FORECASTING AND ASSESSING STRUCTURAL CHANGE USING THE BAYESIAN DYNAMIC LINEAR MODEL:

THE CASE OF PROMOTIONAL CAMPAIGNS

Andy Pole

ABSTRACT

We discuss and illustrate modeling and analysis of structural change in time series data. The modeling framework is the normal dynamic linear model (Pole, West and Harrison 1994). We utilize basic ideas of componentwise model structuring and subjective intervention to provide a natural, easy-to-implement system for dealing with promotional campaigns.

I. INTRODUCTION

We discuss and illustrate the modeling and analysis of structural change in time series data. The topic is treated in the context of a general

Advances in Econometrics, Volume 11, Part B, pages 185–211.
Copyright © 1996 by JAI Press Inc.
All rights of reproduction in any form reserved.
ISBN: 1-55938-974-5

forecasting system employing dynamic linear models (Pole et al. 1994). Specifically, our concern is with modeling short-lived promotional campaigns when forecasting sales of a regularly purchased consumable such as a foodstuff.

Existing approaches are characterized by extensions (such as indicator variables) to standard models (Box and Tiao 1975), both model and extension being poorly suited to coping with evolutionary change, or complicated nonlinear structures, which are extremely difficult to apply (Migon and Harrison 1985). By contrast, the methodology proposed here is a coherent generalization of an existing model class, the dynamic linear model, and is both conceptually and practically straightforward. Other benefits are unambiguous interpretation of estimated model components and the resulting clear picture of overall sales performance and advertising effectiveness.

We begin the discussion with an example of sales forecasting, highlighting the setting of formal models as one component in a complete forecasting system. Next, the statistical modeling framework of the dynamic linear model (DLM) is described, setting into a formal modeling context a general approach to dealing with nonroutine events. Then the specific proposal for modeling and assessing short-lived promotional campaigns in sales forecasting utilizing the DLM is detailed. The method is applied successfully to sales of a soft drink product.

A. Forecasting Systems

Consider forecasting sales for a product in a competitive market. While the market is in equilibrium a firm will forecast its future monthly sales using the company's routine forecasting model. Now consider what happens when a market disequilibrium occurs. As a particular example, suppose that the manufacturing plant of a competitor is destroyed. There is no possibility (sabotage aside) that such an event will be part of the firm's routine forecasting model, and therefore no way that the model can reflect the new market conditions. The forecaster, however, is able to make such adjustments. Assessing existing knowledge of pertinent factors such as the competitor's market share and availability of alternative production facilities will enable the forecaster to form a judgment about the likely effect of the competitor's misfortune on his own firm's sales. Cast in terms of probability statements about the forecast series, the assessment facilitates a new forecast, one that combines the routine model information with the implications of the unusual event.

Two points from the example are notable. The first is that formal models are only one component of a forecasting system. The second is that subjective judgment is very important. The following discussion expands on both points.

A forecast system may include several models but, crucially, it also includes the forecaster. The forecaster is human, an intelligent and knowledgeable overseer. A forecasting system is not tangible in the sense of it being possible to buy it off the shelf. In the same way that an airplane is not a transport system until a pilot takes control, a forecast model is not a forecast system until the forecaster takes control.

The parallel drawn between pilot and forecaster offers further illumination. In routine flying an airplane operates on automatic, the pilot keeping an eye on things from time to time. When bad weather strikes, the pilot takes personal control of the airplane, overriding the suggestions of the automatic pilot as he sees fit. In routine forecasting, model forecasts are used like the autopilot in an airplane. When an unusual event occurs, however, the forecaster takes direct control. Circumstances surrounding the event are analyzed and the model forecast is adjusted in an appropriate way. There is no fully detailed formal model for the quantification of the overriding action—that would just be an elaboration of the routine mechanistic model—although the method of overriding certainly is formal. Just as the method of pilot overriding the autopilot is formal: the pilot can bank, adjust the yaw, or change flight level—all of which the autopilot could do if it had determined such action appropriate. The point is that the pilot will do these things to different degrees than the model (autopilot) according to best judgment, moment by moment, about the weather. The pilot's judgment could, of course coincide with what the autopilot would have done. Similarly, reacting to unusual circumstances, the forecaster may choose to go with what the model is saying.

A pilot does not just sit in the cockpit and wait for bad weather to cause passengers to spill their coffee before taking action. The pilot is actively on the lookout for indications of potentially unfavorable conditions in order that timely preventive action may be taken. Warned of a storm in the flight path at the current flight level, the pilot may elect to move to a higher or lower flight level where the air is less disturbed. The pilot can simply program the change into the autopilot and return to a supervisory role. In the same way, a forecaster may become aware of changed market conditions in advance of the onset of those changes. An obvious example, and the one with which this chapter is concerned, is a marketing

campaign. Surely the company knows ahead of time about an advertising campaign it is planning. Just as surely this information will be communicated to the forecast system to change the currently prevailing environment on which the forecasts are predicated. There may be no change to the model form, perhaps a simple change to the expected level of sales over the next quarter. The model's projections will then include the extra information, just as the reprogrammed autopilot does, and the forecaster may resume the overseeing role.

Naturally, one should not take the parallel with the pilot-airplane example too literally. Uncertainties in the marketing case will typically remain greater as a result of the advertising activity; the forecaster will be closely scrutinizing forecast-outcome comparability and constantly revising judgment as information becomes available. The pilot does so too, but his scenario, as constructed here, is subject to less expected uncertainty.

The principle of *management by exception* has been coined to encapsulate the intelligent forecasting system—with both proactive and retroactive intervention in formal models—just described.

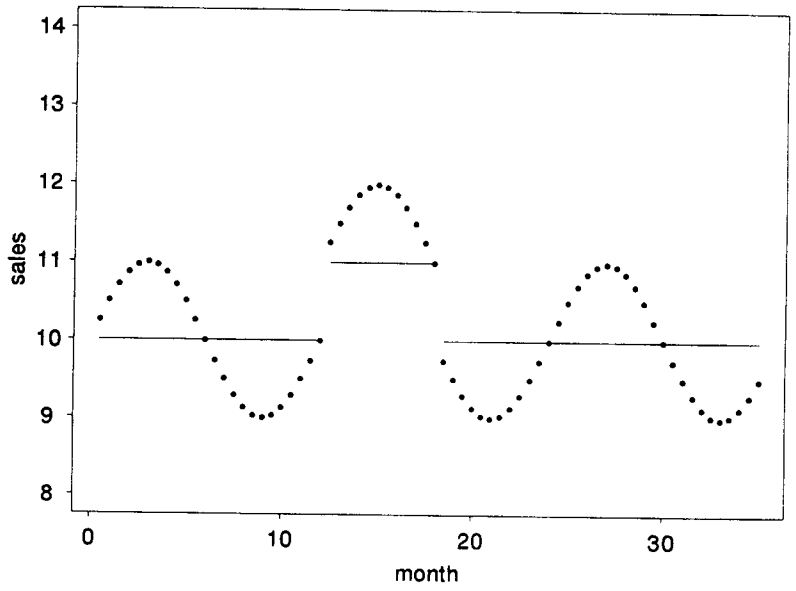

Figure 1. Seasonal sales series with a level increase corresponding to a short-term price reduction.

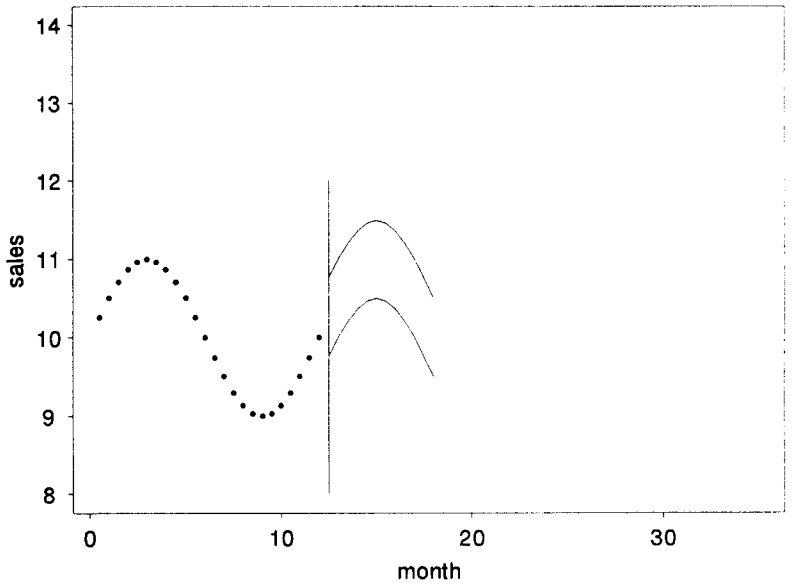

Figure 2. Sales forecast (range): before price reduction decision.

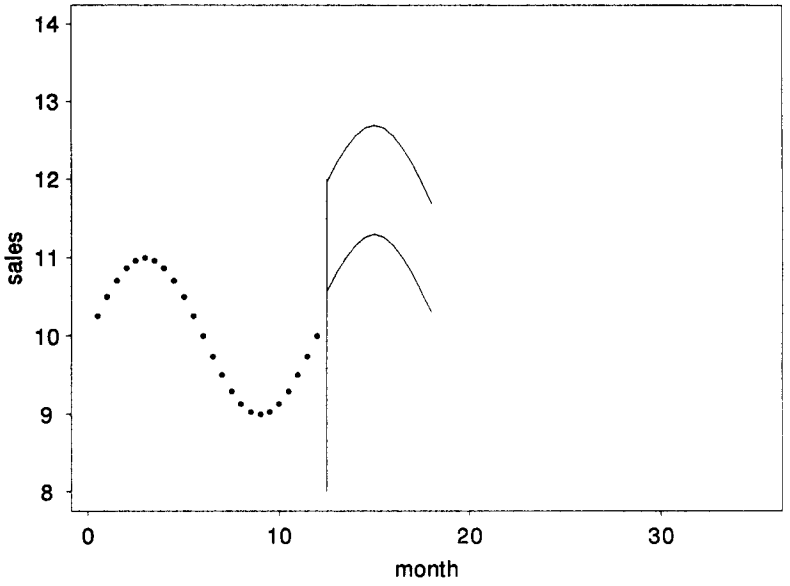

Figure 3. Modified sales forecast (range): including best guess of the effect of a price reduction.

Figure 1 shows an archetypal example. A seasonal sales series, with summer peak and winter trough, is perturbed by a positive response to a price reduction. Subsequently, with the price reduction removed, sales return to prepromotion levels. Figure 2 shows the routine model forecasts before the price reduction is made. Based upon the seasonal model form and fitting to historical data, the estimated annual cycle is simply propagated forward. Figure 3 shows a new set of forecasts, made after including the best judgment of the likely effect of the planned price reduction. The judgment in this case is that the underlying level of sales will increase but that the seasonal pattern will remain unaltered. Individual point forecasts are therefore simply increased by the amount of the level increase judged to be appropriate. Uncertainties are also increased, reflecting the lack of precision in the estimated level change.

B. Structural Change

Instances of change abound in commercial and economic time series. A few examples give the general picture.

A specific instance of a substantial change in seasonal pattern occurred in the series for numbers of marriages in the United Kingdom following changes in the tax laws announced in the 1968 budget. In the early 1960s if, in a given tax year, a couple married, then they were able to reclaim tax paid during that year as single people (the U.K. tax regime favoring considerably married couples over single people). This was a great incentive for couples to marry in the first quarter of a calendar year, which is the end of a tax year. The effect of the tax change was to remove that incentive, with the result that the yearly pattern of marriages changed to reflect the climate and holiday pattern in the United Kingdom, resulting in a pronounced trough in the first quarter (winter) where before there had been a seasonal peak. The underlying trend of marriages did not change at all. Analyses of this data are given in Harrison (1988) and West and Harrison (1989).

A dramatic example of trend change with seasonal stability, typically more common than the previous example, occurred during the 1970s, when the member countries of the oil-producing cartel, OPEC, quadrupled the price of crude oil. Among other notable effects was the marked impact on the number of miles traveled by car in the United Kingdom. Official figures clearly show a dramatic drop in level and a stifling of growth but no change in the yearly seasonal pattern. A related example is the number of injuries resulting from road traffic accidents in the

United Kingdom: there was a sharp drop in underlying level following the introduction of compulsory wearing of seat belts in 1983. The injury series is analyzed in Pole et al. (1994).

More radical types of structural change involve changes in the qualitative form of appropriate model rather than simply the abrupt quantitative changes of a constant qualitative form exemplified above. A common example is the life cycle of a consumable product: initial market growth is often linear as the product struggles to gain market recognition, then growth switches to an exponential pattern, and finally a return to a linear pattern as the market becomes saturated. Other examples include the changing nature of market dynamics, which cause previously useful explanatory variables to become unimportant and other variables to become influential. Changes in form such as these can be more or less enduring according to the nature of the underlying causes and subsequent market changes.

In the sequel we discuss formal modeling of anticipated, short-lived, structural change in time series. Specifically, modeling and analysis with dynamic linear models is extended to analysis of sales of a product that is subject to advertising campaigns. The development differs from existing discussion of intervention in dynamic linear forecast models. The existing literature (summarized in West and Harrison 1989) focuses on one-off changes to quantification of existing model components, much as we described for the change in level in the sales example and in the injuries series above. (Mention is also made of more general interventions, where the model form is altered, but no specific example or modeling has appeared to date.) It is noted in the cited references that an analysis may contain multiple interventions, but that is rather different from the situation we are concerned with here because those interventions are unrelated, except perhaps in a reactionary sense.

In the illustration (recall Figures 1–3) we deliberately omitted any discussion of activities at the end of the promotional year. Obviously one can go through the same procedure as the promotion draws to a close as was done before it was instigated: adjust the existing component quantifications in ways that seem appropriate—a reduction in underlying level, for instance. However, treating promotional campaigns by making interventions on existing model components in this way is somewhat obfuscatory, particularly when campaign duration is measured in periods of much less than a year. The underlying sales level during the campaign period is contaminated by the additional sales attributable directly to the campaign. Ideally, what we would like to do is to form an explicit

quantified view of the campaign contribution to sales at any point, in addition to a view of the underlying non-campaign-related sales pattern. We present an extension of the dynamic linear model that achieves precisely that goal. A transitory component designed specifically for the campaign effects is added to the routine model for just the period of the campaign. In this way, it is possible to clearly determine sales attributable to the campaign while maintaining a coherent view of the underlying sales model. Moreover, when the campaign has run its course, the standard retrospective analysis of the DLM provides estimates of what the campaign actually achieved, over and above what (most likely) would have been achieved in the absence of the campaign. Such information is extremely valuable for comparing the effectiveness of alternative types of promotion, or promotions in different markets.

The chapter proceeds as follows. In Section II we present the structure and analysis of the dynamic linear model. Fair detail is included since the modeling extensions described in Section III build directly on the basic DLM. A complete discussion is given in the book by Pole et al. (1994); that reference includes a copy of the BATS (Bayesian analysis of time series) computer program, which provides all of the modeling and analysis of the DLM as described in the book. In Section III we present the formal modeling extensions for analysis of transitory promotional campaigns. In Section IV we apply the model to an analysis of sales data for a soft drink product for which there was a month-long advertising campaign. Finally, in Section V we present a summary of this work.

II. THE DYNAMIC LINEAR MODEL

The general univariate dynamic linear model is written as

$$\text{Observation equation: } Y_t = F_t'\theta_t + \nu_t, \quad \nu_t \sim N[0, V_t],$$

$$\text{System equation: } \theta_t = G_t\theta_{t-1} + \omega_t, \quad \omega_t \sim N[0, W_t],$$

where Y_t denotes the observation series at time t; F_t is a vector of known constants (the regression vector); θ_t denotes the vector of model state parameters; ν_t is a normally distributed stochastic error term with zero mean and variance V_t; G_t is a matrix of known coefficients that defines the systematic evolution of the state vector across time; and ω_t is a normally distributed stochastic error term with zero mean and covariance matrix W_t.

The two stochastic series $\{\nu_t\}$ and $\{\omega_t\}$ are assumed to be temporally independent and mutually independent; that is, the covariances $\text{Cov}[\nu_t,\nu_s]$, $\text{Cov}[\omega_t,\omega_s]$ for all $t \neq s$, and $\text{Cov}[\nu_t,\omega_s]$ for all t, s are zero. These assumptions are made for convenience rather than necessity, since models with correlated stochastic terms can always be transformed into the independence form.

A. Prior Information

Prior information on the state vector for time $(t + 1)$ is summarized as a normal distribution with mean a_{t+1} and covariance R_{t+1},

$$\theta_{t+1} \mid D_t \sim N[a_{t+1}, R_{t+1}],$$

where D_t denotes the state of knowledge at time t.

From the prior information, forecasts are generated using the observation equation. The forecast distribution for one step ahead has the normal form $Y_{t+1} \mid D_t \sim N[f_{t+1}, Q_{t+1}]$, where $f_{t+1} = F_{t+1}'\, a_{t+1}$ and $Q_{t+1} = F_{t+1}'\, R_{t+1} E_{t+1} + V_{t+1}$.

Forecasting several steps ahead requires the prior information to be projected into the future through repeated application of the system equation. Given the prior for time $(t + 1)$, the implied prior for time $(t + 2)$ from the same standpoint, with no additional information, is $p(\theta_{t+2} \mid D_t)$. This prior is obtained by applying the system equation as in

$$\theta_{t+2} = G_{t+1}\theta_{t+1} + \omega_{t+2}, \quad \omega_{t+2} \sim N[0, W_{t+2}].$$

Linearity once again ensures that this two-step-ahead prior will be normal.

B. Likelihood

The observation equation defines the likelihood for an observed value, $Y_t = y_t$, to have the normal form

$$L(\theta_t \mid Y_t = y_t, V_t) \sim N[F_t'\theta_t, V_t].$$

C. Posterior Information

The prior information is combined with information in the observation (the likelihood) using Bayes' theorem to yield the posterior distribution on the state,

$$p(\theta_t \mid D_{t-1}, y_t) \propto p(Y_t = y_t \mid \theta_t, V_t) p(\theta_t \mid D_{t-1})$$

[the normalizing constant, $p(Y_t = y_t)^{-1}$, typically being ignored].

The state posterior is the product of two normal density functions as just seen, yielding another normal density:

$$\begin{aligned} p(\theta_t \mid D_{t-1}, y_t) &\propto \exp\{-0.5 V_t^{-1}(y_t - F_t'\theta_t)^2\} \\ &\quad \times \exp\{-0.5(\theta_t - a_t)' R_t^{-1}(\theta_t - a_t)\} \\ &\propto \exp\{-0.5(\theta_t - m_t)' C_t^{-1}(\theta_t - m_t)\}, \end{aligned}$$

where the defining moments are obtained as

$$m_t = a_t + A_t e_t, \quad C_t = R_t - A_t A_t' Q_t,$$

$$A_t = R_t F_t / Q_t, \quad e_t = y_t - f_t.$$

The posterior mean is adjusted from the prior value by a multiple of the one-step-ahead forecast error. The amount of adjustment is determined by the adaptive factor A_t, which is, in turn, determined by the relative size of the state prior variance and the observation variance: the larger the observation variance compared with the state prior variance, the smaller the adaptive factor. The intuition here is that if the observation variance is "large," then commensurately large forecast errors are quite compatible with the model quantification. Therefore, the underlying state estimate is adequate and does not require substantial movement. Conversely, if the state prior variance is large compared with the observation variance, then the observation has a lot of relevant information for the state and adjustment from the prior to the posterior should properly reflect that position. This will be seen in the application in Section 3: before a campaign starts, uncertainty about the likely effect is large.

Once an observation is made and posterior descriptions are calculated, concern moves to consideration of the next time. Given the posterior distribution for the state at time $(t-1)$ as normally distributed with mean m_{t-1} and covariance C_{t-1}, direct application of the system evolution equation leads to the prior for time t. Once again a linear combination of normally distributed quantities yields a normal distribution,

$$\theta_t \mid D_{t-1} \sim N[a_t, R_t],$$

where the moments are defined by

$$a_t = G_t m_{t-1}, \quad R_t = G_t C_{t-1} G_t' + W_t.$$

This completes the cycle of prior to forecast to posterior to next prior. These stages characterize the routine on-line updating analysis of the DLM.

D. Forward Intervention

Intervening in an on-line analysis is achieved by adjusting the model-based prior $p(\theta_t \mid D_{t-1})$ to a new prior $p(\theta_t \mid D_{t-1}, I_t)$. Here I_t denotes the extra pertinent information available when making the forecast. Recalling the example, suppose it is judged that the firm's sales will double as a result of the competition's production difficulties. If our routine prior belief on sales level is 200,

$$\theta_t \mid D_{t-1} \sim N[200, R_t],$$

and it is doubled to 400, then the postintervention prior is determined directly as

$$\theta_t \mid D_{t-1}, I_t \sim N[400, R_t].$$

(Typically the prior variance will be increased too, reflecting greater uncertainty.) Whatever the values decided upon, there is a new quantitative description of the prior on which to base future forecasts. It is worth reiterating that the procedure of arriving at the doubling figure may (and most likely will, at least in part) be informal. However, having made the judgment, the information is incorporated into the model in a formal, structured manner.

Formally, subjective intervention of this nature can be modeled by including an additional stochastic term in the system evolution,

$$\theta_t = G_t \theta_{t-1} + \omega_t + \zeta_t, \quad \zeta_t \sim N[h_t, H_t],$$

where the intervention variable ζ_t is uncorrelated with the state θ_{t-1} and system innovation ω_t. This extended form of the system equation retains a linear normal structure so the postintervention prior is normal—only the defining moments are altered as was seen a moment ago. This formalism is mentioned because it provides the mechanism for retrospective assessment (see the next subsection) of interventions.

E. Smoothing

Time series analysis makes use of fitted-model estimates using information from all observations to give a view of component development over time. Such a view has the benefit of hindsight. After a campaign has run its course, looking back one has a clearer picture of what actually happened during the campaign period than one did at any time during the campaign. Later observations contain information about the earlier days, and therefore revised estimates are more accurate and more certain than initial estimates. Such backwards evaluation is variously called smoothing, filtering, or retrospective analysis.

For a dataset comprising observations from time 1 to time n, the retrospective view utilizes the filtered distributions $p(\theta_t \mid D_n)$ for $t = 1, \ldots, n$. Information from observations subsequent to a given time is filtered back to that time in a manner similar to the forward filtering of the on-line analysis. The information filtering mechanism is once again Bayes' theorem, with the system equation providing the necessary cross-temporal joint distributions. Details are given in Pole et al. (1994).

F. Component Forms

The formal analysis detailed in Section 2 thus far is framed in terms of the general DLM. The updating recurrences apply quite generally, regardless of the detailed internal structure of a model. Building a suitable model for a time series is facilitated by knowledge of such model structure. Here we describe models for trends and seasonal cycles, both of which are included in the campaign application.

1. Polynomial Trend Components

The simplest trend model is the first-order polynomial trend, or level. Observed series values are stochastically distributed about a time varying value:

$$Y_t = \mu_t + \nu_t, \quad \mu_t = \mu_{t-1} + \omega_t.$$

The system equation defines the level to be a simple random walk through time. Identifying the terms of a defining DLM, it is easy to see that the regression vector and system evolution matrix are constants,

$$F_t = 1, \quad G_t = 1.$$

A second-order polynomial trend model allows for systematic growth or decline in level, the additional parameter quantifying the time to time change in level (for details, refer to Pole et al. 1994). In the sales application to be discussed below, a first-order polynomial trend is used.

2. Seasonal Component Models

Modeling seasonal patterns in time series requires a component form that is periodic. The most natural representation of such patterns is the *seasonal-factor model*, where a different parameter—a factor—is defined for each point in a cycle. For example, quarterly data exhibiting an annual cycle are modeled with a set of four factors, one for each quarter.

An alternative representation isolates an underlying trend from periodic movement about that trend. Over a complete cycle the effects sum to zero since the trend, which is just the average of the factors over the cycle, contains the overall series movement over that time span. This *seasonal-effects model* defines parameters to measure seasonal departures from a trend. To illustrate, a set of seasonal factors 100, 140, 80, and 120 is equivalent to a constant trend of 110 and seasonal effects −10, 30, −30, and 10. In practice, the effects model is preferred over the factors model because it facilitates separation of underlying trend and seasonal variation about the trend. The trend, of course, may be more complex than a simple constant form. For example, cyclical behavior may be superimposed on a generally increasing level.

For the quarterly, annual cycle, seasonal-effect DLM, the system state comprises four effects parameters, one for each quarter. As time proceeds, the state vector simply rotates by one element each period. If the last quarter was quarter three, then the state is ordered as

$$\theta_{t-1} = \begin{pmatrix} \text{qtr 3} \\ \text{qtr 4} \\ \text{qtr 1} \\ \text{qtr 2} \end{pmatrix}.$$

This quarter (fourth) the state is rotated once to give the new ordering

$$\theta_t = \begin{pmatrix} \text{qtr 4} \\ \text{qtr 1} \\ \text{qtr 2} \\ \text{qtr 3} \end{pmatrix}.$$

The cycling state is expressible in the DLM system equation by setting the evolution matrix to have the cyclic form

$$G_t = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

To complete the model definition, the observation equation picks off the current quarter effect from the state, so the regression vector is simply $F_t = (1\ 0\ 0\ 0)'$.

For a general seasonal component with period p, the definitions of the regression vector and system matrix extend naturally to

$$F_t = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad G_t = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

With a DLM seasonal component model structured in this effect form, there is one complicating factor in the statistical analysis. The zero sum constraint on the effects must be maintained in all of the probabilistic descriptions, priors and posteriors, otherwise the model is difficult to interpret. Without the constraint the underlying trend becomes confounded with the average of the "effects"—which are no longer effects, of course.

If the constraint is satisfied by the prior distribution, then it is guaranteed to be satisfied by the posterior distribution by the rules of the probability calculus. A constraint is just a condition. Unless it is explicitly removed, it is maintained through Bayes' theorem. Therefore, so long as the prior satisfies the constraint, so will the posterior.

It remains only to ensure that the prior description of a seasonal component satisfies the constraint at any time. Priors are either directly specified by the investigator through intervention or are derived from the previous period's posterior from the state evolution. In direct specification, individual seasonal-effect priors are typically quantified independently. Moreover, it is usual for information to be limited to a description of the marginals—moments of individual effects—with no information forthcoming on covariances between pairs of effects. In that circumstance the zero sum constraint must be applied in constructing the

joint prior distribution for the set of effects. The procedure is simple; the BATS program does the calculations automatically.

The prior distribution obtained from the system evolution applied to the previous posterior will satisfy the constraint if the innovation term ω_t meets the constraint conditions. (The system matrix G_t is just a permutation matrix, which serves only to reorder the individual state elements, so it does not affect the relationship between effects.) This term has zero mean, so attention need only focus on the variance W_t. In practice, this variance is specified using the discount approach, so that W_t is a fraction of the previous posterior variance C_{t-1}. Since the posterior distribution satisfies the condition $C_{t-1}1 = 0$, so does the innovation, $W_t 1 = (\delta^{-1} - 1)C_{t-1}1 = 0$.

a. Harmonic analysis. Mathematical functions such as sines and cosines exhibit well-known cyclical behavior. The cosine function $\cos(\omega(t-1))$ defines a periodic function with period $2\pi/\omega$. When $\omega = \pi/6$, the cosine function has period 12. If a monthly time series with an annual seasonal pattern is well approximated by the cosine form, then a very parsimonious model can be used in modeling. Instead of 11 effect parameters (one for each month plus the zero sum constraint) only one is required,

$$Y_t = a_t \cos\left(\frac{\pi(t-1)}{6}\right) + \nu_t.$$

The parameter a_t models an arbitrary amplitude (seasonal peak to trough variation). A simple generalization of the basic cosine wave allows the cycle maximum and minimum to be phase shifted. The seasonal peak can be translated from the default January position to any other month by using a sum of sine and cosine terms with the same frequency,

$$Y_t = a_t \cos\left(\frac{\pi(t-1)}{6}\right) + b_t \sin\left(\frac{\pi(t-1)}{6}\right) + \nu_t.$$

This model has two parameters, combinations of which may be defined to give a cosine wave of period 12 with any desired amplitude and phase shift. A DLM formulation of this single harmonic function is obtained using a two-parameter state, with the regression vector and system matrix defined as follows (both are time invariant so the time subscript is omitted):

$$F = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad G = \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix}.$$

The system matrix is p-cyclic: like the permutation matrix used in the seasonal-effects model, it repeats itself after p periods, $G^p = G$.

More complicated cyclical patterns are obtained by considering additional cosine waves with higher frequencies. The function $\cos(2\omega(t-1))$ is similar to the function just examined with the exception that it completes two full cycles in the time it takes the first to complete one. Its period is half that of the first or, equivalently, its frequency is twice as great. With $\omega = \pi/6$ this higher-frequency function completes two full cycles in 12 time periods. Once again, a linear combination of sine and cosine terms at this frequency allows for an arbitrary peak-to-trough range and phase shift.

Adding two such periodic functions together enables a much wider range of cyclical behavior to be modeled than is possible with either function individually—and still with only four parameters,

$$Y_t = [a_1 t \cos\omega(t-1) + b_1 t \sin\omega(t-1)]$$

$$+ [a_2 t \cos 2\omega(t-1) + b_2 t \sin 2\omega(t-1)] + \nu_t.$$

The regression vector and system matrix for this two harmonic model are

$$F = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad G = \begin{pmatrix} \cos\omega & \sin\omega & 0 & 0 \\ -\sin\omega & \cos\omega & 0 & 0 \\ 0 & 0 & \cos 2\omega & \sin 2\omega \\ 0 & 0 & -\sin 2\omega & \cos 2\omega \end{pmatrix}.$$

Continuing to add higher-frequency terms, more and more complex seasonal patterns can be obtained through the harmonic formulation. With a full set of harmonics any arbitrary pattern can be represented.

b. Why harmonic analysis? When appropriate, a restricted harmonic model for seasonal variation can be very parsimonious. However, relating the coefficients of sine and cosine terms to the observation series is not easy. In fact, knowledge of the full harmonic machinery is necessary and even then interpretation of the coefficients is fraught with difficulty. Imagine trying to explain to a company sales manager that his

product exhibited yearly seasonal variation described by a cosine wave with amplitude such and such and phase shift so and so.

It is emphasized that practical application of harmonic models is not carried out in terms of the Fourier coefficients. As technical analysts we may utilize a harmonic model for our convenience, but communication and interpretation is always made in terms of meaningful quantities, namely seasonal effects. Technical details of the relationship between Fourier coefficients and seasonal effects are straightforward (Pole et al. 1994).

G. Superposition: Block Structured Models

Component forms for trends, seasonal patterns, and regressions are the building blocks for constructing models of complex time series behavior. A seasonal-effects DLM will typically require a trend component—only if the series has a fixed zero underlying level will this not be so. The linear, additive structure of the DLM enables component models to be brought together in a straightforward manner. A level trend plus single harmonic seasonal model is a DLM with regression vector and system matrix defined by

$$\theta_t = \begin{pmatrix} \theta_T \\ \theta_S \end{pmatrix}_t, \quad F = \begin{pmatrix} F_T \\ F_S \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},$$

$$G = \begin{pmatrix} G_T & 0 \\ 0 & G_S \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{pmatrix}.$$

Note that the regression vector simply stacks the individual regression vectors on top of each other; similarly the system state stacks up the component states. The system matrix is structured block diagonally. (Naturally, for all three quantities the stacking order must be the same, although the particular order chosen does not matter.)

In this way, models with any number of components may be constructed. A more complex seasonal pattern—with multiple harmonics—just adds in more of the harmonic components (with different frequencies); regressions on exogenous predictors just tag on one at a time.

H. Extensions

The DLM class of models is extended by Migon and Harrison (1985) and West, Harrison and Migon (1985) to include approaches for nonnormal and nonlinear models. Pole (1988) and Pole and West (1990) introduce quadrature techniques for classes of nonnormal and nonlinear dynamic models. Jacquier, Polson, and Rossi (1994) and Müller and Pole (1993) analyze nonlinear dynamic models with sophisticated sampling techniques. Pole (1994) surveys numerical approaches to nonlinear and nonnormal dynamic modeling.

III. CAMPAIGNS

In constructing mathematical/statistical models for forecasting and analysis it is vital to maintain links with the application context. Model elements should be recognizable to the investigator in respect of the system being investigated, providing understandability and facilitating interpretation of parametric estimates and forecast components. As illustrated in the discussion in the previous section, the DLM and Bayesian framework provide precisely these features for general models by combining, in a straightforward way, submodels individually designed for specific identifiable features of a series. In the remainder of the chapter we describe and illustrate an extension of these model structuring ideas for analyzing short-lived promotional campaigns.

Interest centers on two key aspects: adjusting routine sales forecasts in the light of advance knowledge of a campaign and, crucially, learning as much as possible about the actual effects of the campaign on sales once it has run its course. Predictions are adjusted for campaigns so that potential sales are not lost because of depleted inventories, and also to minimize production disruption and overstocking. The learning process facilitates improved predictions of the likely outcomes of future similar campaigns, as well as providing the means for comparative analysis of alternative forms of promotional activity and identifying factors that are important in determining different market reactions to promotions.

A. Nature of Promotional Campaigns

The basic characteristic of a promotional campaign is that it is a departure from the norm, introducing additional factors in the market, altering perceptions and plans, and that (for the most part) it is not sustained for a great length of time. A natural way of modeling effects

arising from such a campaign is to introduce an additional explanatory component into the model for precisely the period affected by the campaign (this begs the question of determining the "period of effect," about which we will say more later). In this way, effects on the series of interest (sales in our illustration) caused by the promotion can be isolated, forecast, and subsequently estimated given the observed data. This isolation is of crucial practical importance. It means that the instabilities, changes, and uncertainties associated with a campaign can be prevented from affecting the non-promotion-related model components, such as the underlying level, so that postcampaign forecast performance does not deteriorate. Moreover, the separation of market outcomes into "what we would have expected to happen without the campaign" and "explicit campaign-related change" provides the information needed to assess the effectiveness of promotional activity and thereby improve forecast performance during future promotions.

Sustained changes attributed to a promotional campaign, or arising for some other reason confounded with a promotion, can also be handled in the general framework of subjective intervention and campaigns. However, we shall ignore this for the present since it would otherwise cloud the discussion.

B. Modeling Campaigns in the DLM

To make matters concrete, we consider sales of a soft drink product. Analysis of the recent history and market conditions for this product has resulted in choice of a model comprising three basic components: a steady trend (single-parameter random walk); a regression on the price of the product; and a seasonal component. Identifying θ_{it}, $i = 1, 2, 3$, with each of the latter components respectively, we may write (assuming commensurate dimensionality of the corresponding terms),

$$\theta_t = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}_t, \quad G = \begin{pmatrix} G_1 & 0 & 0 \\ 0 & G_2 & 0 \\ 0 & 0 & G_3 \end{pmatrix}, \quad F_t = \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix}_t,$$

where $G_1 = 1$, $G_2 = 1$, G_3 is the block diagonal matrix of sine and cosine terms for the harmonic representation of a full harmonic model, $F_1 = 1$, F_{2t} = price at time t, and F_3 is the regression vector for a harmonic component.

Table 1. Campaign Prior Estimates

Month	May	June	July	Aug	Sept	Oct
Mean effect	35	40	40	40	20	5
s.d.	25	20	20	20	20	25

Now, the company is to introduce an advertising campaign in May 1993 that is to run until October. For this product, previous experience with promotions suggests that sales will only be affected during the months of the campaign; there is no residual carryover to succeeding months. Initially we assume that the same behavior will be characteristic of the present promotion. It remains to quantify the expected effects of the promotion on the sales for May to October. On introducing the campaign, the marketing department subjectively assesses the effect on sales and their considerable uncertainty about the assessment as summarized in Table 1.

For the duration of the promotion we need to extend the basic DLM to include the campaign effect. This is done in a natural way by specifying a campaign component that reflects the required structure and augmenting the existing DLM with this component for the months of May to October. The campaign component structure used here has a particularly simple form, namely, temporally independent effects, which may be written as

$$\theta_{4t} = \delta_t,$$

with $F_4 = 1$ and $G_4 = 0$. The augmented model becomes

$$\theta_t = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix}_t, \quad G = \begin{pmatrix} G_1 & 0 & 0 & 0 \\ 0 & G_2 & 0 & 0 \\ 0 & 0 & G_3 & 0 \\ 0 & 0 & 0 & G_4 \end{pmatrix}, \quad F_t = \begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \end{pmatrix}_t$$

The individual campaign effects for months May through October have prior information given by the subjective assessment in Table 1. Although we have assumed it here, independence of the monthly effects is not a requirement of the model. In this application, the nature of the product and analysis of comparable previous promotions convinced the marketing department that independent effects on sales during the campaign period was a reasonable a priori choice of model structure. Other markets have different response characteristics and require alternative structural forms for a campaign, modified random walks (for memory effects) or regressions on campaign expenditures for instance. The proposed modeling procedure allows for the full range of basic DLM components to be included as campaign elements.

Augmenting the model structure at the onset of a campaign requires that the state vector prior means and covariances be suitably extended to enable forecasting from the model and subsequent data analysis. The extension procedure is the "obvious" one. The mean vector extends to include the mean effect of the campaign as given in Table 1; the covariance between the campaign and other components is set to zero. Crucially, this covariance prior setting is maintained for the entire campaign period. If D_t represents all information available to time t, then the prior covariance matrix $R_t = \mathrm{Var}(\theta_t \mid D_{t-1})$ for the state parameter θ_t has the block diagonal form

$$R_t = \begin{pmatrix} R_a & 0 \\ 0 & R_c \end{pmatrix}_t,$$

where R_{at} is the covariance of the precampaign state vector as computed in the usual manner from the posterior value at time $(t-1)$ using the component discount strategy (see Pole et al. 1994); R_{ct} is simply the square of the relevant standard deviation in Table 1.

Setting the prior covariance between the campaign component and the other model components to zero throughout the promotional period follows the general component intervention strategy. The importance of the procedure should not be understated. The posterior correlation between campaign and other components during the campaign time will, in general, be nonzero because of the usual correlation inducing effect of observing all components combined in the single sales value. The prior specification protects the non–campaign components from propagation of any such spurious correlations.

C. Example Analysis

Figure 4 shows one year's forecasts of sales of the CSDRINKS product. The forecasts are made in April 1993 for each of the next 12 months based upon analysis of the "routine" model thought to be appropriate for the data up to that time. Clearly the forecasts are inadequate in

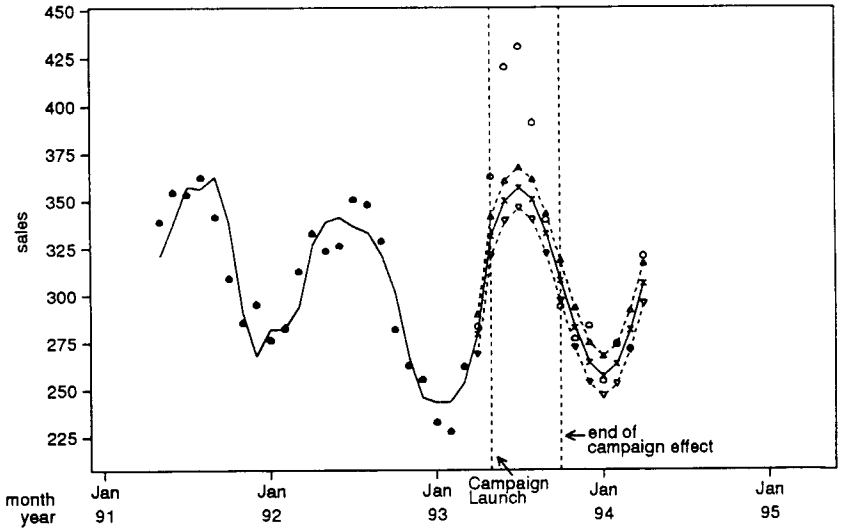

Figure 4. Forecasts for 12 months from April 1986: mean, 2/3 probability limits, and observed values.

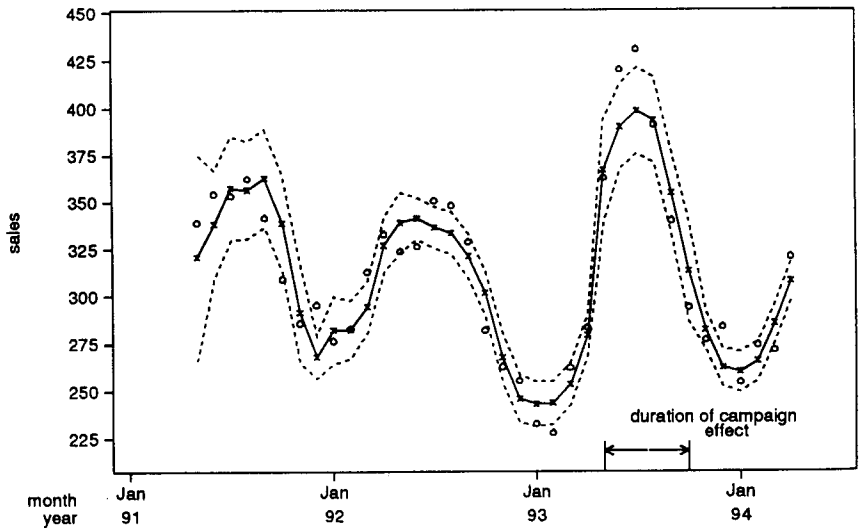

Figure 5. One-step-ahead forecasts (with 2/3 probability interval) for campaign model.

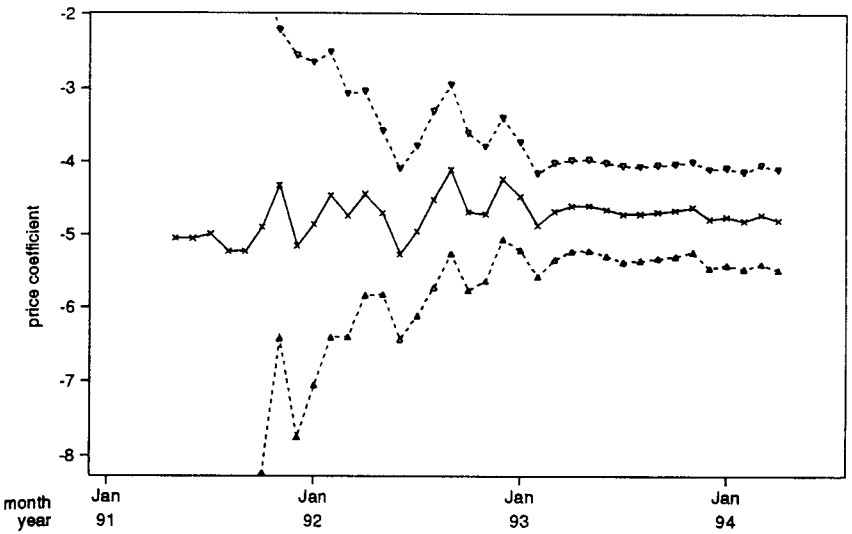

Figure 6. On-line estimate of regression coefficient (with 2/3 probability interval) on price.

the short term because they do not take into account the promotional campaign launched in May and continuing until October. Considerable additional uncertainty is introduced by the promotion, as described by the prior views expressed in Table 1. Figure 5 shows the one-step-ahead forecasts (with 2/3 probability limits) from the augmented model including the campaign component and the prior information. There is much greater forecast uncertainty associated with the campaign period than with the immediately preceding and succeeding months (compare Figure 4). Figure 6 shows the updated on-line (*not* fitted) estimates of the price elasticity of the product, demonstrating how, during the campaign period, these estimates are protected from the large forecast errors that are properly attributed to the campaign. A similar picture can be drawn for the on-line seasonal estimates.

On completion of the campaign we wish to estimate what actually happened to sales. Figure 7 and Table 2 show the retrospective estimates of the campaign effect showing it to have had much greater impact than initially anticipated (compare Table 1). Note also the offsetting negative impact in the final month of the campaign. That feature was *not* present in the original prior information that the marketing department supplied based on their previous experience and is therefore quite important for

Table 2. Retrospective Estimate of Campaign Effect

Month	May	June	July	Aug	Sept	Oct
Mean effect	28	63	66	40	8	-12
s.d.	9	9	9	9	9	9

future reference. Many tenable explanations may be posited for the effect, including:

1. The nature of the market (including competitors' response) was different in some important aspect on this occasion as compared with previous campaigns analyzed.
2. Marketing was "optimistic" interpreting end effects of past campaigns.
3. The campaign differed from previous campaigns in some important aspect.

The possibilities must be considered by the marketing department as they compare the retrospective estimates of the current campaign with those on which the prior estimates for this analysis were based. Prior estimates for future campaigns will be suitably modified according to the information gleaned from the examination.

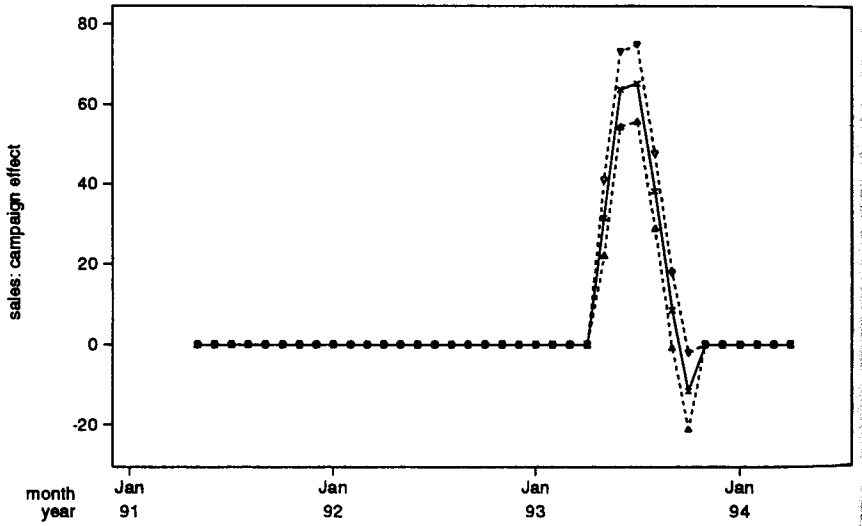

Figure 7. Retrospective estimate of campaign effect: mean and 2/3 probability interval.

The message from the analysis in this section is clear. The model component structure permits a decomposition of sales into readily identifiable, practically meaningful component parts. It enables a clear and unambiguous estimation of individual effects, and presents an easily interpreted picture of performance. Such information breakdown is important not only in assessing the effects of a single campaign as done here, but also in comparative analysis of alternative forms of promotion.

IV. FURTHER DISCUSSION

In the preceding application we specified at the outset both the duration of the campaign's effects on sales and the initially expected magnitude of the effect for each month, together with uncertainty about the estimates. In a live forecasting situation we have rather more flexibility than this implies. We can modify the campaign prior estimates continually as we observe new data each month: after observing the actual sales for May we might wish to update our prior estimates for the effect on sales for June and so on. This would be especially crucial if a particular type of campaign had previously been associated with more than one pattern of response, but where the critical "environmental" distinguishing features giving rise to the differences have not been sufficiently well identified. Choosing which pattern of behavior is most likely to occur would then be particularly dependent on analysis of consumer response and competitors' reactions in the early part of the campaign.

A decision on the campaign's end-of-effect time does not have to be made until it has arrived and will be intimately bound up with decisions on how to proceed with the campaign. It is not unheard of for a promotion to be stopped in its tracks by consideration of unforeseen external events, including adverse reaction to the campaign itself. A campaign effect period specified too far into the future will, if maintained, result in posterior estimates of zero for the effect at the later times but forecast variance there will be inflated by the amount of prior uncertainty on the campaign component. On the other hand, curtailing the presumed effect of the campaign too early will lead to the resulting (large) forecast errors feeding through into the non-campaign components. The former situation is clearly preferable although the latter case would typically be flagged by the forecast system's monitoring facilities, allowing remedial action to be taken at the time.

None of these operational considerations alters the basic analysis that has been presented.

Finally, note that the proposed modeling extension of the DLM is readily generalizable in formal terms to multivariate models, useful where several related products are being forecast simultaneously. The operational reality is complicated, though, because there are more difficult problems of cross-correlation structure to consider when formulating prior information.

V. SUMMARY

Alternative approaches to analyzing sales for consumables subject to promotional activity involve more sophisticated modeling of consumer response to advertising resulting in complex, nonlinear models (see, for example, Migon and Harrison 1985; West and Harrison 1986, 1989). Analysis of such models requires considerably more work and greater mathematical ability than the analysis we have described in this chapter. That is not necessarily a disadvantage—indeed, more analysis may often confer additional benefits in terms of greater explanatory power of the models and reduced forecast variance. However, such analysis is not always feasible. Resource constraints—especially time—may be prohibitive. In such situations the approach we have presented provides a simple, practical alternative.

A more pernicious problem with nonlinear techniques is the potential for disaster, even in the hands of experienced practitioners. For the novice, such models are a minefield best approached with caution and preferably with the benefit of insight from first applying the simpler, linear models we have described.

Using basic ideas of model component structure, temporary model augmentation, and formal subjective intervention, we have demonstrated how the normal DLM framework can successfully capture the effects of campaign-induced structural change. The soft drink application illustrated the practicability of the approach, demonstrating the range of information available from the analysis.

The software package BATS (Pole et al. 1994) is currently being extended to include analysis of campaign effects.

ACKNOWLEDGMENTS

The author thanks Jeff Harrison for the idea, Mike West for helpful discussion on the first draft of Section 3, and Tom Fomby for many improvements to the general presentation.

REFERENCES

Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*, 2nd edition. San Francisco: Holden-Day.

Box, G. E. P. and G. C. Tiao (1975). "Intervention Analysis with Applications to Economic and Environmental Problems." *Journal of the American Statistical Association* 70, 70–79.

Harrison, P. J. (1965). "Short Term Sales Forecasting." *Applied Statistics* 15, 102–139.

Harrison, P. J. (1988). "Bayesian Forecasting in O.R." In *Developments in Operational Research 1988*, edited by N. B. Cook and A. M. Johnson. Oxford: Pergamon.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business and Economic Statistics* 12, 371–388.

Migon, H. S. and P. J. Harrison (1985). "An Application of Non-Linear Bayesian Forecasting to Television Advertising." In *Bayesian Statistics II*, edited by J. M. Bernardo et al. Valencia: Valencia University Press.

Müller, P. and A. Pole (1993). "Monte Carlo Posterior Integration in Garch Models." ISDS Discussion Paper 93-A28, Duke University.

Pole, A. (1988). "Transfer Response Models: A Numerical Approach." In *Bayesian Statistics 3*, edited by J. M. Bernardo et al. Oxford: Oxford University Press.

Pole, A. (1994). "Numerical Techniques for Bayesian Inference in Nonlinear Sequential Models." Warwick Research Report, Department of Statistics, University of Warwick, U.K.

Pole, A. and M. West (1990). "Efficient Bayesian Learning in Nonlinear Dynamic Models." *Journal of Forecasting* 9, 119–136.

Pole, A., M. West, and P. J. Harrison (1994). *Applied Bayesian Forecasting and Time Series Analysis*. New York: Chapman & Hall.

West, M. and P. J. Harrison (1986). "Advertising Awareness Response Model: Micro-Computer APLPLUS/PC Implementation." Warwick Research Report, 118. Department of Statistics, University of Warwick, U.K.

West, M. and P. J. Harrison (1989). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.

West, M., P. J. Harrison, and H. S. Migon (1985). "Dynamic Generalized Linear Models and Bayesian Forecasting (with Discussion)." *Journal of the American Statistical Association* 80, 73–97.