Annals of Economic and Social Measurement, 6/2, 1977

IMPULSE RESPONSE IDENTIFICATION AND CAUSALITY DETECTION FOR THE LYDIA-PINKHAM DATA*

BY P. E. CAINES, S. P. SETHI AND T. W. BROTHERTON**

The Lydia-Pinkham data is analysed using a recently developed system identification algorithm. For an observed time series this yields an estimate of the process impulse response which we argue is a more robust modeling device than the traditional autoregressive moving average model for econometric time series. Our results are compared with some earlier results for the Lydia-Pinkham time series. Further, a new multivariate causality test on the data dramatically reveals a uni-directional causal relationship from log-advertising expenditure to sales.

1. INTRODUCTION

One of the most important marketing problems concerns the determination of the best advertising policy. A crucial element in this problem is the relationship between advertising and sales. To quote Bass [4]:

> "There is no more difficult, complex, or controversial problem in marketing than measuring the influence of advertising on sales. There is also probably no more interesting or potentially profitable measurement problem than this one."

The purpose of this paper is to identify and estimate the dynamic sales-advertising relationships from the Lydia-Pinkham time-series data and draw conclusions about the lag structure and the direction of causality in these relationships.

The Lydia-Pinkham data was first analysed systematically by Palda [23] and subsequently a number of times by other workers. For these and other related studies, the reader is referred to a survey by Clarke [14]. In this paper we use a new statistical system identification technique that is especially suited to identifying systems driven by correlated noise. The correlation between residuals was, of course, a problem confronting previous analyses of the Lydia-Pinkham data. Further, we believe that a large class of estimation techniques actually identify the so-called impulse response of the systems generating the observations [7]. As a result we compare our estimates of the impulse response between advertising expendi-

*Various parts of this research were supported by Grant S76-0342 from the Canada Council, Grant A-9328 from the National Research Council of Canada and Grant N00014-75-C-0648 from the Office of Naval Research under the Joint Services Electronics Program.

**The authors are at the Division of Engineering and Applied Physics, Harvard University (formerly with the University of Toronto), Faculty of Management Studies, University of Toronto, and Department of Electrical Engineering, University of Hawaii (formerly with the University of Toronto), respectively.

ture and sales for the Lydia-Pinkham data with those implicitly estimated by Palda. This comparison results in the general agreement that was expected. Finally, we employ a multivariable identification technique to analyze the causality relationship between sales and advertising for the Lydia-Pinkham time series using a recently completed theory of causality [8, 9, 10, 17, 30]. This analysis dramatically reveals a uni-directional causality from the logarithm of advertising expenditures in dollars (possibly a surrogate of effective advertising) and sales measured directly in dollars.

2. Previous Analyses of the Lydia-Pinkham Data

Palda's work [23] was the first important empirical research which found support for the existence of carry-over effects of advertising. For his analysis Palda chose the Lydia-Pinkham data and Koyck's distributed lag scheme [20].[1] This scheme simplified the linear distributed lag model considerably by converting a large number of (lagged) exogenous variables into a model with only one lagged and one non-lagged exogenous variable. While the scheme oversimplifies the dynamics of the sales-advertising relationship, it may have been necessary to employ such a scheme given the state-of-the-art in statistical estimation at the time of Palda's study. The use of a greatly simplified model avoids the issue of the proper number of lagged terms to include in the regressions.

Of the large number of regressions which he ran, Palda selected those having the best fitting estimates, i.e., with the largest coefficients of determination R^2. These were referred to as KOYK, KOYL1, KOYL2, and KOYLDIF and are described in Section 5 below. For the final comparison, he also included KOYL2 Yless for its superior forecasting performance. The criterion for this was the Measure of Percentage Error (MPE). Since this was computed on the basis of one observation only it is judged, however, to be lacking any statistical meaning.

Clarke and McCann introduced a 'current-effect' model [15, p. 136] to challenge Palda's results. Given the negative signs of regression coefficient when lagged advertising variables are included [23, p. 90; 15, p. 135], Clarke and McCann suspected the validity of the Koyck scheme for the effect of advertising on sales. In their current effect model, Clarke and McCann assumed a serially correlated noise structure to account for carryover effects. Using the method of frequency domain regression (FDR), they concluded that Palda's coefficient (.537) of current advertising in his KOYK model was 17% lower than theirs (.642) and that Palda's

[1] The Koyck scheme assumes geometrically declining future effects: i.e., it assumes $\alpha_{i+1} = \lambda\alpha_i$, $0 < \lambda < 1$, in $S_t = k + \sum_{i=0}^{\infty} \alpha_i U_{t-i}$.

coefficient of lagged sales (.628) indicated a longer carryover than one year obtained with FDR.

In an answer to Clarke and McCann, Houston and Weiss [19] developed a model based on the theoretical rationale provided by Kuehn, McGuire and Weiss [21]. According to them, the coefficient of lagged sales is interpreted as the proportion of consumers who are habitually repeating a purchase. Thus, the carry-over effect is modelled directly rather than through a geometric decay of advertising effectiveness as in Palda. Houston and Weiss use a nonlinear least squares method to obtain the maximum likelihood estimates of parameters. They claim to find (i) the presence of serially correlated error terms and (ii) important carry-over effects associated with advertising expenditures, although not necessarily of Koyck type.

The dynamics of the sales-advertising relationship is still not fully understood. This is possibly because (i) the estimation techniques used to date have not dealt adequately with correlated errors (ii) autoregressive moving-average (ARMA) models rather than impulse responses have been the object of the estimation exercises and (iii) the data blocks are very short considering the number of parameters that may be necessary to represent the dynamical relationships involved. A new technique which is now being made available is the Cholesky Least Squares (CLS) method [7]. This algorithm is described in the next section and is applied to the Lydia-Pinkham data in Section 4.

3. Modelling Methodology

3.1 *Impulse Responses and Rational Transfer Functions*

We take as our basic system model a linear system with the input-output relationship

$$(3.1.1) \qquad Y_t = \alpha_0 U_t + \alpha_1 U_{t-1} + \alpha_2 U_{t-2} + \dots, \quad \text{for all } t,$$

where Y is the output process and U the input process. Clearly a unit impulse input sequence $U = \{1, 0, 0, \dots\}$ commencing at the instant t yields the impulse response $S = \{\alpha_0, \alpha_1, \alpha_2, \dots\}$ at the system output commencing at the instant t. This, incidentally, is also the starting point for Palda's discussion of distributed lag systems.

We can compactly represent the entire history of the output sequence in response to the entire history of the input sequences using z-transforms. Let

$$\alpha(z) \triangleq \sum_{i=0}^{\infty} \alpha_i z^i, \quad Y(z) \triangleq \sum_{i=-\infty}^{\infty} Y_i z^i$$

and

$$U(z) \triangleq \sum_{-\infty}^{\infty} U_i z^i,$$

then it is easy to see that by equating powers of z^i we obtain

$$Y(z) = \alpha(z) U(z). \tag{3.1.2}$$

Notice that the system (3.1.2) *must* be non-anticipative because $\alpha(z)$ contains no negative powers of z (see e.g. [13, 11]).

The obvious problem with (3.1.1) and (3.1.2) from a conceptual and computational viewpoint is that the sequence $\alpha = \{\alpha_0, \alpha_1, \ldots\}$, equivalently the power series $\alpha(z)$, requires an infinite sequence of real numbers for its description. There are two useful solutions to this problem: (i) Assume that the magnitude of the α_i terms decay and hence truncate $\alpha(z)$ to $\alpha_M(z) = \sum_{i=0}^{M} \alpha_i z^i$, where $|\alpha_i| < \epsilon$ for $i > M$ for some small $\epsilon > 0$, (ii) Assume that the function $\alpha(z)$ is a rational function of z and hence that there exists a numerator polynomial $\beta(z) = \sum_{i=0}^{n} \beta_i z^i$ and denominator polynomial $\gamma(z) = \sum_{i=0}^{n} \gamma_i z^i$ (where n denotes the maximum degree of the two polynomials) such that $\alpha(z) = \beta(z)/\gamma(z)$. Notice that alternative (i) may be viewed as a special case of (ii) by setting $\beta(z) = \alpha_M(z)$ and $\gamma(z) = 1$.

If we assume that the system output (in our case sales) is disturbed by a noise process which has a rational spectrum, we obtain [11] the basic *rational transfer function representation*

$$Y(z) = \frac{\beta(z)}{\gamma(z)} U(z) + \frac{\mu(z)}{\nu(z)} \epsilon(z) \tag{3.1.3}$$

for the output process Y in terms of the observed input process U and the unobserved input noise process ϵ, where $E\epsilon_t\epsilon_s = \delta_{ts}$, where $\delta_{ts} = 1$ if $t = s$ and 0 otherwise.

When (3.1.3) is written in the form

$$\gamma(z)\nu(z) Y(z) = \beta(z)\nu(z) U(z) + \mu(z)\gamma(z)\epsilon(z) \tag{3.1.4}$$

we see that (3.1.4) is just the familiar auto-regressive moving average (ARMA) system model [6].

Most parameter estimation techniques for stochastic processes possessing ARMA representations also yield estimates of the systems impulse response. It is argued in [7] that estimates of a system impulse response are "robust" with respect to large alterations in the ARMA model orders; see also [24]. In complete contrast to this, estimates of ARMA model parameters vary greatly with respect to order specifications. The object, of course, in altering the orders of the polynomials in an ARMA model is to achieve the important modeling goal of better prediction performance and 'whiter' residuals. But we wish to emphasize that the *result* of this process is in fact better estimates of the underlying system *impulse response* and *not* closer estimates of the true ARMA model which may possess a totally different structure. This observation would appear to have consequences concerning the behavioural interpretation of ARMA models in several areas of econometrics in addition to our immediate concern of advertising models. This is discussed further in Section 5.3.

3.2 *The Cholesky Least Squares Algorithm*

The identification experiments on the Lydia-Pinkham data described in this paper use a recently developed parameter estimation algorithm. It is called the Cholesky Least Squares (CLS) algorithm and is described in greater detail in [7]. In this paper, we merely give a very brief description of its operation.

Following the notation in the previous subsection, assume that we have a single input single output system defined as follows:

$$a(z)Y(z) = b(z)U(z) + c(z)\epsilon(z) \tag{3.2.1}$$

where $\{Y(z)\}$ and $\{U(z)\}$ are the output and input sequences respectively, $\{\epsilon(z)\}$ is a white noise sequence, and, as before, $a(z)$, $b(z)$, $c(z)$ are the polynomials that describe the AR, MA, and noise MA coefficients of the system, respectively. Suppose $c(z) \not\equiv 1$; if we try to make estimate of the $a(z)$ and $b(z)$ polynomials using the original data and some standard least squares technique we will introduce bias into the estimates and the estimates are not consistent. A way to overcome this problem is to transform the data by operating with a filter $\hat{c}^{-1}(z)$, where $\hat{c}(z)$ is an estimate of $c(z)$. Let $Y^F(z) \triangleq \hat{c}^{-1}(z)Y(z)$ and $U^F(z) = \hat{c}^{-1}(z)U(z)$, then $\hat{a}(z)$ and $\hat{b}(z)$ are estimated by least squares on the filtered data $Y^F(z)$, $U^F(z)$; this results in

$$\hat{a}(z)Y^F(z) = \hat{b}(z)U^F(z) + \hat{\epsilon}(z) \tag{3.2.2}$$

where $\hat{\epsilon}(z)$ is the z-transform of the resulting residual sequence. If $\hat{c}(z) = c(z)$ and the orders of $\hat{a}(z)$ and $\hat{b}(z)$ equal the true orders of $a(z)$ and $b(z)$, respectively, then standard least squares theory shows that the resulting estimates of $a(z)$ and $b(z)$ are consistent.

In the CLS method a sequence of estimates $\hat{c}(z)$ of $c(z)$ is obtained by directly computing the Cholesky factors[2] of the autocovariance matrix of the residuals $w(z)$ [26, 22, 7], where $w(z)$ is defined by

$$w(z) = \hat{a}(z)Y(z) - \hat{b}(z)U(z), \tag{3.2.3}$$

[2] A Cholesky factor of a symmetric positive definite matrix R is that unique upper triangular matrix L with positive elements on the main diagonal such that $LL^T = R$.

and the autocovariance matrix R_w is such that the i^{th} diagonal is given by an estimate of $E(w(j)\,w(j+i))$. At each iteration of the algorithm the filter $\hat{c}(z)$ generated at its previous iteration is used to generate $Y^f(z)$ and $U^f(z)$. $\hat{a}(z)$ and $\hat{b}(z)$ are then estimated from (3.2.2) using a least squares algorithm. $w(z)$ and $\hat{c}(z)$ are then generated from (3.2.3) and the entire process repeated until a convergence criterion is satisfied. We set $\hat{c}(z) = 1$ initially.

Clearly when the orders of $a(z)$ and $b(z)$ are set to incorrect values this method, like all others, cannot be consistent. A theoretical analysis of the consistency of the technique and the behavior of the resulting estimates of the impulse response is presently being carried out. Highly satisfactory simulation results are presented in [7].

Model Order Determination

Several techniques have recently been devised to determine the appropriate model orders. These techniques are based essentially on some scheme that weighs the variance of the residuals of the fitted model against the total number of parameters fitted [1, 24, 12]. We have adapted Akaike's Final Prediction Error technique (FPE) [1] to fit into our CLS algorithm. The estimated FPE measure is computed as

$$FPE_{\hat{p}\hat{q}\hat{r}} = \frac{N + \hat{p} + \hat{q} + \hat{r} + 1}{N - \hat{p} - \hat{q} - \hat{r} - 1}\,\hat{\sigma}_\epsilon^2, \tag{3.2.4}$$

where N is the number of data points: $\hat{p}$ is the estimate of p, the true order of the AR part of the model; $\hat{q}$ is the estimate of q, the true order of the MA part of the model: $\hat{r}$ is the estimate of r, the true order of the noise regression, and $\hat{\sigma}_\epsilon^2$ is the sampled variance of the filtered data residuals, i.e.,

$$\hat{\sigma}_\epsilon^2 = \frac{1}{N}\sum_{t=1}^{N} \hat{\epsilon}_t^2 = \frac{1}{N}\sum_{t=1}^{N}\left[\frac{\hat{a}(z)}{\hat{c}(z)}\,Y(z) - \frac{\hat{b}(z)}{\hat{c}(z)}\,U(z)\right]_t^2$$

In the complete CLS algorithm p and q are alternately increased up to a value of 5 and for each pair of (p, q) the FPE is computed for values of r up to 4. (The values 5 and 4 are, of course, arbitrary.) The set of values yielding the lowest valued FPE are taken as the estimates of p, q and r.

Multi-Output Systems

A multivariable version of the CLS algorithm is used in the causality test in Section 4. It estimates the polynomials $a_1(z), \ldots, a_{n_0}(z)$ and the matrix of polynomials $B(z)$ by identifying row by row the model

$$A(z)\,Y(z) = \operatorname{diag}(a_i(z))\,Y(z) = B(z)\,U(z) + C(z)\epsilon(z)$$

using the single output program repeatedly, where n_0 is the dimension of the output process Y. The matrix $C(z)$ is then estimated by taking a Cholesky factor of the covariance of the vector residual sequence $W(z) \triangleq A(z)Y(z) - B(z)U(z)$ with a suitable number of diagonals.

4. Computational Results

In this section, we present the results of the application of the Cholesky Least Squares algorithm to the Lydia-Pinkham data [23, p. 23]. This data set consists of annual end of year sales and advertising for the years 1907 to 1960 inclusive. This provides a set of 54 data points. In terms of thousands of dollars the sales have a maximum value of 3438, a minimum value of 921 and a mean value of 1840. For the advertising expenditures these figures are 1941, 339 and 941, respectively.

In addition to advertising expenditure, we followed Palda in taking three dummy, or "off"–"on", variables as causal factors. These corresponded to three successive periods in which the quality of the advertising copy for the vegetable compound was significantly different. In turn these periods correspond to times at which the company made different claims in its advertising copy in response to varying policies of the FDA and later the FTC. Palda's three dummy variables were labeled $D1$–$D3$ and were prescribed as follows: $D1$ has the value 1 from 1908 to 1914 and was zero otherwise, $D2$ has the value 1 from 1915 to 1925 and was zero otherwise and $D3$ has the value 1 from 1926 to 1940 and was zero otherwise.

We did not follow Palda in taking either disposable income or a time trend variable as causative or explanatory variables. This was due to the fact that out of Palda's five best models with which we were concerned, only two contained time trend, and of these only one contained disposable income.

All the time series which we used were centered by subtracting off their mean values before any regressions were computed. This merely has the effect of removing the constant term which appears on the right hand side of Palda's equations.

4.1 *Log-advertising Related to Sales*

In this section U denotes the centered log-advertising variable and S the centered sales variable. The D^i variables are also centered. Note that in the present version of the CLS algorithm, standard errors of the parameter estimates are not available. These will be generated by a forthcoming modified version of the algorithm.

Order and Parameter Estimation Using the Cholesky Least Squares Technique

(i) For the initial values $p = 1, q = 0$ we obtained $r = 4$ and the model:

(4.1.1)
$$S_t - 0.615S_{t-1} = 822.175U_t - 179.859D_t^1 + 146.461D_t^2 - 65.683D_t^3 + 145.355\epsilon_t + 40.408\epsilon_{t-1} + 7.857\epsilon_{t-2} + 0.561\epsilon_{t-3} + 5.799\epsilon_{t-4}$$

(ii) The next model was obtained when the Akaike FPE attained a local minimum [7] and the residual error sequence was found to be uncorrelated at the 5% level [2, 6]. The values of p, q, r were 3, 1, 4 and the resulting model was

(4.1.2)
$$\begin{aligned} S_t - 1.054S_{t-1} + 0.149S_{t-2} + 0.173S_{t-3} &= 1434.133U_t \\ &-1108.522U_{t-1} - 331.682D_t^1 + 169.021D_{t-1}^1 \\ &-342.027D_t^2 + 438.634D_{t-1}^2 - 65.544D_t^3 + 63.876D_{t-1}^3 \\ &+ 159.934\epsilon_t - 56.259\epsilon_{t-1} - 36.503\epsilon_{t-2} - 6.774\epsilon_{t-3} \\ &- 0.941\epsilon_{t-4} \end{aligned}$$

(iii) The final model was obtained when both the Akaike FPE and the prediction error variance attained global minima [7]. However, this model did not pass the whiteness test in contrast to model (ii) above. The values of p, q, r were 5, 3, 4.

(4.1.3)
$$\begin{aligned} S_t - 0.770S_{t-1} + 0.010S_{t-2} + 0.210S_{t-3} - 0.167S_{t-4} \\ + 0.173S_{t-5} = +1375.339U_t - 623.683U_{t-1} \\ + 64.019U_{t-2} - 150.958U_{t-3} - 190.836D_t^1 \\ + 45.885D_{t-1}^1 - 278.126D_{t-2}^1 + 222.961D_{t-3}^1 \\ -201.367D_t^2 - 135.333D_{t-1}^2 + 508.116D_{t-2}^2 \\ -45.705D_{t-3}^2 + 4.104D_t^3 - 176.684D_{t-1}^3 + 277.849D_{t-2}^3 \\ - 92.182D_{t-3}^3 + 122.630\epsilon_t + 13.065\epsilon_{t-1} - 9.353\epsilon_{t-2} \\ - 36.522\epsilon_{t-3} - 9.947\epsilon_{t-4} \end{aligned}$$

4.2 *Causality Experiments*

There has been some discussion in the literature on the topic of the direction of causality between sales and advertising expenditures [4].

Over the last few years a rigorous theory of causality between stationary vector time series has been produced. We shall not go into that in detail here but shall refer the reader to the papers by Granger [17], Sims [30], Pierce and Hough [25], Caines [8], Caines and Chan [9, 10], and Wall [31].

Granger's [21] original definition of causality between two stationary stochastic processes is as follows: the process U drives, or is causal to, the process S if and only if

$$\hat{U}_{t+k} \mid \{U^t, S^t\} = \hat{U}_{t+k} \mid \{U^t\}, \qquad \text{for all } k \geq 1,$$

where $\hat{X}_{t+k} \mid \{Z^t\}$ denotes the linear least squares estimate of the process X at the instant $t + k$ given the observations $\{Z^t\} \triangleq (Z_t, Z_{t-1}, Z_{t-2}, \ldots)$. This condition has the interpretation that U is causal to S if and only if estimates of future behavior of U given the past history of U and S are equal to the estimates given only to past history of U. Sims [30], Caines [8], and Caines and Chan [9,10] produced a set of equivalent formulations of this notion. One of these (see [9, 10]) states that U is causal to S if and only if the non-anticipative linear least squares (Wiener) filter estimating S_t from the observations $\{U^t\}$ is identical to the anticipative filter which uses the observations $\{U^\infty\}$ i.e.

$$\hat{S}_t \mid \{U^t\} = \hat{S}_t \mid \{U^\infty\}.$$

In [9, 10] an important equivalent operational definition is given in terms of the innovations representation of the joint $\begin{bmatrix} S \\ U \end{bmatrix}$ process. Consider the innovations [32] or Wold [33] representation of the bivariate process $\begin{bmatrix} S \\ U \end{bmatrix}$ with respect to the bivariate orthogonal process ϵ:

$$\begin{bmatrix} S \\ U \end{bmatrix}_t = \epsilon_t + \Phi^1 \epsilon_{t-1} + \Phi^2 \epsilon_{t-2} + \cdots \tag{4.2.1}$$

with $E\epsilon_t \epsilon_s^T = \Sigma \delta_{ts}$, where Σ is a 2×2 matrix, and where the superscript T denotes transpose. Then the process U drives, or is causal to, the process S if the representation (4.2.1) has an upper triangular structure, i.e.

$$\begin{bmatrix} S \\ U \end{bmatrix}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \epsilon_t + \begin{bmatrix} \phi_{11}^1 & \phi_{12}^1 \\ 0 & \phi_{22}^1 \end{bmatrix} \epsilon_{t-1} + \begin{bmatrix} \phi_{11}^2 & \phi_{12}^2 \\ 0 & \phi_{22}^2 \end{bmatrix} \epsilon_{t-2} + \cdots.$$

We remark that in [8, 9, 10] this relationship is called the "feedback-free" relation for reasons which are explained in those papers.

The bivariate version of the CLS algorithm [7] identifies uniquely defined models of the form

$$\begin{bmatrix} a_1(z) & 0 \\ 0 & a_2(z) \end{bmatrix} \begin{bmatrix} S \\ U \end{bmatrix} = \begin{bmatrix} \phi_{11}(z) & \phi_{12}(z) \\ \phi_{21}(z) & \phi_{22}(z) \end{bmatrix} \begin{bmatrix} \epsilon^1 \\ \epsilon^2 \end{bmatrix}, \tag{4.2.2}$$

with $$E\begin{bmatrix}\epsilon^1\\ \epsilon^2\end{bmatrix}_t [\epsilon^1 \quad \epsilon^2]_s = \Sigma\delta_{ts},$$

where all the displayed functions of z in (4.2.2) are polynomials. Now, as observed also by Wall [31], it is clear that U drives S if and only if $\phi_{21}(z)$ is zero in this model.

We wish to emphasize that in both the theory and application of this causality criterion the universe of observed stochastic processes is just the set $\{S, U\}$. The questions whether (i) driving occurs via an intermediate process, or (ii) both S and U are in some sense driven by a third process, are not answered within this framework. Such questions may be the most important concerns for an experimenter [16]. The fact that the unique (cannonical) innovations representation of the process $\begin{bmatrix}S\\ U\end{bmatrix}$ is used in the definition above means that the theoretical question of the direction of causality between S and U may be answered unambiguously. In particular, the "observational equivalence" ambiguities in the causality tests cited by Basmann [3] do not occur with our formulation. Finally, we wish to point out that the identification of a model of the form (4.2.2) does not in any way prejudice the conclusion concerning the direction of causality. In fact both hypotheses "S causes U" and "U causes S" are nested within the hypothesis "S and U are in general relation" (i.e., all $\phi_{ij}(z)$ in (4.2.2) are non-zero). We refer the reader to Haavelomo [18] for an earlier discussion of the value of nested hypotheses in model estimation when the direction of causality is an issue.

In our experiment the CLS algorithm was used to estimate a model of the form (4.2.2) for the centered bivariate sales and log-advertising series. This resulted in the following model of the form (4.2.2):

$$\begin{bmatrix}(1-0.830z) & 0\\ 0 & (1-0.784z)\end{bmatrix}\begin{bmatrix}S\\ U\end{bmatrix}_t = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\epsilon^1\\ \epsilon^2\end{bmatrix}_t + \begin{bmatrix}0.289 & 1309.170\\ 0.000 & 0.022\end{bmatrix}\begin{bmatrix}\epsilon^1\\ \epsilon^2\end{bmatrix}_{t-1} + \begin{bmatrix}0.214 & 608.297\\ 0.000 & -0.187\end{bmatrix}\begin{bmatrix}\epsilon^1\\ \epsilon^2\end{bmatrix}_{t-2} + \begin{bmatrix}0.252 & -390.781\\ 0.000 & -0.153\end{bmatrix}\begin{bmatrix}\epsilon^1\\ \epsilon^2\end{bmatrix}_{t-3} + \begin{bmatrix}0.156 & -113.200\\ 0.000 & 0.159\end{bmatrix}\begin{bmatrix}\epsilon^1\\ \epsilon^2\end{bmatrix}_{t-4}, \tag{4.2.3}$$

with $$E\begin{bmatrix}\epsilon^1\\ \epsilon^2\end{bmatrix}_t [\epsilon^1\,\epsilon^2]_t = \begin{bmatrix}28135.129 & 6.266\\ 6.266 & 0.007\end{bmatrix}$$

Despite the small variance shown by the Σ_{22} term, we regard this as a very important result. It dramatically reveals a causal effect between the centered log-advertising and centered sales time series.

In the causality detection experiments on Gross Domestic Product and Unemployment described in [8, 9, 10], various statistical tests were employed. Analogous tests cannot be carried out here because of the lack of estimates of standard errors of parameter estimates in (4.2.3). The desired statistical tests will be performed later using a modified version of the CLS algorithm. This should also reveal whether the causal relationship detected above is strong or weak (see [8] where strong causality is defined as a causal relationship without instantaneous feedback.) We would like to remark, however, that the causality effect in (4.2.3) appears to be more powerful than the corresponding relationship from GDP to Unemployment obtained in [9, 10].

An identical experiment using the centered advertising and centered sales series does not reveal a significant causal or driving effect. This should not cause too much surprise since the causal relationship we are investigating is defined in terms of linear process representations. We remark that a nonlinear generalization of this theory has recently been produced [34]. The fact that the logarithmic transformation of advertising expenditures yields such a strong causal relationship shows that it yields a suitable variable for a *linear* causal relationship. This suggests the hypotheses that (i) sales are related in a direct linear fashion to a measure of *effective* advertising and *not* dollar advertising; (ii) a measure of this effective advertising is a diminishing returns to scale function such as logrithm. (The diminishing returns to scale effect is, of course, generally accepted as being reasonable.)

We remark that other empirical studies, including Benjamin and Maitland [5], have used the logarithmic transformation of advertising expenditures as a resonable measure of effective advertising. Futher, Sethi [28] has obtained optimal advertising policies for this type of model.

5. Interpretation of Results and Conclusion

5.1 *Log-Advertising related to Sales*

We next compare our log-advertising to sales models with the KOYLI and KOYLDIF models of Palda. With U denoting the log-advertising variable, the KOYLI Model is as follows:

$$(5.1.1)\qquad \underset{}{S_t} - \underset{(0.150)}{0.633S_{t-1}} = -2924 + \underset{(564)}{1226\, U_t} - \underset{(186)}{20D_t^1} + \underset{(130)}{215D_t^2} - \underset{(126)}{164D_t^3}$$

Comparing our CLS model (4.1.1) with (5.1.1) we see that they all possess an a_1 coefficient in the neighborhood of -0.6. In fact the estimates of a_1 differ only by 3% in the models (4.1.1) and (5.1.1).

TABLE 1

COMPARISON OF LOG-ADVERTISING TO SALES TRUNCATED IMPULSE RESPONSES

ARMA Model Order = (p, q, r) \ S_t =	Impulse Response Coefficients*									
	U_t	U_{t-1}	U_{t-2}	U_{t-3}	U_{t-4}	ϵ_t	ϵ_{t-1}	ϵ_{t-2}	ϵ_{t-3}	ϵ_{t-4}
KOYL1 (5.1.1) (1.0.0)	1226	776	491	311	197					
KOYLDIF (5.1.2)	1326	697	366	193	102					
Model (4.1.1) (1.0.4)	822	507	311	191	118	145	130	88	54	39
Model (4.1.2) (3.1.4)	1434	403	211	−86	−191	160	112	58	10	−27
Model (4.1.3) (5.3.4)	1375	435	385	−147	21	123	107	72	−8	−19

*In this table U denotes the log-advertising variable. Here we ignore the D^i terms for each of the models.

The KOYLDIF model is:

$$(5.1.2) \qquad S_t - 0.370S_{t-1} = \underset{}{-1903} + \underset{(0.170)}{0.527}\,(S_{t-1} - 0.370S_{t-2})$$

$$+ \underset{(552)}{1326}\,(U_t - 0.370\,U_{t-1}) - \underset{(168)}{41}D^1 + \underset{(120)}{165}D^2 + \underset{(114)}{108}D^3$$

In order to compare the respective truncated impulse responses of all the log-advertising models we present them in Table 1.

The reader should notice that the operation in the KOYLDIF model of differencing both the S and U series by $(1 - 0.370z)$ leaves the U to S impulse response unchanged. The purpose of this differencing in Palda's model is to obtain a suitably white residual sequence for the data on which to apply the least squares technique. However, as explained in Section 3.2, such operations are carried out in a more flexible and automatic manner by the CLS method.

From Table 1 we see that the impulse response of (4.1.1) differs significantly from those of (5.1.1) and (5.1.2), whilst the impulse responses of these latter two models are seen to resemble one another.

A comparison of the impulse response of the CLS model (4.1.2) with those of the models (5.1.1) and (5.1.2) shows a difference of 15% and 18% respectively between the leading terms, as a fraction of this term in (4.1.2). These terms are, of course, identical to the b_0, or the 'feedthrough' terms, in the respective models.

The final model (4.1.3) chosen by Akaike's FPE criterion would appear to be over parameterized. However, it is significant that the impulse responses of (4.1.2) and (4.1.3) agree to within 4% on the first term and 7%

on the second term. Furthermore, the leading terms of (4.1.3) and (5.1.1) and (5.1.2) differ by 11% and 4% respectively, and the third terms, i.e., the coefficient of U_{t-2} differ by 8% and 2% respectively, where all percentages are taken with respect to the leading term of (4.1.3).

The conclusion we draw from the results above is that there is rough agreement between the models described by Palda and the models (4.1.2), (4.1.3) presented above, given the limited amount of data available. Note that the Akaike criterion, which is known to work excellently for large samples, chose high order models for the CLS algorithm. On the face of it the high order $a(z)$ polynomials which were chosen deny the Koyck hypothesis of a first order geometric decay of the impulse response.

We have said that we view all of the estimation methods in this paper as being basically impulse response identification methods. Therefore we regard it as an important fact that the decay of the impulse response for the models (4.1.2) and (4.1.3) chosen by the CLS algorithm was approximately 30% after one year and 15–30% in the second year as a fraction of the leading terms. Furthermore, the succeeding term is negative in both models, but is an order of magnitude smaller than the leading term. We could therefore say that these two models indicate a positive carryover effect of two years beyond the current year. (It is an important fact that this observation is corroborated by the bivariate model (4.2.3)). The estimated impulse responses for models (4.1.2) and (4.1.3) imply that advertising effectiveness decays in a different fashion than the constant 60% rate of decay obtained by Palda.

It is interesting to note that the immediate gain factor of the impulse responses of (4.1.2) and (4.1.3) is approximately 1350. This means, for example, that an increment from the mean in advertising expenditure of \$10,000 results in, approximately, an immediate *increase* in sales, from its mean, of \$50,000. The respective figure for an advertising expenditure increment of \$100,000 is \$103,000. This has implications for optimal advertising policies. The reader is referred to Sethi [29] for a comprehensive survey of dynamic optimal control models in advertising.

5.2 *Advertising Related to Sales*

An identical experiment was performed with respect to advertising expenditure in thousands of dollars. We shall not describe this experiment in detail but merely give the results in Table 2. This is for two reasons: (i) space limitations and (ii) the fact that significant causality results were not obtained for this case.

5.3 *Predictive versus Explanatory Models*

In this conclusion a remark is in order concerning "explanatory" versus "predictive" models. If it is taken as a basic assumption that the

TABLE 2

COMPARISON OF ADVERTISING TO SALES TRUNCATED IMPULSE RESPONSES

ARMA Model Order = (p, q, r) \ $S_t =$	Impulse Response Coefficients*									
	U_t	U_{t-1}	U_{t-2}	U_{t-3}	U_{t-4}	ϵ_t	ϵ_{t-1}	ϵ_{t-2}	ϵ_{t-3}	ϵ_{t-4}
KOYK (1,0,0)	0.537	0.337	0.212	0.133	0.084		—	—	—	
Model A (1,0,4)	0.381	0.245	0.157	0.101	0.065	175	161	104	76	59
Model B (2,0,4)	0.426	0.384	0.215	0.075	0.001	174	157	80	44	41
Model C (3,1,3)	0.533	0.119	−0.061	−0.047	−0.015	170	142	95	69	49

*In this table, U denotes the advertising expenditure variable. Here also, we ignore the coefficients of the D^t terms for each of the models.

advertising to sales impulse response cannot have negative terms, i.e., if it is assumed that those terms of behavioral significance in the explanatory models cannot be negative, then techniques such as least squares or Cholesky Least Squares estimation cannot in general be used. This is because, in general, rational transfer functions will yield impulse responses unrestricted in sign. The notable exception is that in which the $a(z)$ polynomial is of order one, with a_1 negative, and in which $b(z)$ has positive coefficients. Notice that all of Palda's models fall in this special class. In general, one must either abandon the attachment to "explanatory" models in the search for models which constitute the best predictors, or carry out a constrained parameter search when finding the best predictive model which (i) lies in a certain model class (i.e. for given orders of the terms in the ARMA representation) *and* (ii) has an impulse response which is positive for a prescribed number of terms. In principle this task can be accomplished by using constrained optimization routines in the identification algorithms. Of the two alternative courses of action described above we have obviously chosen the former. However, as this point is not apparently discussed in the literature, it is perhaps worth reiterating that the choice of ARMA models with the degree of $a(z)$ greater than or equal to 2 will in general yield models that are open to the objection that they do not satisfy the required positivity assumption of the "explanatory" models.

5.4 *Causality*

Finally we come to the impressive results of the causality analysis. As we said in Section 4.2, we regard the numerical results for the causality detection experiment between log-advertising and sales as being quite remarkable. They show an extremely strong linear driving relationship between log advertising and sales. The behavioural interpretation is that, measured in terms of the logarithmic transformation, advertising expenditures in thousands of dollars drive sales measured directly in thousands of dollars. This would indicate that advertising expenditure decisions by the Lydia-Pinkham management caused variations in their sales performance while the converse did not occur. This is in contrast to the practice of basing advertising expenditures on a fixed percentage of sales, as has been shown [27] to occur in the automobile industry.

Harvard University
University of Toronto
University of Hawaii

Revised December 1976

REFERENCES

[1] Akaike, H., "Statistical Predictor Identification," *Ann. Inst. Statist. Math.*, 22, 1970, pp. 203–217.

[2] Åström, K. J. & Eykhoff, P., "System Identification: A Survey," *Automatica*, Vol. 7, 1971, pp. 123–169.

[3] Basmann, R. L., "A Note on the Statistical Testability of 'Explicit Causal Chains' Against the Class of 'Interdependent' Models," *JASA*, 1965, pp. 1080–1093.

[4] Bass, F. M., "A Simultaneous Equation Regression Study of Advertising and Sales of Cigarettes," *Journal of Marketing Research*, Vol. VI, August 1969, pp. 291–300.

[5] Benjamin, B. & Maitland, J., "Operational Research and Advertising: Some Experiments in the Use of Analogies," *Operations Research Quarterly IX*, Sept. 1958, pp. 207–17.

[6] Box, G. E. P. & Jenkins, G. M., *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco, 1970.

[7] Brotherton, T. W. & Caines, P. E., "Impulse Response Identification Via the Cholesky Least Squares Method," Control Systems Report No. 7606, University of Toronto, 1976. Presented at *Conference on Decision and Control*, Clearwater Beach, Florida, December 1976.

[8] Caines, P. E., "Weak and Strong Feedback Free Processes," *IEEE Transactions on Automatic Control*, Vol. AC-21, No. 5, October 1976.

[9] Caines, P. E., & Chan, C. W., "Feedback Between Stationary Stochastic Processes," *IEEE Trans. Auto. Control*, Vol. AC-20, No. 4, August, 1975.

[10] Caines, P. E. & Chan, C. W., "Estimation, Identification and Feedback," *System Identification: Advances and Case Studies* (D. G. Lainiotis & R. K. Mehra, Eds.), Marcel Dekker, Inc., New York, 1976.

[11] Caines, P. E. & Wall, K. D., "Theoretical Foundations for Methods of Estimation and Policy Optimization in Economic Management," Chapter 13 in *Modelling the Economy* (G. A. Renton, Ed.) Crane, Russak and Company, Inc., New York, December 1974.

[12] Chan, C. W., Harris, C. J. & Wellstead, P. E., "An Order Testing Criterion for Mixed Autoregressive Moving Average Processes," Inst. of Science and Technology Report No. 229, University of Manchester, December 1973.

[13] Chen, C. T., *Introduction to Linear System Theory*, Holt Rinehart & Winston, 1970, New York.

[14] Clarke, D. G., "Econometric Measurement of the Duration of the Advertising Effect on Sales," Working Paper, Division of Research, Harvard Business School, Boston, 1975.

[15] Clarke, D. G. & McCann, J. M., "Measuring the Cumulative Effects of Advertising: A Reappraisal," *Combined Proceedings* (ed. T. V. Greer), American Marketing Association, Chicago, 1973.

[16] Cochran, W. G., "The Planning of Observational Studies in Human Populations," *J. Royal Stat. Soc.*, Series A, Part 2, 1965, pp. 234–266.

[17] Granger, C. W. J., "Economic Processes Involving Feedback," *Information and Control*, Vol. 6, 1963, pp. 28–48.

[18] Haavelomo, T., "Methods of Measuring the Marginal Propensity to Consume," *JASA*, 1947, pp. 105–122.

[19] Houston, F. S. & Weiss, D. L., "Cumulative Advertising Effects: The Role of Serial Correlation," *Decision Sciences*, Vol. 6, 1975, pp. 471–481.

[20] Koyck, L. M., *Distributed Lags and Investment Analysis*, North-Holland, Amsterdam, 1965.

[21] Kuehn, A., McGuire, T. W. & Weiss, D. L., "Measuring the Effectiveness of Advertising," in *Science Technology and Marketing* (R. M. Hass, Ed.), American Marketing Association, Chicago, 1967.

[22] Morf, M., "Fast Algorithms for Multivariable Systems," Ph.D. Dissertation, Stanford University, California, August 1974.

[23] Palda, K. S., *The Measurement of Cumulative Advertising Effects*, Prentice-Hall, Englewood-Cliffs, New Jersey, 1964.

[24] Parzen, E., "Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes." Technical Report No. 23, State University of New York at Buffalo, July 1975.

[25] Pierce, D. A. & Hough, L. D., "The Assessment and Detection of Causality in Temporal Systems." Report of Division of Research and Statistics, Federal Reserve System, Washington, January 1975.

[26] Rissanen, J., "Algorithms for Triangular Decomposition of Block Hankel and Toeplitz Matrices with Applications to Factoring Positive Matrix Polynomials." *Mathematics of Computation*, Vol. 27, No. 21, January 1973.

[27] Schmalensee, R., *The Economics of Advertising*, North-Holland, Amsterdam, 1972.

[28] Sethi, S. P., "Optimal Control of a Logarithmic Advertising Model." *Operational Research Quarterly*, Vol. 26, No. 2, 1975, pp. 317–319.

[29] Sethi, S. P., "Dynamic Optimal Control Models in Advertising: A Survey." Forthcoming in *SIAM Review*.

[30] Sims, G. A., "Money, Income and Causality." *American Economic Review*, Vol. 62, 1972, pp. 540–552.

[31] Wall, K. D., "An Application of Simultaneous Estimation to the Determination of Causality Between Money and Income." Report of the Programme for Research into Econometric Methods, University of London, March 1974.

[32] Wiener, N. & Masani, P., "The Prediction Theory of Multivariate Stochastic Processes, I." *Acta Mathematica 98*, 1957, pp. 111–150.

[33] Wold, H., *A Study in the Analysis of Stationary Time Series*, (2nd ed.). Almquist and Wiksell, Stockholm, 1954.

[34] Zames, G., Private Communication.