ changes from ping to ping. Assume additive Gaussian noise with a known spectrum, the problem is to remove the systematic bias introduced by a change in the medium.

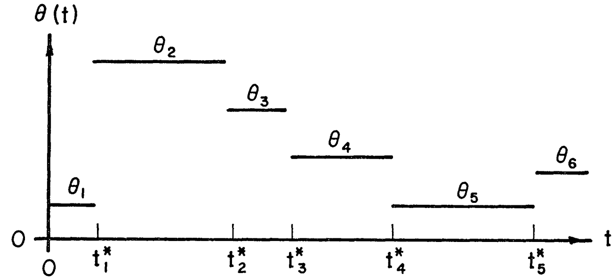

FIG. 1. Randomly shifting mean parameter

Chernoff and Zacks [2] used Bayesian methods to deal with this problem. Since their estimator is a complicated function of the observations they present a simpler ad-hoc procedure for at most one change, and test it by Monte Carlo computations. Page [9] deals with the problem of testing for a change in the mean using a test statistic similar to the one developed here. Our variances have not been compared with those of the Chernoff-Zacks estimators.

In practice, n successive time points, $t_1 < \cdots < t_n$ are observed so that $t_{i+1} - t_i$ constitutes the natural time unit in the problem. The sample points form 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. Since the parameter is a step function,

$$\theta(t) = \begin{cases} \theta_1 & \text{if} \quad t < t^* \\ \theta_2 & \text{if} \quad t > t^* \end{cases} \tag{2}$$

where t^* is the time of a random jump in the mean from θ_1 to θ_2. $N(t)$ is Gaussian noise with known covariance $\sigma_{ij} = E[N(t_i)N(t_j)]$.

It might happen, for example, that $t^* > t_n$ and thus there is no jump in the mean during the sampling period. If $t_0 < t^* < t_1$, then a jump occurs in the start of the record. Since t^* is purely random, it can fall anywhere within the sample with equal probability for any interval. γ, the rate of occurrence of jumps, is the probability for each of the n events $t_i < t^* < t_{i+1}$ for $i = 0, \cdots, n - 1$. Thus the probability of no jump in the sampling interval is $1 - n\gamma$. If $\gamma^{-1} \gg n$, it is likely that the mean does not change. We can express the probability of each possible state of nature by defining

$$S_j\theta = \{\theta_{j-n+2}, \theta_{j-n+3}, \cdots, \theta_{j+1}\}, \tag{3}$$

where for any integer k

$$\theta_k = \begin{cases} \theta_1 & \text{if} \quad k \leqq 1 \\ \theta_2 & \text{if} \quad k \geqq 2 \end{cases} \tag{4}$$

and $j = 0, \cdots, n$. Thus for each $j = 1, \cdots, n$

(5) $\quad E\{X_1, \cdots, X_n\} = S_j\theta$ with probability γ and

(6) $\quad E\{X_1, \cdots, X_n\} = S_0\theta = \{\theta_1, \cdots, \theta_1\}$ with probability $1 - n\gamma$.

For example, if $n = 5$,

$$\begin{aligned}
S_1\theta &= \{\theta_1, \theta_1, \theta_1, \theta_1, \theta_2\} && \text{with probability } \gamma \\
S_2\theta &= \{\theta_1, \theta_1, \theta_1, \theta_2, \theta_2\} && \text{with probability } \gamma \\
S_3\theta &= \{\theta_1, \theta_1, \theta_2, \theta_2, \theta_2\} && \text{with probability } \gamma \\
S_4\theta &= \{\theta_1, \theta_2, \theta_2, \theta_2, \theta_2\} && \text{with probability } \gamma \\
S_5\theta &= \{\theta_2, \theta_2, \theta_2, \theta_2, \theta_2\} && \text{with probability } \gamma \\
S_0\theta &= \{\theta_1, \theta_1, \theta_1, \theta_1, \theta_1\} && \text{with probability } (1\text{–}5\gamma).
\end{aligned}$$

Thus 5γ is the probability of a change occurring during the observation period and the conditional probability distribution as to when the change occurs is uniform over the period.

From (3), (4), (5), and (6), the vector

$$X = \begin{cases} N + S_1\theta & \text{with probability } \gamma \\ N + S_2\theta & \text{with probability } \gamma \\ \quad\vdots & \\ N + S_n\theta & \text{with probability } \gamma \\ N + S_0\theta & \text{with probability } 1\text{–}n\gamma, \end{cases} \tag{7}$$

where $N = \{N_1, \cdots, N_n\}$ has a multivariate normal distribution with mean zero and covariance matrix $\Sigma = (\sigma_{ij})$. If $f(x \mid \theta)$ is the density function of X given the parameter vector $\theta = \{\theta_1, \theta_2\}$, then

$$f(x \mid \theta) = (1 - n\gamma)n(x \mid S_0\theta, \Sigma) + \gamma \textstyle\sum_{j=1}^{n} n(x \mid S_j\theta, \Sigma) \tag{8}$$

where

$$n(x \mid \xi, \Sigma) = (2\pi)^{-n/2} \mid \Sigma \mid^{-1|2} \exp\left[-\tfrac{1}{2}(x - \xi)\Sigma^{-1}(x - \xi)'\right] \tag{9}$$

is a n-dimensional normal density with mean vector $\xi = \{\xi_1, \cdots, \xi_n\}$ and covariance matrix Σ. The density function $f(x \mid \theta)$, a convex combination of multivariate normal densities but not in general multivariate normal itself, is approximated by expansion in Taylor series in the neighborhood of zero.

If we assume that

$$\theta_1^2 + \theta_2^2 \ll E\{N^2(t)\},$$

the noise can be normalized so that $E\{N^2(t)\} = 1$, and thus $\sigma_{ii} = 1$ for all $i = 1, \cdots, n$.

The Information Matrix

The Cramér-Rao inequality (Cramér [3]) gives a bound for the variance of unbiased estimators in terms of the inverse of the "information matrix." Given a parameter vector $\phi = \{\phi_1, \cdots, \phi_n\}$ and a random variable X with density function $p(x \mid \phi)$, the information matrix for p and ϕ, $I(\phi)$, is defined by:

$$I(\phi) = E_\phi[\partial \log p(X \mid \phi)/\partial\phi]'[\partial \log p(X \mid \phi)/\partial\phi] \tag{10}$$

where $I(\phi)$ is an $n \times n$ matrix,

$$[\partial \log p(x \mid \phi)/\partial \phi] = \{\partial \log p(x \mid \phi)/\partial \phi_1, \cdots, \partial \log p(x \mid \phi)/\partial \phi_n\},$$

and E_ϕ is the matrix of expectations with respect to $p(x \mid \phi)$.

If $E_\phi T(X) = \phi$, $T(X)$ is an unbiased estimator of ϕ, a matrix

$$K_T = E_\phi[T(X) - E_\phi T(X)]'[T(X) - E_\phi T(X)]$$

can be defined as the covariance of T, so

$$K_T = E_\phi [T(X) - \phi]'[T(X) - \phi]. \tag{11}$$

The Cramér-Rao inequality shows if T is an unbiased estimator of ϕ, then $K_T \geqq I^{-1}(\phi)$, where the inequality between matrices means that for any n-dimensional vector v

$$vK_Tv' \geqq vI^{-1}(\phi)v'. \tag{12}$$

$vI^{-1}(\phi)v'$ is thus the lower bound for any unbiased estimator of the parameter $\phi v' = \sum_{i=1}^n v_i\phi_i$.

The standard definition of efficiency for an unbiased estimator, T, is that its covariance matrix $K_T = I^{-1}(\phi)$. However, our criterion for unbiased estimators uses a weaker concept called efficiency near ϕ_0 defined so that for every vector v,

$$\lim_{\phi \to \phi_0} vI^{-1}(\phi)v'/vK_Tv' = 1. \tag{13}$$

The information matrix $I(\theta)$ with the ij^{th} element is defined as:

$$\begin{aligned} I_{ij}(\theta) &= E_\theta\left[\frac{\partial \log f(X \mid \theta)}{\partial \theta_i} \frac{\partial \log f(X \mid \theta)}{\partial \theta_j}\right] \\ &= \int \frac{\partial f(x \mid \theta)}{\partial \theta_i} \frac{\partial f(x \mid \theta)}{\partial \theta_j} \frac{1}{f(x \mid \theta)}\, dx_1, \cdots, dx_n \\ &= E_0\left\{\left[\frac{1}{f(X \mid 0)} \frac{\partial f(X \mid \theta)}{\partial \theta_i}\right]\left[\frac{1}{f(X \mid 0)} \frac{\partial f(X \mid \theta)}{\partial \theta_j}\right]\frac{f(X \mid 0)}{f(X \mid \theta)}\right\}. \end{aligned} \tag{14}$$

By expanding in Taylor series about $\xi = 0$ we have, from (9) with

$$\| \xi \|^2 = \sum_{i=1}^n \xi_i^2,$$

the likelihood ratio:

$$\begin{aligned} n(x \mid \xi, \Sigma)/n(x \mid 0, \Sigma) = 1 + x\Sigma^{-1}\xi' + \tfrac{1}{2}\xi\, \Sigma^{-1}(x'x - \Sigma)\Sigma^{-1}\xi' \\ + \| \xi \|^3 K^*(x, \xi) \end{aligned} \tag{15}$$

where $| K^*(x, \xi)| \leqq de^{t_1^*|x_1| + \cdots + t_n^*|x_n|}$ for some $t_i^* > 0$ and $d > 0$. Thus $E_0[K^*(x, \xi)]^r$ exists and is bounded by some number independent of ξ for each $r \geqq 0$.

Setting $\xi = S_j\theta$ in (15) and summing, we have from (8),

$$
\begin{aligned}
f(x \mid \theta)/f(x \mid 0) = 1 &+ (1 - n\gamma)\theta_1(x\Sigma^{-1} 1') \\
&+ \gamma x\Sigma^{-1}(\textstyle\sum_{j=1}^{n} S_j\theta)' \\
&+ \tfrac{1}{2}\theta_1^2[(x\Sigma^{-1}1')^2 - 1\Sigma^{-1}1'] \\
&+ \gamma 0(\| \theta \|^2)K(x, \theta)
\end{aligned} \tag{16}
$$

where 1 is the n-dimensional unit row vector. From (6)

$$S_0\theta = \theta_1 1, \text{ so}$$

$E_0[K(x, \theta)]^r$ exists and is bounded by some number independent of θ for each $r \geqq 0$. Moreover, the coefficient of $K(x, \theta)$ in (16) does not involve x. Using (5), and defining a as the vector $\{1, 2, \cdots, n\}$

$$\textstyle\sum_{j=1}^{n} S_j\theta = n\theta_1 1 + (\theta_2 - \theta_1)a. \tag{17}$$

Using (17) in (16) and setting $\mu = (\theta_2 - \theta_1)$ (the size of the jump of the mean),

$$
\begin{aligned}
f(x \mid \theta_1, \mu)/f(x \mid 0) = 1 &+ (x\Sigma^{-1}1')\theta_1 + (x\Sigma^{-1}a')\mu \\
&+ \tfrac{1}{2}\theta_1^2[(x\Sigma^{-1}1')^2 - 1\Sigma^{-1}1'] \\
&+ \gamma 0(\| \theta \|^2)K(x, \theta).
\end{aligned} \tag{18}
$$

The results in equation (18) can be used to estimate the parameter vector $\{\theta_1, \mu\}$. If $\phi^* = \phi A$, where A is an $n \times n$ nonsingular matrix, then by applying the chain rule to (10),

$$I(\phi^*) = A^{-1}I(\phi)(A^{-1})' \tag{19}$$

where $I(\phi^*)$ is the information matrix of ϕ^*. Thus,

$$I^{-1}(\phi^*) = A'I^{-1}(\phi)A \tag{20}$$

Furthermore, given an unbiased estimator, T,

$$T^*(X) = T(X)A$$

is an unbiased estimator of ϕ^* with covariance

$$K_T^* = A'K_TA. \tag{21}$$

Thus if T is efficient, from (20) and (21) T^* is also efficient. Now

$$\{\theta_1, \theta_2\} = \{\theta_1, \mu\}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and thus if $\hat{\theta}_1$ and $\hat{\mu}$ are efficient estimators of θ_1 and μ, then $(\hat{\theta}_1 + \hat{\mu})$ is an efficient estimator of θ_2.

$I(\theta_1, \mu)$ can be approximated using (18) in (14) with a change of parameters. The Taylor expansion argument which gave (18) also yields

$$
\begin{aligned}
(1/f(x \mid 0))(\partial f(x \mid \theta_1, \mu)/\partial\theta_1) &= x\Sigma^{-1}1' + 0(\theta_1) \\
(1/f(x \mid 0))(\partial f(x \mid \theta_1, \mu)/\partial\mu) &= \gamma x\Sigma^{-1}a' + \gamma 0(\| \theta \|)
\end{aligned} \tag{22}
$$

where again $a = \{1, 2, \cdots, n\}$. The proof of Lemma 3 in Hinich [6] is used to bound $f(x \mid 0)/f(x \mid \theta)$ in (14), and using (22) in (14),
Lemma 1:

$$(23) \qquad I(\theta_1, \mu) = \begin{bmatrix} 1\Sigma^{-1}1' & \gamma 1\Sigma^{-1}a' \\ \gamma 1\Sigma^{-1}a' & \gamma^2 a\Sigma^{-1}a' \end{bmatrix} + \begin{bmatrix} 0(\theta_1) & \gamma 0(\| \theta \|) \\ \gamma 0(\| \theta \|) & \gamma^2 0(\| \theta \|) \end{bmatrix}$$

The main part of the information matrix is nonsingular since by the Schwarz Inequality,

$$(1\Sigma^{-1}a')^2 < (1\Sigma^{-1}1')(a\Sigma^{-1}a').$$

The inverse of (23) yields
Lemma 2:

$$I^{-1}(\theta_1, \mu) = ((1\Sigma^{-1}1')(a\Sigma^{-1}a') - (1\Sigma^{-1}a')^2)^{-1} \begin{bmatrix} a\Sigma^{-1}a' & -\gamma^{-1}1\Sigma^{-1}a' \\ -\gamma^{-1}1\Sigma^{-1}a' & \gamma^{-2}1\Sigma^{-1}1' \end{bmatrix} + \begin{bmatrix} 0(\| \theta \|) & \gamma^{-1}0(\| \theta \|) \\ \gamma^{-1}0(\| \theta \|) & \gamma^{-2}0(\| \theta \|) \end{bmatrix}$$

An example is presented in the Appendix.

Estimation of θ_1 and μ

Efficient estimators of θ_1 and μ are linear combinations of $X\Sigma^{-1}1'$ and $\gamma X\Sigma^{-1}a'$ where $a = \{1, 2, \cdots, n\}$. In order to demonstrate efficiency we need the mean, M, and covariance matrix, K, of $\{X\Sigma^{-1}1', \gamma X\Sigma^{-1}a'\}$.

From (8) and (17),

$$(24) \qquad \begin{aligned} E_\theta\{X\Sigma^{-1}1'\} &= (1 - n\gamma)\theta_1(1\Sigma^{-1}1') + \gamma \textstyle\sum_{j=1}^{n} (S_j\theta)\, \Sigma^{-1}1' \\ &= (1\Sigma^{-1}1')\theta_1 + \gamma(1\Sigma^{-1}a')\mu \end{aligned}$$

and similarly

$$(25) \qquad E_\theta\{X\Sigma^{-1}a'\} = (1\Sigma^{-1}a')\theta_1 + \gamma(a\Sigma^{-1}a')\mu.$$

From (24) and (25),

$$(26) \qquad M = E_\theta\{X\Sigma^{-1}1', \gamma X\Sigma^{-1}a'\} = \{\theta_1, \mu\}\,[I^*(\theta_1, \mu)]$$

where $I^*(\theta_1, \mu)$ is the main part of the information matrix $I(\theta_1, \mu)$ in (23).

From (8), the second moment is

$$(27) \qquad \begin{aligned} E_\theta\{1\Sigma^{-1}X'X\Sigma^{-1}1'\} &= 1\Sigma^{-1}1' + (1 - n\gamma)\theta_1^2(1\Sigma^{-1}1')^2 \\ &\quad + \gamma \textstyle\sum_{j=1}^{n}[1\Sigma^{-1}(S_j\theta)']^2 = 1\Sigma^{-1}1' + 0(\| \theta \|) \\ K &= E_\theta \begin{bmatrix} 1\Sigma^{-1}X'X\Sigma^{-1}1' & a\Sigma^{-1}XX\Sigma^{-1}1' \\ 1\Sigma^{-1}X'X\Sigma^{-1}a' & a\Sigma^{-1}XX\Sigma^{-1}a' \end{bmatrix} - M'M \end{aligned}$$

so from (26) and (27)

$$(28) \qquad K = \begin{bmatrix} 1\Sigma^{-1}1' & 1\Sigma^{-1}a' \\ 1\Sigma^{-1}a' & a\Sigma^{-1}a' \end{bmatrix} + \begin{bmatrix} 0(\| \theta \|) & \gamma 0(\| \theta \|) \\ \gamma 0(\| \theta \|) & \gamma^2 0(\| \theta \|) \end{bmatrix}$$

The estimators $\hat{\theta}$ and $\hat{\mu}$ are

$$\hat{\theta}_1 = [(a\Sigma^{-1}a')(X\Sigma^{-1}1') - (1\Sigma^{-1}a')(X\Sigma^{-1}a')]/[(1\Sigma^{-1}1')(a\Sigma^{-1}a') - (1\Sigma^{-1}a')^2]$$

(29)

$$\hat{\mu} = \gamma^{-1}[(1\Sigma^{-1}1')(X\Sigma^{-1}a') - (1\Sigma^{-1}a')(X\Sigma^{-1}1')]/[(1\Sigma^{-1}1')(a\Sigma^{-1}a') - (1\Sigma^{-1}a')^2].$$

From (24) and (25)

$$E_\theta\,\hat{\theta}_1 = \theta_1 \quad \text{and} \quad E_\theta\hat{\mu} = \mu$$

so $\hat{\theta}_1$ and $\hat{\mu}$ are unbiased estimators.
Rewriting (29),

$$\{\hat{\theta}_1, \hat{\mu}\} = \{X\Sigma^{-1}1', \gamma X\Sigma^{-1}a'\}[I^*(\theta_1, \mu)]^{-1}, \tag{30}$$

and using (28), (29) and Lemma 2,

(31)

$$K_T = [I^*(\theta_1, \mu)]^{-1}K[I^*(\theta_1, \mu)]^{-1} = I^{-1}(\theta_1, \mu) + \begin{bmatrix} 0(\|\theta\|) & \gamma^{-1}0(\|\theta\|) \\ \gamma^{-1}0(\|\theta\|) & \gamma^{-2}0(\|\theta\|) \end{bmatrix}$$

For any vector v, then,

$$\lim_{\|\theta\|\to 0} vI^{-1}(\theta_1, \mu)v'/vK_Tv' = 1,$$

and $\hat{\theta}_1$ and $\hat{\mu}$ are thus efficient near zero for θ_1 and μ.

As an example, suppose the covariance matrix is the identity matrix. Then from (A1) and (A2) we can write (29) as

(32)

$$\hat{\theta}_1 = 2/n(n-1)[(2n+1)\Sigma_jX_j - 3\Sigma_jjX_j] \quad \text{and}$$

$$\hat{\mu} = 6\gamma^{-1}/n(n-1)[(2/n+1)\Sigma_jjX_j - \Sigma_jX_j],$$

where the simple sums are self explanatory and the Σ_jjX_j are weighted terms related to exponential smoothing techniques.

An Alternative Formulation

The property of efficiency of the estimator $\hat{\mu}$ results from the assumption that μ is small compared to the variance of the noise, not from any assumption about θ_1. By expanding $f(x \mid \theta)$ in powers of μ and bounding the non-linear terms, the information matrix can be approximated for θ_1 and μ. The efficient estimators of θ_1 and μ (for small values of γ and μ) are the same linear functions of $X\Sigma^{-1}1'$ and $X\Sigma^{-1}a'$ as were developed in the immediately preceding section. The approximation approach produces estimators which are easy to calculate and which contain most of the information about the parameters in a range.

III. Applications to Brand Switching Problems

The techniques developed in the last section can be used to study parameter changes in hypothesized processes generating consumer panel data. We might, for example, take a sequence of purchases and divide it into a number of non-overlapping subsets. (When to sample and how to divide the sequences are important issues; while the techniques described here are useful in handling only the statistical problems, it will be shown later that some clues are available from the structure of the problem, about when to sample.) We might be interested, for example, in changes in the probability of a family's buying a particular brand, brand A. In each sub-sequence a variable Y_i can be defined as 1 if A is purchased at the i^{th} observation and as 0 otherwise. The statistic

$$\sum_{i=1}^{r} Y_i/r,$$

where the summation is over the r observations in a given subset, can then be viewed as an observation of the variable $X(t)$ needed for the analysis. Further, a binominal approximation argument leads to normal error associated with values of $X(t)$ constructed in this way, for samples small enough to be manageable. An arscine transformation (Brownlee [1]) provides, under the model, the required known variance-covariance matrix with $1/r$ on the diagonal and zeros elsewhere.

The inference model might then be used in two types of closely-related situations:

1) Investigation of various types of cross-sectional stability conditions. Tests can be developed, for example, to see whether $\mu = 0$—that is whether there is no change in the mean purchase probability during some arbitrarily defined time period. This could lead to analyses of whether the mix of brands bought over a long period is stable with respect to shorter sub-periods. The same techniques might be even more useful in the study of store switching, where much more data are available for a short time period than for brand switching in any single product class. For these purposes, sampling might be on some natural basis like seasons, on the basis of some arbitrary time divisions such as three or six-month periods, or on the basis of some external data like the date of a change of address.

2) Another application for this model is to test "before and after conditions" where the test period is divided at a point in time when some type of shock has been introduced into the system—perhaps a special promotional campaign was launched. Here, we may even have some good idea about special properties of μ. For instance, some influences may be subject to deterioration patterns, as with advertising effects where the effect of a given campaign may be monotone decreasing in t. Here $\theta_1(t)$ is the influence of a base rate of advertising, $\theta_2(t)$ is the influence of a new level of expenditure, and μ is the effect attributable to a given campaign. One-shot advertising campaigns might thus be viewed as producing a value for θ_2 which declines exponentially in t, (Kuehn [7]) while a pulsed type of advertising strategy may produce a series of exponentials.

Ultimately, this type of inference model may help improve methods for building and testing models of consumer brand choice behavior or conceptually similar processes like inventory systems with demand rates subject to random shocks. Inference models such as these are important steps in developing empirical tests for more complex models—an area where statistical inference has to some extent lagged behind development of modeling techniques (Farley and Kuehn [4]).

Appendix: An Example for Lemma 2 Where $N(t)$ is "White Noise"

Suppose the noise is white in the sense that the errors are independent and homoscedastistic. Given the normalization $E\{N^2(t)\} = 1$, Σ is $n \times n$ identity matrix because

$$\sigma_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$

Then $\Sigma^{-1} = I$ and

$$\text{(A1)} \qquad \begin{aligned} 1\Sigma^{-1}1' &= n \\ 1\Sigma^{-1}a' &= \textstyle\sum_{j=1}^{n} j = n(n+1)/2 \\ a\Sigma^{-1}a' &= \textstyle\sum_{j=1}^{n} j^2 = n(n+1)(2n+1)/6 \end{aligned}$$

and

$$\text{(A2)} \qquad X\Sigma^{-1}1' = \Sigma X_j \qquad X\Sigma^{-1}a' = \Sigma jX_j$$

Using (A1), (A2) and Lemma 2,

$$\text{(A3)} \quad I^{-1}(\theta_1, \mu) = \frac{2}{n(n-1)} \begin{bmatrix} 2n+1 & -3\gamma^{-1} \\ -3\gamma^{-1} & \dfrac{6\gamma^{-2}}{n+1} \end{bmatrix} + \begin{bmatrix} 0(\| \theta \|) & \gamma^{-1}0(\| \theta \|) \\ \gamma^{-1}(\| \theta \|) & \gamma^{-2}0(\| \theta \|) \end{bmatrix}$$

is the lower bound for the covariance of unbiased estimators of θ_1 and μ.

To carry this example further, suppose that γ^{-1}, the average distance between jumps is $4n$, and thus $\gamma n = \frac{1}{4}$ which is the probability that there is a jump somewhere in the record of duration nS time units. From (A3),

$$I^{-1}(\theta_1, \mu) = \frac{2}{n-1} \begin{bmatrix} 2 + \dfrac{1}{n} & -12 \\ -12 & 96\dfrac{n}{n+1} \end{bmatrix} + \cdots$$

The error terms for I^{-1} are not precise enough to show the power of the method when γ is a function of $1/n$. However, more detailed analysis on the higher order terms in the Taylor series expansion with respect to n shows that the linear term contains most of the information about the parameters.

References

1. Brownlee, K., *Statistical Theory and Methodology in Science and Engineering*, New York: John Wiley & Co.
2. 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, December 1964.
3. Cramer, H., *Mathematical Methods of Statistics*, Princeton, N. J.: Princeton University Press, 1946.

4. FARLEY, J. AND KUEHN, A., "Stochastic Models of Brand Switching," in George Schwartz, ed., *Science in Marketing*, New York: John Wiley and Co., 1964.
5. FRANK, R., "Brand Choice as a Probability Process," *Journal of Business*, January 1962.
6. HINICH, M., "Large-Sample Estimation of an Unknown Waveform Which is Randomly Repeating in Gaussian Noise," *Ann. Math. Statist.*, Vol. 36, April 1965.
7. KUEHN, A., "How Advertising Depends on Other Marketing Factors," *Journal of Advertising Research*, December 1962.
8. ——, "Effectiveness of Consumer Deals and Advertising," in S. Greyser, ed., *Toward Scientific Marketing*, American Marketing Association, 1963.
9. PAGE, E., "A Test for A Change in A Parameter Occurring at an Unknown Point," *Biometrika*, Vol. 42, 1955.
10. ROHLOFF, A. "New Ways to Analyze Brand-to-Brand Competition," in S. Greyser, ed., *Toward Scientific Marketing*, American Marketing Association, 1963.