Volume Title: Annals of Economic and Social Measurement, Volume 4, number 2

Volume Author/Editor: NBER

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm75-2

Publication Date: April 1975

Chapter Title: Random Walk Models of Advertising, Their Diffusion Approximation, and Hypothesis Testing

Chapter Author: Charles Tapiero

Chapter URL: http://www.nber.org/chapters/c10399

Chapter pages in book: (p. 293 - 309)

Annals of Economic and Social Measurement, 4/2, 1975

RANDOM WALK MODELS OF ADVERTISING, THEIR DIFFUSION APPROXIMATION AND HYPOTHESIS TESTING

BY CHARLES S. TAPIERO

Hypotheses concerning market behavior are shown to lead to stochastic process models of advertising. Using diffusion approximations, these models are transformed to stochastic differential equations which are used for determining optimum approximate filter estimates and for hypothesis testing. Using a result given by Appel (1), *the stochastic differential equation of the likelihood ratio of two hypotheses is found. This ratio is used to accept or reject specific equations as models of economic behavior. For demonstration purposes, a numerical example, using the Lydia Pinkham sales-advertising data* (9), *is used to test the hypotheses of the Nerlove–Arrow* (7) *and Vidale–Wolfe* (19) *type models.*

1. INTRODUCTION[1]

To date, advertising models with carry-over effects have assumed that sales reflect past advertising efforts as well as the "forgetting" of these efforts over time. Notable examples are the Nerlove–Arrow model (7), Stigler (14), Gould (4), and Vidale–Wolfe models (19). The basic assumption of these models is that sales response to advertising is deterministic. That is, given an advertising rate, given the effects of an advertising effort on sales, and given the parameters describing "forgetting" of past advertising by consumers, a resultant sales level can be uniquely determined by solving one or a system of differential equations. Each of these differential equations, implicitly and sometimes explicitly, makes specific assumptions concerning market memory mechanisms and advertising effectiveness functions. The choice of an advertising model, therefore, presupposes implicitly market behavior which is for the most part untested.

The purpose of this paper is to propose random walk models of advertising which render explicit the assumptions made concerning a market's behavior. This approach allows a probabilistic interpretation of advertising effectiveness and forgetting. It will also be shown that under specific hypotheses concerning the advertising process, we obtain Nerlove–Arrow and Stigler diffusion models as mean evolutions. Further research is required, however, to determine the implications of such models for optimum advertising policies. Given random walk models of advertising, we provide several solutions—in terms of conditional evolution of probability distributions and conditional probability moments. For empirical parameter estimation purposes, diffusion approximations are used to transform the random walk models into nonlinear filtering problems. Well-known algorithms for the approximate optimum filters are then suggested (2, 5, 6, 10, 11).

The basic assumption of this paper is that in economic and social science, every stochastic model is a hypothesis concerning behavior. This hypothesis

[1] This research was supported in part by a grant from the Kaplan School of Social Science, Hebrew University. The author is grateful to Professor Julian Simon for useful comments and discussion.

usually taken for granted must be rendered explicit, must be tested, and statistical tools of analysis must be developed to provide confidence levels and criteria for acceptance or rejection of the model on the basis of empirical evidence. By determining the likelihood ratio of say two competing models of economic behavior, the statistical acceptance and rejection of a model describing behavior can be drawn from empirical data. For brevity, essential results are summarized in Tables, and a numerical example using the Lydia Pinkham data (9) is used to compare the hypotheses of market behavior.

2. Random Walk Models of Advertising

We assume that advertising expenditures affect the probability of sales and that in a small time interval Δt, the probability that sales will increase by one unit is a function of this advertising rate. Similarly, in a time interval Δt, the probability that sales will decrease by one unit is a function of the forgetting rate. Thus, the advertising model we construct is a random walk model (3).

Consider a line taking the values $x = 0, 1, 2, 3, \ldots M$ where x represents a level of sales and M is the total market potential. Denote by $P(x, t)$ the probability of selling x at time t. At time $t + \Delta t$, the probability of selling x is given by:

$$
\begin{aligned}
\text{(2.1)} \qquad P(x, t + \Delta t) = {} & P(x + 1, t)m(x + 1)\,\Delta t \\
& + P(x, t)[1 - m(x)\,\Delta t][1 - q(M, x, a(t))\,\Delta t] \\
& + P(x - 1, t)q(M, x - 1, a(t))\,\Delta t
\end{aligned}
$$

where $m(x)\,\Delta t$ is the probability that a unit of sales is lost by forgetting. This probability is given as a function m of the aggregate sales x. The probability that a unit sales is generated by an advertising effort $a(t)$ in a time interval Δt is given by the function $q(M, x, a(t))\,\Delta t$ where M denotes the magnitude of a potential demand, x is the sales at time t, and $a(t)$ is the advertising rate at time t. When Δt is very small, equation (2.1) with appropriate boundary restrictions on x, reduces to (2.2):

$$
\begin{aligned}
dP(x, t)/dt &= m(x + 1)P(x + 1, t) - [m(x) + q(M, x, a(t))]P(x, t) \\
&\quad + q(M, x - 1, a(t))P(x - 1, t) \\
\text{(2.2)} \qquad dP(0, t)/dt &= m(1)P(1, t) - q(M, 0, a(t))P(0, t) \\
dP(M, t)/dt &= -[m(M) + q(M, M, a(t))]P(M, t) \\
&\quad + q(M, M - 1, a(t))P(M - 1, t).
\end{aligned}
$$

A solution of (2.2)—the Kolmogorov forward equations (3)—will yield the probability of selling x units at time t as a function of the advertising rate $a(t)$ and the forgetting rate m. A general solution to this equation requires that specific assumptions be made concerning the functional forms m and q. *These assumptions can in fact be construed as explicit hypotheses concerning a market's behavior.* Therefore, specification of the transition probabilities m and q provide a model of market behavior. We shall consider below two hypotheses (see Table 1).[2]

[2] Other models can of course be considered by assuming other transition probabilities. For such models, see Gould (4) for example.

TABLE I

	Nerlove–Arrow Model I	Vidale–Wolfe Model II
Hypothesis	mx $q(a)$	mx $q(a)(M - x)$
Probability Generating Function $F(z, t)$	$\partial F/\partial t = (z - 1)(q(a)F - m\,\partial F/\partial z)$ $F(z, 0) = z^{q^0}$	$\partial F/\partial t = q(a)M(z - 1)F + (m + q(a)(1 - z)\,\partial F/\partial z$ $F(z, 0) = z^{s^0}$
Mean Evolution	$dy/dt = -my + q(a)$ $y(0) = q^0$	$ds/dt = -ms + q(a)(M - s)$ $s(0) = s^0$
Variance Evolution	$dv/dt = mq + q(a) - 2v$ $v(0) = 0$	$dv/dt = q(a)M + (m - q(a))s - 2v(m + q(a))$ $v(0) = 0$

For the first hypothesis, which we call the Nerlove–Arrow hypothesis,[3] $x(t)$ is interpreted in units of goodwill. It assumes that the probability of losing a unit by forgetting is proportional to the goodwill level $x(t)$ at time t. The advertising effectiveness function, expressed as the probability of increasing goodwill by one unit is proportional to some (possibly nonlinear) function of the advertising rate irrespective of the market size which is assumed to be potentially infinite. We can also show (see Table 1) that the probability distribution of goodwill has a mean evolution equivalent to that of Nerlove–Arrow (7) (see also 16, 17).

The second hypothesis is called the diffusion hypothesis.[4] It assumes a finite market and an advertising effectiveness proportional to the remaining market potential $M - x(t)$ and to some function (possibly nonlinear) of the advertising rate. This model can be shown to lead to a mean evolution given by Vidale–Wolfe (19) and Stigler (14) (see also 15, 17).

Given these hypotheses, we substitute the corresponding transition probabilities into (2.2) and solve for $P(x, t)$—the probability of selling x at time t. An explicit solution of $P(x, t)$ is difficult. Nonetheless, by determining the probability generating function of (2.2), an evolution of the probability moments under both hypotheses can be found. For brevity, Table 1 includes both models, the partial differential equation of the probability generating functions and a mean-variance evolution of the random variable $x(t)$. Given these (and higher order) moments, a "certainty equivalent" advertising strategy can theoretically be selected to reflect both managerial motives and attitudes towards risk.

In practice, the transition probabilities reflecting market hypotheses can hardly be assumed known. Further, sales are only probabilistically defined. For this reason, it is necessary to obtain methods estimating sales and testing the effects of the transition probabilities. This paper considers an approach which reduces random walk models (by diffusion approximations) to stochastic differential equations. Application of approximate filtering techniques, for example, will then yield optimum sales response estimates to advertising programs. Further, the filter estimates can be used to compute the likelihood ratio of two competing alternative hypotheses. This likelihood ratio may then be used to accept or reject a model of market behavior on the basis of empirical observations.

3. The Diffusion Approximation and Optimum Approximate Filters

A diffusion approximation of (2.2) is found by replacing $P(x + 1, t)$ and $P(x - 1, t)$ by the first three terms of a Taylor series expansion about $P(x, t)$. The resultant equation is a Fokker–Plank partial differential equation whose solution is a stochastic integral equation, given by:

$$\text{(3.1)} \qquad x(t) - s^0 = \int_{t_0}^{t} [-mx(\tau) + q(a(\tau))]\, d\tau + \int_{t_0}^{t} [mx(\tau) + q(a(\tau))]^{1/2}\, dw(\tau)$$

$$x(t) \geq 0$$

[3] The model described however, has not been derived by Nerlove and Arrow (7). Rather, we find a mean evolution which is structurally similar to that of Nerlove and Arrow.

[4] This is based on a market share hypothesis. Thus, advertising has an effect on the market share of a firm (Stigler (14)).

with $dw(\tau)$ a standard Wiener process;

$$E\,dw(\tau) = 0$$

$$E\,dw(t)\,dw(\tau) = \delta(t - \tau)$$

and

$$\delta(t - \tau) = \begin{cases} 1 & \text{if } t = \tau \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

As usual in stochastic control, we assume that (3.1) is satisfied with probability one, and therefore a stochastic differential equation in the sense of Itô can be defined. Because of the reflecting barriers (at $x = 0$ and $x = M$), we replace the initial conditions by inequality constraints. For both the Nerlove–Arrow (7) and Vidale–Wolfe (19) models, the diffusion approximations are given in Table 2.[5] For simplicity, assume that $x \geq 0$ and let $Y(t)$ be a sales time series with continuous measurements $y(t)$,

$$Y(t) = \{y(\tau) | \tau \leq t\}. \tag{3.3}$$

Conditional mean estimates for $x(t)$ are given by:

$$\hat{x} = E(x|Y) = \int_{-\infty}^{\infty} xP(x|Y)\,dx. \tag{3.4}$$

An algorithm for generating such sales estimates and the corresponding error variance are found by non-linear filtering techniques. For simplicity, a first order solution algorithm with known advertising strategy and appropriate measurement model yields, for example, the optimum goodwill and sales estimates given in Table 2. Greater accuracy can be reached by using higher order approximations and other non-linear filtering techniques. It is also evident that a wide variety of approximations can be suggested since we can also consider alternate models of advertising as indicated by the use of Itô's differential rule.[6] Specifically, if $h(x)$—a function of goodwill—denotes sales, the Nerlove–Arrow stochastic differential equation can be transformed (using Itô's differential rule) to a non-linear stochastic differential equation of sales.[7] Next, we consider the problem of hypothesis testing which is of central interest to this paper. The results briefly summarized thus far are required for the hypothesis testing on and of the models outlined above.

[5] We include in this Table additional equations to be discussed below.

[6] The Itô differential rule is defined as follows. Given a random variable x and given the stochastic differential equation:

$$dx = f(x, t)\,dt + g(x, t)\,dw$$

with dw a Wiener process, then the transformed variable $y = h(x, t)$ is described by the stochastic differential equation:

$$dy = \left\{ \frac{\partial h}{\partial t} + f(x, t)\frac{\partial h}{\partial x} + \tfrac{1}{2}g^2(x, t)\frac{\partial h}{\partial x^2}\,dt + g(x, t)\frac{\partial h}{\partial x}\,dw \right\}$$

[7] Specifically the change of variables

$$Q = \int \frac{[mx + q(a)]^{1/2}}{\partial h/\partial x}\,dx$$

will transform the Nerlove–Arrow model into a non-linear model with additive disturbances.

TABLE 2

	Nerlove–Arrow Model I	Vidale–Wolfe Model II
Diffusion Approximation	$dx = (-mx + q(a))\,dt + (mx + q(a))^{1/2}\,dw$ $x \geq 0$	$dx = (-mx + q(a)(M - x))\,dt + (mx + q(a)(M - x))^{1/2}\,dw$ $x \geq 0 \quad x \leq M$
Measurement Model	$y = h(x) + \eta$ $E(\eta) = 0, \quad E(\eta(t)\eta(\tau)) = \delta(t - \tau)\theta^2$	$y = x + \varepsilon$ $E(\varepsilon) = 0, \quad E(\varepsilon(t)\varepsilon(\tau)) = \delta(t - \tau)\psi^2$
First Order Mean Estimates	$d\hat{x}/dt = -m\hat{x} + q(a) + V_{\hat{x}}(\partial h(\hat{x})/\partial\hat{x})(y - h(\hat{x}))/\theta^2$	$d\hat{x}/dt = -m\hat{x} + q(a)(M - \hat{x}) + V_{\hat{x}}(y - \hat{x})/\psi^2$
First Order Variance Estimates	$dV_{\hat{x}}/dt = -2mV_{\hat{x}} + (m\hat{x} + q(a)) - V_{\hat{x}}^2(\partial h(\hat{x})/\partial\hat{x})^2/\theta^2$	$dV_{\hat{x}}/dt = -2(m + q(a))V_{\hat{x}} + m\hat{x} + q(a)(M - \hat{x}) - V_{\hat{x}}^2/\psi^2$

4. Hypothesis Testing

The stochastic advertising models defined earlier are now considered as hypotheses concerning market memory mechanisms and advertising effectiveness. The functional form of the transition probabilities renders explicit the implicit assumptions included in the advertising models.

The number of hypotheses one may test is of course very large. These include hypotheses concerning parameters, functional forms of transition probabilities (i.e. process models), measurement models etc. Further, we may distinguish between cases where available evidence (i.e. the data) is itself drawn from a stochastic (or non stochastic) model. We shall consider four types of problems below and treat one in detail in the next section. A summary of these problems can be found in Table 3.

The first two problems assume a random sales-advertising process, and hypotheses are built upon the qualitative and quantitative sales effects of advertising. Specifically, the first problem assumes a random Nerlove–Arrow process, and establish hypotheses on the probable relationships between the measurement of sales and goodwill. The second problem,[8] on the other hand assumes some general random process of sales and advertising and uses the Nerlove–Arrow and Vidale–Wolfe models as sales measurement hypotheses. Empirical evidence may then be brought to bear on each of these hypotheses. The third and fourth problems in Table 3, assume deterministic sales advertising processes. These processes although unknown are given by sales and advertising time series. Tests of hypotheses are then conducted on two advertising effectiveness functions $q_0(a)$ and $q_1(a)$ (problem 3) and the Nerlove–Arrow and Vidale–Wolfe models (problem 4). To test these hypotheses, we use empirical evidence as given by the sales and advertising time series $Y(t)$ and $A(t)$ respectively, and compute the likelihood ratio $\Lambda(t)$. These are defined below:

(4.1)
$$Y(T) = \{y(\tau)|\tau \leq T\}$$
$$A(T) = \{a(\tau)|\tau \leq T\}$$

(4.2)
$$\Lambda(T) = \frac{\pi_0 P[H_1|Y(T), A(T)]}{(1-\pi_0)P[H_0|Y(T), A(T)]}.$$

Here π_0 and $(1 - \pi_0)$ are the *a priori* probabilities of the null and alternative hypotheses H_0 and H_1 respectively and $P[H_j|Y(T), A(T)]$ $(j = 0, 1)$ are therefore the conditional probabilities of hypothesis H_j $(j = 0, 1)$ on the time series (4.1). With binary hypotheses, of course, we have

(4.3)
$$P[H_1|Y(T), A(T)] + P[H_0|Y(T), A(T)] = 1.$$

For computational purposes, it is more convenient to compute the log likelihood ratio, $z(T)$

(4.4)
$$z(T) = \log \Lambda(T)$$

and use it to reach a decision concerning each of the hypotheses.

[8] In other words, we assume that the sale-advertising process is random and use tests of hypotheses on random measurement models. Problem 2, is however, an unsolved problem.

TABLE 3

Model	Null Hypothesis H_0	Alternative Hypothesis H_1	Comments
1. $dx = (-mx + q(a))\,dt + (mx + q(a))^{1/2}\,dw$	$s = h_0(x, t) + v$	$s = h_1(x, t) + v$	Measurement of sales in the Nerlove–Arrow hypothesis
2. $dx = f(x, a, t)\,dt + dv$	$ds = (-mx + q_0(a))\,dt + (mx + q_0(a))^{1/2}\,dw$	$ds = (-mx + q_1a(M - x))\,dt + (mx + q_1a(M - x))^{1/2}\,dw$	Testing Nerlove–Arrow and Vidale–Wolfe models using an empirical model
3. Time series of sales and advertising	$ds = (-mx + q_0(a))\,dt + (mx + q_0(a))^{1/2}\,dw$	$ds = (-mx + q_1(a))\,dt + (mx + q_1(a))^{1/2}\,dw$	Testing advertising effectiveness functions
4. Time series of sales and advertising	$ds = (-mx + q_0(a))\,dt + (mx + q_0(a))^{1/2}\,dw$	$ds = (-mx + q_1a(M - x))\,dt + (mx + q_1a(M - x))^{1/2}\,dw$	Testing Nerlove–Arrow and Vidale–Wolfe model using time series

In the nonlinear model defined by problem 1 for example, it can be shown[9] (see Appel (1)) that $z(t)$ satisfies a stochastic differential equation given by;

$$(4.5)\quad dz/dt = [E_0 h_0(x, t) - E_1 h_1(x, t)] \cdot \{s - \tfrac{1}{2}[E_0 h_0(x, t) + E_1 h_1(x, t)]\}/\theta^2$$
$$z(0) = 0$$

where θ^2 is the error variance of $v(t)$ in problem 1, and where E_0 and E_1 denote conditional expectations with respect to probability distributions $p(s, t|H_0, Y(t), A(t))$ and $p(s, t|H_1, Y(t), A(t))$ respectively. These conditional expectations are precisely the mean (filter) estimates given in Table 2 under both hypotheses H_0 and H_1. If h_0 and h_1 are two non-linear functions, Taylor series approximations yield;

$$(4.6)\quad h_j(x, t) \approx h_j(\hat{x}, t) + \frac{\partial h_j}{\partial x}(\hat{x}, t)(x - \hat{x}) + \frac{1}{2}\frac{\partial^2 h_j}{\partial x^2}(\hat{x}, t)(x - \hat{x})^2.$$

Inserting (4.6) into (4.5) yields a log likelihood ratio stochastic differential equation given by;

$$(4.7)\quad dz/dt = \left[\Delta h - \frac{\Delta\partial^2 h}{2\partial x^2} V_{\hat{x}}^j\right]\left\{s - \frac{1}{2}\sum_{j=0}^{1}\left(h_j + \frac{\partial^2 h_j}{2\partial x^2} V_{\hat{x}}^j\right)\right\}\Big/\theta^2$$
$$z(0) = 0$$

where $\Delta h = h_1(x, t) - h_0(x, t)$ and the subscripts x and t are implied in h. Also, $V_{\hat{x}}^j$ denotes the error variance under both hypotheses as denoted in Table 2.

When h_j are linear functions, the stochastic differential equation in (4.5) is a quadratic stochastic differential equation and a solution for $z(t)$ although difficult is possible. When h_j are non-linear, a solution for $z(t)$ is almost impossible. In such a case we turn to approximations.

If instead of problems 1 and 2 we consider problems 3 and 4, a general solution for $z(t)$ can be found. Specifically, consider the discrete time version of problem 3

$$(4.8)\quad \begin{aligned} \text{Null } H_0:\quad & \Delta s = [-m_0 x + q_0(a)]\,\Delta t + [m_0 x + q_0(a)]^{1/2}\,\Delta w \\ \text{Alternative } H_1:\quad & \Delta s = [-m_1 x + q_1 a(M - x)]\,\Delta t \\ & \qquad + [m_1 x + q_1 a(M - x)]^{1/2}\,\Delta w \end{aligned}$$

where Δs are sales increments, Δt the time interval is taken to equal one, and Δw is therefore a standard normal distribution. Thus $\Delta s(i)$, $i = 1, \ldots T$ is a normal random vector with mean vector N_j and variance-covariance matrix K_j under the null ($j = 0$) and alternative ($j = 1$) hypotheses. Given sales and advertising measurement $x(i), a(i)$, $i = 1, \ldots T$ respectively, the likelihood ratio of the two

[9] Proof of this equation is found by computing conditional estimates probability distributions and using Itô's differential rule. For brevity, the proof is deleted.

hypotheses in (4.8) is now desired. We let;

$$N_j = \{n_j(1), n_j(2), \dots n_j(T)\} \qquad j = 0, 1$$

$$n_0(i) = -m_0 x(i) + q_0(a(i))$$

$$n_1(i) = -m_1 x(i) + q_1 a(i)[M - x(i)]$$

$$K_j \triangleq E[(\Delta s - N_j)(\Delta s - N_j)'|H_j]$$

$$(4.8) \qquad K_j = \left\| \begin{matrix} k_j(1) & & & \\ & k_j(2) & & 0 \\ & & \ddots & \\ 0 & & & \ddots \\ & & & k_j(T) \end{matrix} \right\|$$

$$k_0(i) = m_0 x(i) + q_0(a(i))$$

$$k_1(i) = m_1 x(i) + q_1 a(i)(M - x(i)).$$

Computations of the variance-covariance matrices K_j ($j = 0, 1$) in (4.8) can be easily proved by noting that $E(\Delta w(t)\, \Delta w(\tau)) = 0$ for $t \neq \tau$. Now define the likelihood ratio of the two hypotheses:

$$(4.10) \qquad \Lambda(T) = \frac{|K_0|^{1/2} \exp\left[-\frac{1}{2}(\Delta s - N_1)' Q_1 (\Delta s - N_1)\right]}{|K_1|^{1/2} \exp\left[-\frac{1}{2}(\Delta s - N_0)' Q_0 (\Delta s - N_0)\right]}$$

where $Q_j = K_j^{-1}$—the inverse matrix of K_j. The log likelihood ratio is clearly given by $z(T)$

$$z(T) = l_1(T) - l_0(T)$$

$$(4.11) \qquad l_j(T) = -\tfrac{1}{2}(\Delta s - N_j)' Q_j (\Delta s - N_j) - \tfrac{1}{2} \ln |K_j| \qquad j = 0, 1$$

Or[10]

$$(4.12) \qquad l_j(T) = -\frac{1}{2}\left\{ \sum_{i=1}^{T} \frac{[\Delta s(i) - n_j(i)]^2}{k_j(i)} + \ln k_j(i) \right\}.$$

In continuous time (when Δs becomes very small), (4.12) is reduced to:

$$(4.13) \qquad l_j(T) = -\frac{1}{2}\int_0^T \left\{ \frac{[ds(t) - n_j(t)]^2}{k_j(t)} + \ln k_j(t) \right\} dt \qquad j = 0, 1.$$

The log likelihood ratio is used next to accept or reject hypotheses. (For a thorough study of this problem see Van Trees (18)). For simplicity, we shall consider a decision threshold F, then

$$(4.14) \qquad \begin{aligned} &\text{If } z(T) > F \quad \text{accept } H_1 \\ &\text{If } z(T) \leq F \quad \text{accept } H_0. \end{aligned}$$

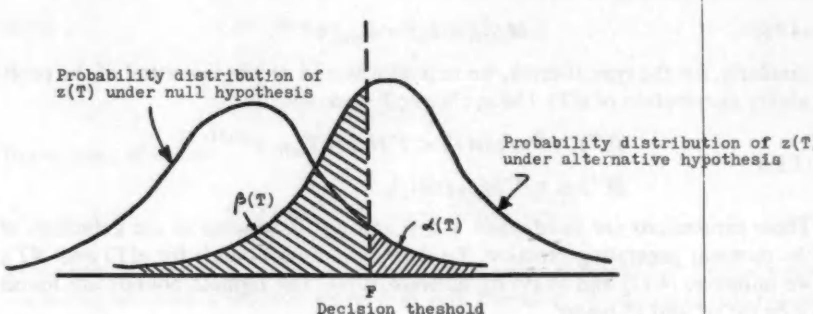
[10] This is easily proved by noting that Qj—the inverse matrix, is given by: $q_{jj} = 1/k_{jj}$ and $q_{ij} = 0$ for $i \neq j$.

This threshold, standard in statistics (e.g. Wald (20)) is calculated in terms of type I and type II errors. Namely, consider the test of the hypothesis at time T (see Figure 1) and define

$\alpha(T)$: Type I error at time T (or false alarm probability)

$\beta(T)$: Type II error at time T

$$\alpha(T) = \text{Prob}\,[z(T) \geq F|H_0] \qquad (4.15)$$
$$\beta(T) = \text{Prob}\,[z(T) \leq F|H_1].$$



The choice of the threshold level is an important and fundamental one in statistics. In hypothesis testing, it is common to fix the type I error to a predetermined level and solve for F. Given F, the type II error is also determined. By balancing these two errors, an appropriate threshold level can be found. To determine the threshold level F from $\alpha(T)$ and the corresponding error $\beta(T)$, however, it is necessary to compute the probability distribution of $z(T)$ under both the null and alternative hypotheses. Equation (4.5) expressing $dz(t)/dt$ is a diffusion process whose solution as we noted may be difficult. A possible approximation consists in computing the mean-variance evolutions of $z(t)$ and supposing that these are the parameters of a normal probability distribution. Taylor series approximations may also be used in computing the mean-variance evolutions of $z(t)$ (see (4.7)). If we let $\mu_i(t)$ be the conditional mean (normal) estimates under both hypotheses and let $\sigma_i^2(t)$ be the corresponding variances, then the type I and II errors are given by (see Figure 1):

$$\alpha(T) = \text{erfc}\left[\frac{F + \mu_0(T)}{\sigma_0(T)}\right] \qquad (4.16)$$
$$\beta(T) = \text{erfc}\left[\frac{\mu_1(T) - F}{\sigma_1(T)}\right]$$

where

$$\text{erfc}\,(y) = \int_y^\infty \frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2}\,dx.$$

Given $\mu_0(T)$, $\sigma_0^2(T)$ and $\alpha(T)$, it is evident that F can be found by using Tables for the erfc function.

If the normal approximation is not acceptable, we can solve for F using Chernoff bounds (see also (18)). Recall that the type I error is given by:

$$\alpha(T) = \text{Prob}\,(z(T) > F|H_0) \le E_{z|H_0}\, e^{w(z(T)-F)} \tag{4.17}$$

or

$$\alpha(T) \le e^{-Fw} M_{z(T)}(w|H_0)$$

where $E_{z|H_0}$ is expectation of z under the null alternative, $w \ge 0$, and $M_{z(T)}$ is the moment generating function of $z(T)$.

$$M_{z(T)}(w|H_0) = E_{z|H_0}\, e^{wz(T)}. \tag{4.18}$$

Similarly, for the type II error, we require a bound on the lower tail of the probability distribution of $z(T)$. Using Chernoff's bounds

$$\begin{aligned} \beta(T) &= \text{Prob}\,(z(T) < F|H_1) \le E_{z|H_1}\, e^{v(z(T)-F)} \\ \beta(T) &\le e^{-Fv} M_{z(T)}(v|H_1). \end{aligned} \tag{4.19}$$

These expressions are valid when $w \ge 0$ and $v \le 0$, because of the definition of the moment generating function. To determine tight bounds for $\alpha(T)$ and $\beta(T)$, we minimize (4.17) and (4.19) by differentiation. The tightest bounds are found to be for w^* and v^* where

$$\begin{aligned} \frac{d/dw\, M_{z(T)}(w^*|H_0)}{M_{z(T)}(w^*|H_0)} &= F \\ \frac{d/dv\, M_{z(T)}(v^*|H_1)}{M_{z(T)}(v^*|H_1)} &= F. \end{aligned} \tag{4.20}$$

These equations may then be used to determine F, $\alpha(T)$, and $\beta(T)$.[11]

Extensions to sequential tests are straightforward by using Wald's (20) Sequential Probability Ratio Test (SPRT). It is then necessary to compute two bounds F_0 and F_1, and the decision test becomes

$$\begin{aligned} &\text{If } z(T) \ge F_1 \quad \text{accept } H_1 \\ &\text{If } z(T) \le F_0 \quad \text{accept } H_0 \\ &\text{If } F_0 < z(T) < F_1 \quad \text{continue Data Collection.} \end{aligned} \tag{4.21}$$

To determine F_0 and F_1, we use the fundamental relation given by Wald, and note that:

$$\begin{aligned} \frac{\beta(T)}{1-\alpha(T)} &\le e^{F_1} \\ \frac{\alpha(T)}{1-\beta(T)} &\le e^{-F_0}. \end{aligned} \tag{4.22}$$

[11] Rather we compute the bounds on $\alpha(T)$ and $\beta(T)$. First we assume an upper bound for $\alpha(T)$ and solve for w^* in (4.17). We use the first part of equations (4.20) to compute F, the second part to compute v^* and finally, use (4.19) to compute the upper bound on $\beta(T)$.

These inequalities, of course, provide only upper limits for $\alpha(T)$ and $\beta(T)$. In summary, given F (or F_0 and F_1), the hypotheses we have considered can be tested on-line. As additional data is accumulated, a decision can then be made regarding the acceptance or rejection of the hypothesis.

When the model is non-stochastic (as problems 3 and 4 in Table 3), the log likelihood ratio in (4.11) consists (because of the diffusion approximation) in the difference of two non-central chi-squared random variables. The test of the hypothesis is thus;

$$G(T) = \sum_{j=0}^{1} \frac{(-1)^j}{2} (\Delta s - N_j)' Q_j (\Delta s - N_j)$$

(4.23)
$$F^* = F + \tfrac{1}{2} \ln |K_1| - \tfrac{1}{2} \ln |K_0|$$

$$G(T) \begin{cases} > F^* & \text{accept } H_1 \\ \leq F^* & \text{accept } H_0. \end{cases}$$

In our case, of course,

$$G(T) = \sum_{i=1}^{T} \sum_{j=0}^{1} \frac{(-1)^j}{2} \frac{[\Delta s(i) - n_j(i)]^2}{k_j(i)}.$$

In the special case of zero-cost when the right decision is reached and equal costs if a wrong decision is reached, we have $F = 0$. Therefore, the decision rule to test the hypothesis is;

(4.24)
$$\begin{aligned} l_1(T) &> l_0(T) \quad \text{accept } H_1 \\ l_1(T) &\leq l_0(T) \quad \text{accept } H_0 \end{aligned}$$

where $l_j(T)$ $(j = 0, 1)$ are given by (4.12) or (4.13). If an $\alpha(T)$ error is specified, it is evident that $G(T)$ is given by the difference of two chi-squared distributions. Under the null hypothesis, $[\Delta s(i) - n_0(i)]/\sqrt{k_0(i)}$ is a standard normal distribution Thus, the sum of the squares has a central chi-square distribution of degree T. For the second sum, we note that under the null hypotheses, these have a non-central chi-squared distribution (i.e., resulting in the sum of independently normal distributed random variables with mean $[n_0(i) - n_1(i)]/\sqrt{k_1(i)} = \Delta n(i)/\sqrt{k_1(i)}$ and variance $k_0(i)/k_1(i)$. We make the approximation

(4.25)
$$k_0(i)/k_1(i) \approx \sigma^2$$

and define the noncentrality parameter λ^2;

(4.26)
$$\lambda^2 = \sum_{i=1}^{T} [\Delta n(i)]^2 / k_1(i)$$

where σ^2 is a constant for all $i = 1, \ldots T$ (i.e., the ratio of variances under both hypotheses is a constant). The moment generating function of the log-likelihood ratio $G(T)$ is then given by $M_{G(T)}(w|H_0)$;

(4.27)
$$M_{G(T)}(w|H_0) = \left[\frac{1}{(1 + 2w)(1 - 2w\sigma^2)}\right]^{T/2} \exp\left[\frac{w\lambda^2}{(1 - 2w\sigma^2)}\right].$$

The log of M is thus

$$(4.28)\quad \log M_{G(T)}(w|H_0) = -\frac{T}{2}[\log(1 + 2w) + \log(1 - 2w\sigma^2)] + w\lambda^2/(1 - 2w\sigma^2).$$

The mean and variance $G(T)$ under the null hypothesis can then be computed by taking successive derivatives of (4.28). The moment generating function $M_{G(T)}(w|H_1)$ and probability moments of the log likelihood ratio under the alternative hypothesis are similarly found. To obtain a bound on the $\alpha(T)$ error, we take the derivative of (4.28), equate it to F^* (the threshold) and solve for w^*. Thus,

$$(4.29)\qquad F^* = T[\sigma^2/(1 - 2w^*\sigma^2) - 1/(1 + 2w^*)] + \lambda^2/(1 - 2w^*\sigma^2)^2.$$

A bound on the $\beta(T)$ error is obtained by deriving $M_{G(T)}(v|H_1)$ which is also the moment generating function of a difference of chi-squared distributions. To obtain exact results, the moment generating functions $M_{G(T)}(w|H_0)$ and $M_{G(T)}(v|H_1)$ ought to be inverted. Although this is possible (see Omura and Kailath (8)), the resultant distribution is an extremely complicated one.

The importance of the results obtained earlier is now demonstrated by applying them to an examination of advertising effectiveness functions using the Lydia Pinkham data (9).

5. The Lydia Pinkham Case Revisited

The Lydia Pinkham case has been extensively treated in the literature on advertising theory (see (9) for an excellent survey and analysis). Popularity of this case in the advertising literature is essentially due to the availability of extensive sales and advertising time series. Furthermore, the firm, through its long history, has essentially been unaffected by competition and sales have been shown to be extremely sensitive to advertising budgets. We shall therefore use this data in testing advertising effectiveness functions. Specifically, we use (seasonally adjusted and the original) monthly sales-advertising time series[12] for the periods January 1954 to June 1960, to test the hypothesis of economies of scale in advertising. The results we found corroborate studies by Simon (13) and Palda (9) although we use an entirely different procedure. Further research is currently being conducted to test alternative models and data[13] in verifying this and other hypotheses. The sales-advertising model we consider is of the Nerlove–Arrow type[14] and is given by;

$$(5.1)\qquad \Delta s = [-mx + q_0 a^{\delta}] + [mx + q_0 a^{\delta}]\,\Delta w.$$

Several thousand hypotheses were tested using alternative parameter configurations.[15] Maximum likelihood parameter configurations are summarized in Table 4. Results in this Table are given for the first 58, 68 and 78 measurements of the time

[12] Palda [9], pp. 32–3.

[13] Specifically, Schmalense's (12) data on cigarettes as well as other diffusion models.

[14] Here $\Delta t = 1$ and goodwill is equated to sales.

[15] In other words, a large number of parameters (m, q_0, δ) were tested and only the configurations (m^*, q^*, δ^*) with very high likelihood accepted.

TABLE 4

Seasonally Adjusted Data (monthly)					Original Data (monthly)				
Length	m	q	δ	chi	Length	m	q	δ	chi
58	0.400	0.500	1.03	*3.54*	58	0.400	0.400	1.01	*3.52*
58	0.450	0.500	1.05	3.60	58	0.450	0.600	0.97	3.63
68	0.400	0.400	1.07	*3.59*	58	0.550	0.400	1.05	3.64
68	0.400	0.600	1.01	3.61	58	0.600	0.500	1.03	3.69
68	0.450	0.400	1.09	3.69	68	0.450	0.500	0.99	*3.58*
68	0.450	0.500	1.05	3.60	68	0.450	0.600	0.97	3.63
68	0.500	0.500	1.07	3.62	68	0.550	0.700	0.97	3.69
78	0.400	0.400	1.07	3.59	68	0.600	0.500	1.03	3.69
78	0.400	0.500	1.03	*3.54*	78	0.550	0.400	1.05	*3.64*
78	0.400	0.600	1.01	3.61	78	0.550	0.700	0.97	3.69
78	0.450	0.400	1.09	3.69	78	0.600	0.500	1.03	3.69
78	0.500	0.500	1.07	3.62					

series. We note here the δ—the scaling parameter, is extremely close to one. Thus, any competing hypothesis with $\delta > 1$ (or $\delta < 1$) is likely to be rejected compared to the hypothesis that $\delta = 1$. Such hypotheses were in fact tested and rejected. Experiments were also conducted using the Vidale–Wolfe model. This model was found to be insignificant, however.[16] This is to be expected since in the Lydia Pinkham case, the concept of market share, on which the Vidale–Wolfe model is based, makes little sense. Finally, in the analysis of the Lydia Pinkham yearly data we encountered a trend which was not accounted for in the stochastic models constructed in this paper.[17] For empirical analysis purposes, such a trend is necessary to reflect more precisely the effects of forgetting and advertising on sales.

6. Conclusion

One of the first problems in the analysis of dynamical systems is to construct appropriate models which reflect reality. This is particularly important when we consider economic, social, and management applications. In these fields, an equation mapping behavior can be assumed at best to be a hypothesis. The choice of the relevant variables and behavioral hypotheses in fact determine the resultant dynamic models. If this is so, it is imperative that we provide the explicit mathematical and statistical tools for testing the hypotheses we make concerning a behavioral process.

In this paper, a set of advertising models were constructed starting from simple hypotheses concerning market behavior. Using the simple structure of random walk models, hypotheses concerning memory mechanisms and advertising effectiveness were expressed in terms of transition probabilities. Given the corresponding random walk model, diffusion approximations were shown to lead

[16] In other words, in all cases, the log likelihood was found to be large. Further study of the Vidale–Wolfe's model is however currently investigated using the Schmalense cigarettes data.

[17] This is a particularly important point for empirical analyses since the stochastic process model assumed only the effects of forgetting and advertising on sales.

to non-linear stochastic differential equations. This formulation of the problem is standard in non-linear filtering theory.

The models of advertising suggested in this paper have mean evolutions equivalent to the Nerlove–Arrow model (7), Vidale–Wolfe (19) and Stigler (14) models. This particular property of the models points out some explicit hypotheses made by the authors. Evidently, there may be a great number of hypotheses which can be shown to lead to mean evolutions as given in this paper. An interesting and important question would be to consider the inverse problem—that of finding the range of hypotheses giving rise to a particular evolution. This problem is a difficult one and is not in the scope of this paper.

For empirical analysis purposes, we computed the likelihood ratio of hypotheses and thereby obtained a mechanism for testing on-line, models as well as parameter configurations. To demonstrate our results, a numerical example concerning economies of scale in advertising was considered. Maximum likelihood scaling estimates were shown to be in the neighborhood of one, thereby rejecting the hypothesis of economies of scales. Of course this numerical example is merely a preliminary analysis and further empirical research is clearly required.

Columbia University

References

[1] Appel, J. J., Approximate Optimum Nonlinear Filtering, Detection and Demodulation of Non-Gaussian Signals, *Ph.D. Dissertation*, Columbia University, New York, 1970.

[2] Bucy, R. S., "Nonlinear Filtering Theory," *IEEE Transactions on Automatic Control*, AC-10, 1965, 198.

[3] Feller, W., *An Introduction to Probability Theory and Its Application*, Vol. II New York, John Wiley & Sons, 1966.

[4] Gould, J. P., "Diffusion Processes and Optimal Advertising Policy," in E. S. Phelps *et al.* (eds.) *Microeconomic Foundation of Employment and Inflation Theory*, New York, W. W. Norton & Co., 1970, pp. 338–68.

[5] Kalman, R. W., "New Methods in Wiener Filtering" in J. L. Bogdanoff and F. Kozin (Eds.), *Engineering Applications of Random Function Theory and Probability*, New York, Wiley, 1963.

[6] Meditch, J. S., "Formal Algorithms for Continuous-Time Non-Linear Filtering and Smoothing," *Int. J. Control*, 11, 1970, 1061–68.

[7] Nerlove, M. and K. J. Arrow, "Optimal Advertising Policy Under Dynamic Conditions," *Economica*, 39, 1962, pp. 129–42.

[8] Omura, J. and T. Kailath, Some Useful Probability Distributions, Technical Report No. 7050-6, Stanford Electronics Laboratory, Stanford, Cal., September, 1965.

[9] Palda, K. S., *The Measurement of Cumulative Advertising Effects*, Englewood Cliffs, N.J., Prentice Hall, 1964.

[10] Sage, A. P. and J. L. Melsa, *Estimation Theory with Applications to Communications and Control*, New York, McGraw Hill Book Co., 1971.

[11] Schwartz, L., "Approximate Continuous Nonlinear Minimal Variance Filtering," Report No. 67-17, Department of Engineering, University of California, Los Angeles, April 1967.

[12] Schmalensee, R., *The Economics of Advertising*, Amsterdam, North Holland Publishing Co., 1972.

[13] Simon, J. L., "Are There Economics of Scale in Advertising?" *Journal of Advertising Research*, 1965, pp. 15–19.

[14] Stigler, G., "The Economics of Information," *Journal of Political Economy*, 69, 1961, pp. 213–25.

[15] Tapiero, C. S., "On-Line and Adaptive Optimum Advertising Control by a Diffusion Approximation," *Operations Research*, 1975, Forthcoming.

[16] Tapiero, C. S., "Optimal Advertising Control and Goodwill under Uncertainty," Working Paper No. 57, Columbia University, 1974.

[17] Tapiero, C. S., *Managerial Planning; Optimum Control Approach*, Gordon & Breach Science Publishers, Forthcoming, New York.

[18] Van Trees, H. L., *Detection, Estimation and Modulation Theory*, Part I, New York, John Wiley & Sons Inc., 1968.
[19] Vidale, M. L. and H. B. Wolfe, "An Operations Research Study of Sales Response to Advertising," *Operations Research*, 5, 1957, pp. 370–81.
[20] Wald, A., *Sequential Analysis*, New York, John Wiley and Sons, 1947.