Journal of Business & Economic Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ubes20

Bayesian Forecasting With Stable Seasonal Patterns

Robert M. Oliver [a]

[a] Department of Industrial Engineering and Operations Research , University of California , Berkeley , CA , 94720

Published online: 02 Jul 2012.

Journal of Business & Economic Statistics, January 1987, Vol. 5, No. 1

Bayesian Forecasting With Stable Seasonal Patterns

Robert M. Oliver
Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720

A multiplicative seasonal forecasting model for cumulative events in which, conditional on end-of-season totals being given and seasonal shape being known, it is shown that events occurring within the season are multinomially distributed is presented. The model uses the information contained in the arrival of new events to obtain a posterior distribution for end-of-season totals. Bayesian forecasts are obtained recursively in two stages: first, by predicting the expected number and variance of event counts in future intervals within the remaining season, and then by predicting revised means and variances for end-of-season totals based on the most recent forecast error.

KEY WORDS: Seasonal forecasting; Cohort data; Linear Bayes; Multiplicative seasonal; Kalman.

1. INTRODUCTION

In many seasonal forecasting problems the fraction of events that occur by a certain time within the season are stable over time; that is, the pattern and shape of event occurrence is similar from season to season although the end-of-season totals and the cumulative count of events within any given season are uncertain. Stability in seasonal patterns is documented in the time series and forecasting literature (see, e.g., Chang and Fyffe 1971; Green and Harrison 1973), as are many types of longitudinal cohort data (see Marshall and Oliver 1970; Grinold and Marshall 1977). There have been numerous references in the scientific literature to seasonal models in which the demand in a specified period of time (such as a month), is viewed as a fraction of total demand in the entire season (such as a year)—that is, a multiplicative model. In this article a multiplicative model is assumed when arrivals within the season are assumed to be generated by a binomial process conditional on known end-of-season totals; the analysis then uses the information associated with the occurrence of events to yield inferences on the end-of-season totals as well as the counts of future events within the remainder of the season.

One of the early models making use of the stability of longitudinal data was developed to predict student enrollment and attendance at a large university (Marshall and Oliver 1970, 1979). Although their models did not use a Bayesian analysis to provide updates throughout the season, examination of longitudinal cohort data for student attendance patterns indicated surprising stability from year to year. One of the earliest Bayesian seasonal models that made explicit use of stability in the seasonal factors was proposed by Chang and Fyffe (1971) to forecast style goods. Their model assumed that monthly sales equaled a known fraction of unknown, random end-of-season total sales but that the variance of the underlying noise process was stationary and homoscedastic, even though one of their two proposed methods for estimating the seasonal fractions implicitly used a binomial noise variance. The purpose of the Bayesian analysis was to provide an early and accurate forecast of total demand for a seasonal style good, the total life of which was only a few months. A similar analysis using a dynamic linear model (DLM) was developed by Green and Harrison (1973) to predict demands and returns of fashion goods purchased from a mail-order firm. They assumed that the logarithm of customer demand was the sum of the logarithms of known seasonal factors and randomly varying level and trend terms. Although return rates and seasonal factors were estimated by least squares from historical data, Bayesian updates of level and trend were based on the assumption that stationary noise variances were proportional to a power of the demand level.

Another reference to a seasonal fractions model in a Bayesian setting is an article by Fildes and Stevens (1978) in which cumulative monthly sales are expressed as a fraction of annual sales; the model is used to illustrate how a Bayesian analysis can proceed from time zero with little or no data. The analysis and results, however, contain the incorrect assumption that cumulative seasonal errors are *independent* and variances are *stationary* from period to period within the season, both of which are serious in the sense that the updating equations are extremely sensitive to these critical assumptions.

An article by Olson (1982) used a similar model for the number of adopters of a new technology. He implicitly assumes that adopters act independently of one another and that the variance of the resulting noise term is binomial—that is, proportional to the size of the population and the product of $F(t)(1 - F(t))$, with $F(t)$

the probability of adoption by an individual firm on or before time t. His proposal to predict the number of future adopters by a Gaussian approximation is a global calculation (once parameters of the model have been fitted) in the sense that the forecasts are not revised as new data arrive in each period.

2. A DETERMINISTIC MODEL

Consider, in a deterministic setting, how one captures the effect of a stable seasonal pattern upon the occurrence of future arrivals given that $n(s) \leq N$ are known to have occurred at time $s \leq t \leq T$, with T the end of the season. It is assumed that the fraction of the total season demand, N, arriving on or before time t is $F(t)$, known and given. Cumulative arrivals at a future t can therefore be expressed as

$$X(t) = n(s) + (N - n(s))F(t \mid s), \quad s \leq t \leq T$$

$$F(t \mid s) = (F(t) - F(s))(1 - F(s))^{-1}, \tag{1}$$

with $F(t \mid s)$ the renormalized fraction that applies to the subinterval interval (s, T). Unfortunately, the deterministic model does not suggest how one should capture the uncertain arrival of future events or how to estimate the total season demand N in the event that actual data for cumulative counts do not agree with the predictions of $X(t)$ in (1).

3. BINOMIAL EVENTS CONDITIONED ON KNOWN END-OF-SEASON TOTALS

If it is assumed that the cumulative demand is $X(T) = N$ at the end of the season and that events occur independently of one another, and if we have

Assumption 1. An event has a probability $F(t)$ of occurring on or before time t,

Assumption 2. Events occur independently of one another,

then the conditional probability of observing n events on or before t is binomial (bin), that is, $X(t) \sim \text{bin}(N, F(t))$, so

$$p(n \mid N) = \Pr\{X(t) = n \mid X(T) = N\}$$

$$= \binom{N}{n} F(t)^n(1 - F(t))^{N-n}, \qquad 0 \leq n \leq N, \tag{2}$$

with conditional expectation and variance given by $NF(t)$ and $NF(t)(1 - F(t))$, respectively. Similarly, if we condition on the number n that has been observed by time $s \leq t$ as well as the total number at the end of the season, then

$$(X(t) - n \mid X(s) = n, N)$$

$$\sim \text{bin}(N - n, F(t \mid s)), \qquad s \leq t \leq T, \tag{3}$$

with conditional expectation and variance $(N - n)F(t \mid s)$ and $NF(t \mid s)(1 - F(t \mid s))$. Note that the variance at $t = s$ and $t = T$ is zero and that the greatest uncertainty (as measured by variance) occurs when half of the remaining fraction has yet to arrive.

4. THE MULTINOMIAL DISTRIBUTION OF EVENTS IN (0, T)

Let the total number of events at the end of the season be given by N, and let the cumulative counts be $0 \leq n(t_1) \leq n(t_2) \leq \cdots \leq n(t_k) \leq N$ at time points $0 = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_k = t \leq T$. In addition, denote the probability that an event will occur in (t_{j-1}, t_j) by $a_j = F(t_j) - F(t_{j-1})$, $a_1 + a_2 + \cdots + a_k = F(t)$, $a_{k+1} = 1 - F(t)$, $F(t_0) = F(0) = 0$ and $F(T) = 1$. Then the conditional probability $p(i_1, i_2, \ldots, i_k, i_{k+1} \mid N, a_j)$ that i_j occur in (t_{j-1}, t_j) conditional on N is multinomial with parameters N and $a_1, a_2, \ldots, a_k, a_{k+1}$. The expression can be factored into two parts so that

$$p(i_1, i_2, \ldots, i_{k+1} \mid N, a_j) = p(i_1, i_2, \ldots, i_k \mid n, b_j)$$

$$\times \left\{\binom{N}{n} F(t)^n(1 - F(t))^{N-n}\right\}, \tag{4}$$

with $b_j = a_jF(t)^{-1}$; the first factor on the right side of Equation (4) is the multinomial probability for the events that occur in the intervals $(0, t_1)(t_1, t_2) \cdots (t_{k-1}, t_k = t)$ and is independent of N, the total number at the end of the season. The second factor is the binomial probability that $i_1 + i_2 + \cdots + i_k = n$ events occur in $(0, t)$ and the remainder $N - n$ in (t, T). Since N only appears in the second factor, n is seen to be a sufficient statistic for N and a maximum likelihood estimate is easily obtained by finding N^* such that the inequality

$$p(n \mid N^*) \geq \max[p(n \mid N^* - 1); p(n \mid N^* + 1)] \tag{5}$$

is satisfied. Cancellation of like terms in the inequality yields $n/F(t) - 1 \leq N^* \leq n/F(t)$. When $n/F(t)$ is integral there may be two adjacent values of N^* that yield the maximum; when $n/F(t)$ is not an integer a unique N^* equals the integral part of $n/F(t)$. In most cases of interest with large numbers for end-of-season totals the integrality of N is probably not an important issue, in which case N^* has the physical appeal that it is the seasonal total that one would estimate by simply "scaling up" the most recent count of events at time t. In properly treating N as a random variable for the end-of-season total, it will be seen that N^* plays an important role in several cases.

5. CONJUGATE PRIORS: BINOMIAL, POISSON, AND NEGATIVE BINOMIAL

Assume that $p_{N|0}$ is the prior probability of an end-of-season total equal to N at time 0 and that $p_{N|n}$ is the posterior probability given $X(t) = n$. Then it follows that the posterior is proportional to the product of the binomial likelihood in (4) with the prior $p_{N|0}$ because $p(i_1, i_2, \ldots, i_k \mid n, b_j)$ is a common factor to both:

$$p_{N|n} \propto p(i_1, i_2, \ldots, i_k, i_{k+1} \mid N, a_j)p_{N|0}$$

$$\propto \binom{N}{n} F(t)^n(1 - F(t))^{N-n}p_{N|0}. \tag{6}$$

The binomial, Poisson, and negative binomial distributions are conjugate priors for $p_{N|n}$, the posterior probability of end-of-season demand based on the most recent observation of $X(t)$. In the case of the Poisson where $X(T) \mid 0 \sim P(\lambda)$ with λ the expected seasonal total, it follows that the conditional distribution of counts in any future interval $(0 \leq s, t \leq T)$ is also Poisson, that is,

$$(X(t) - X(s) \mid X(s) = n) \sim \mathbf{P}(\lambda(F(t) - F(s))), \quad s \leq t \leq T, \quad (7a)$$

so the posterior distribution of the end-of-season total is a shifted Poisson,

$$\begin{aligned} p_{N|n} &= \Pr\{X(T) = N \mid n\} \\ &= \exp[-\lambda(1 - F(s))][\lambda(1 - F(s))]^{N-n}/(N - n)!, \\ &\qquad N = n, n + 1, \ldots, \quad (7b) \end{aligned}$$

with posterior mean and variance given by $n + \lambda(1 - F(s))$ and $\lambda(1 - F(s))$. In the Poisson case, the cumulative count at time s only adds to the total count at the end of the season; the mean λ is unaffected. Equation (7) also shows that graphically the seasonal total can be estimated at time s by placing a new axis origin at the point $(n(s), s)$ and drawing from that point a cumulative curve that has the same shape as the original in the interval (s, T).

The negative binomial prior is particularly interesting because it is a two-parameter distribution with variance to mean ratio greater than or equal to 1. Given a negative binomial (**NB**) prior $N \mid 0 \sim \mathbf{NB}(M, p)$, then both the marginal distribution of $X(t)$ and the posterior distribution of $X(t) - X(s)$ are negative binomial,

$$\begin{aligned} &(X(t) - X(s) \mid X(s) = n) \\ &\qquad \sim \mathbf{NB}(M + n, 1 - (1 - p)(1 - F(s))). \quad (8) \end{aligned}$$

Although I have not yet seen any practical applications of the binomial prior, the Poisson case (variance to mean ratio of 1) arises whenever end-of-season totals can be viewed as rare events. An example of this behavior occurs in the prediction of donors and donations to a charitable fund (Britto and Oliver 1986) where the number of donors is a small fraction of potential donors. Use of the negative binomial is often appropriate when the variance of the prior distribution is large owing to lack of information about location of the mean.

6. JOINT PROBABILITY AND COVARIANCE OF ARRIVALS IN TWO ADJACENT INTERVALS

Consider the two subintervals (r, s) and (s, t), where $X(r) = m$ and $X(s) = n$; that is, the cumulative counts at the beginning of each interval equal m and n, respectively, with $m \leq n$ and $0 \leq r \leq s \leq t \leq T$. The conditional probability that there are j arrivals in (r, s) given $X(r) = m$ and the end-of-season total N, is binomial with parameters $N - m$ and $F(s \mid r)$. Moreover, the conditional probability that there will be k arrivals in the interval (s, t) given $X(s) = m + j$ and an end-of-season total equal to N is the binomial distribution with parameters $N - m - j$ and $F(t \mid s)$. Thus the joint probability of j in (r, s) and k in (s, t) conditional on m at r is given by

$$\begin{aligned} &\binom{N - m}{j}\binom{N - m - j}{k} (F(s) - F(r))^j(F(t) - F(s))^k \\ &\qquad \times (1 - F(t))^{N-m-j-k}(1 - F(r))^{m-N}, \quad (9a) \end{aligned}$$

from which it follows that arrivals in two adjoining (nonoverlapping) intervals are dependent and negatively correlated; a larger than average number of arrivals in the first interval will lead with high probability to a smaller than average number of arrivals in the following one, since the number that have not arrived is made smaller at the beginning of the second interval. From (9a) it can be shown that

$$\begin{aligned} &\operatorname{cov}[X(s) - X(r), X(t) - X(s) \mid m, N] \\ &\quad = -(N - m)(F(s) - F(r))(F(t) - F(s)) \\ &\qquad \div (1 - F(r))^2. \quad (9b) \end{aligned}$$

One-step-ahead forecast errors, however, are uncorrelated, as can be demonstrated by calculating the covariance of $e(r, s)$ with $e(s, t)$, where $e(r, s) = X(s) - m - (N - m)F(s \mid r)$ and $e(s, t) = X(t) - X(s) - (N - X(s))F(t \mid s)$. Since each error has expectation zero,

$$\begin{aligned} &\operatorname{cov}[e(r, s), e(s, t) \mid m, N] \\ &\qquad = E[e(r, s)e(s, t) \mid m, N], \quad (9c) \end{aligned}$$

and since the conditional and marginal mean of $e(s, t) \mid n, m, N = 0$, it immediately follows that

$$\begin{aligned} &E[e(r, s)e(s, t) \mid m, N] \\ &\quad = E_{X(s)}[e(r, s)E[e(s, t) \mid m, n, N] \mid m, n, N] \\ &\quad = E_{X(s)}[e(r, s) \cdot 0 \mid m, n, N] = 0. \quad (9d) \end{aligned}$$

7. LARGE *N*, GAUSSIAN PRIORS, AND LINEAR BAYES

With large N, the probability in Equation (2) can be approximated by a Gaussian distribution with mean $NF(t)$ and variance $NF(t)(1 - F(t))$; that is,

$$(X(t) \mid X(T) = N) \sim N(NF(t), NF(t)(1 - F(t))). \quad (10)$$

Hence, conditionally on N being given, the increase in the cumulative count in the interval (s, t) can be written as the sum of expected new arrivals plus an error term,

$$(X(t) - X(s) \mid n, N) = (N - n)F(t \mid s) + a(s, t), \quad (11)$$

with $a(s, t)$ being a nonstationary Gaussian error having expectation zero and conditional variance $(N - n) F(t \mid s)(1 - F(t \mid s))$ that depends not only on the number of arrivals, $N - n$, in the remainder of the season but also on the probability that an individual will arrive in (s, t) given no arrival by s. It might appear that the joint distribution of the end-of-season total and the error term is bivariate normal, but it should be pointed out that the joint density of $X(t) = x$ and $X(T) = y$ is approximately

$$p(x, y \mid X(s) = n) = (2\pi)^{-1}[p_s F(t \mid s)(1 - F(t \mid s)]^{-1/2} \times \exp -\{Q(x, y \mid F(t \mid s))\},$$

where

$$2Q(x, y \mid F) = [(x - yF)^2/yF(1 - F)] + (y - m_s)^2/p_s + \ln y, \quad (12)$$

with m_s and p_s the posterior expectation and variance of end-of-season demand conditional on knowing $X(s)$ at time s. We make the following assumption for the initial prior at time 0.

Assumption 3. $(N \mid 0) \sim N(m_0, p_0)$. At time s, the conditional variance of $a(s, t)$ is given by the following assumption.

Assumption 4. $\text{var}[a(s, t) \mid n(s), N] = (m_s - n(s)) F(t \mid s)(1 - F(t \mid s))$. In Assumption 4 it is effectively assumed that the variance of the noise is constant between two successive updates, using the most recent value for the posterior mean to estimate the next binomial variance term. Thus, it follows that the marginal variance of $a(s, t)$ under this linear Bayes assumption is obtained by adding $p_s F(t \mid s)^2$ to the conditional variance:

$$\text{var}[a(s, t)] = (m_s - n(s))F(t \mid s) \times (1 - F(t \mid s)) + p_s F(t \mid s)^2.$$

With this simplifying assumption in the multiplicative model, one can make use of well-known Kalman filtering techniques to obtain the posterior Gaussian distributions for $N - n$ and hence N. At time 0, under Assumptions 3 and 4,

$$(X(s) \mid N) \sim N(NF(s), m_0 F(s)(1 - F(s))). \quad (13)$$

Using Bayes's rule and the fact that the covariance matrix of the joint probability of $X(s)$ and $X(T)$ is now given by

$$\Sigma_{0,s} = \begin{bmatrix} m_0 F(s)(1 - F(s)) + p_0 F(s)^2 & p_0 F(s) \\ p_0 F(s) & p_0 \end{bmatrix}, \quad (14)$$

the posterior distribution of $X(T)$ can be shown to be $N(m_s, p_s)$, where the updating equations for posterior mean and variance are given by

$$m_s = m_0 + K(0, s)(n(s) - m_0 F(s))$$
$$p_s = p_0 - p_0^2 F(s) \times [p_0 F(s) + m_0(1 - F(s))]^{-1}$$
$$K(0, s) = p_s/m_0(1 - F(s)). \quad (15)$$

Assume for the moment that one realization occurs at time s and the next at time t. Then the covariance matrix for the joint probability of $X(t)$, $X(T) \mid X(s) = n(s)$ is easily found and the corresponding updating equations are given by

$$m_t = m_s + K(s, t)[n(t) - (n(s) + [m_s - n(s)]F(t \mid s))]$$
$$p_t = p_s - p_s^2 F(t \mid s)[p_s F(t \mid s) + (m_s - n(s))(1 - F(t \mid s))]^{-1}$$
$$K(s, t) = p_t/(m_s - n(s))(1 - F(t \mid s)), \qquad s \le t \le T. \quad (16)$$

The result generalizes to any number of updates at times $t_1, t_2, \ldots, t_n$ in the interval $(0, T)$ using $s = t_{j-1}$ and $t = t_j$; the proof that Equation (16) is the recursion for posteriors is by induction. The updating equations bear explanation because they have a deceptively simple structure. At time s, $n(s)$ events have occurred; based on this information the posterior mean for $X(T)$ is m_s. Thus $m_s - n(s)$ is the expected number yet to arrive in the interval (s, T); the number of *new* arrivals in the interval (s, t) is this expectation multiplied by $F(t \mid s)$. Since $n(s)$ are already certain, the total number expected by t must be the sum of $n(s)$ and $(m_s - n(s)) F(t \mid s)$. The difference between this sum and $n(t)$ is the forecast error in the interval (s, t). Clearly, the structure of the *time-dependent* and *data-dependent* Kalman gain $K(s, t)$ recognizes the length of the remaining interval, the number of events that have already occurred, and the location of the interval (s, t) within the season $(0, T)$; the latter obviously affects the likelihood of an event occurring in the subinterval, as the shape of $F(t \mid s)$ determines whether one is in an interval with "high" or "low" odds for the occurrence of events. For example, if $F(t \mid s)$ is small then the posterior variance in (16) is reduced less and $K(s, t)$ gives less weight to forecast errors than would have been the case if $F(t \mid s)$ were large. This behavior contrasts with seasonal models whose variance updating rules do not take recent observations or the future shape of the seasonal pattern into account. Finally, it is seen that the posterior calculations at the end of period (s, t) are identical in structure to those at the end of $(0, s)$ provided we use the conditional probability $F(t \mid s)$ in place of $F(s)$ and use $m_s - n(s)$ in place of m_0. Thus, attention focuses on the random number yet to arrive in future intervals.

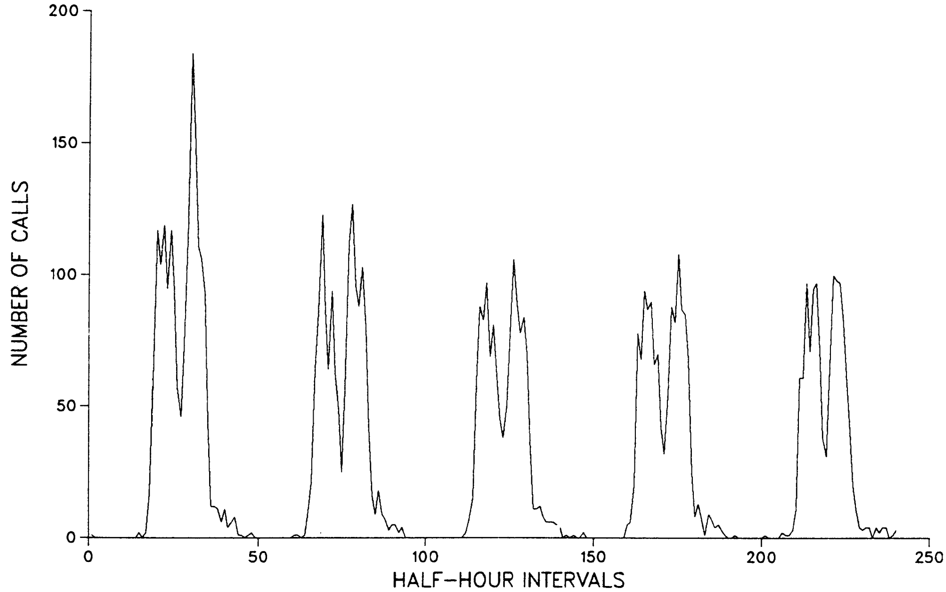

Figure 1. Incoming Half-Hourly Telephone Calls (University of Warwick).

In most applications it is useful to also allow for local random growth in the season total so that $N(t) = N(s) + V(s, t)$, which requires that a time-dependent variance term $\sigma_V^2(s, t)$ be included in the denominator of the variance recursions of Equations (15) and (16). To study local changes, it is common to assume that $V(s, t)$ is proportional to $t - s$.

8. END OF SEASON

It is interesting to see how forecast errors are weighted toward the end of the season and how the posterior variance decreases and our confidence in forecasts increases with increasing numbers of realizations. The coefficient of the forecast error in Equation (16) [$F(t)$ close to 1] is approximated by $p_0/(m_0(1 - F(t)) + p_0F(t)) \approx F(t)^{-1}$. In this case the posterior mean for total season demand approaches the maximum likelihood estimate, $N^* = n(t)F(t)^{-1}$, discussed earlier. If there is high confidence in the prior for the end-of-season total, forecast errors are given little weight, whereas an uninformative prior weights observations more heavily. Thus, in the former case the posterior mean will lie close to the prior mean, provided $F(t)$ is not too close to 1 and the season is not at an end. Rejection of the initial prior is always assured when $F(t)$ is close to 1. This effect can be more easily seen

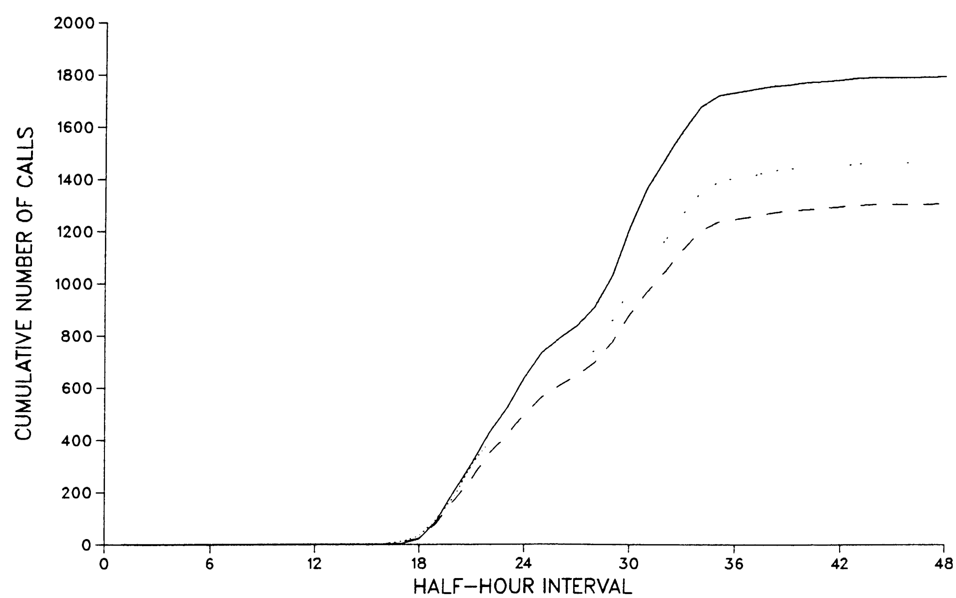

Figure 2. Daily Cumulative Incoming Telephone Calls (Monday, Tuesday, Wednesday).

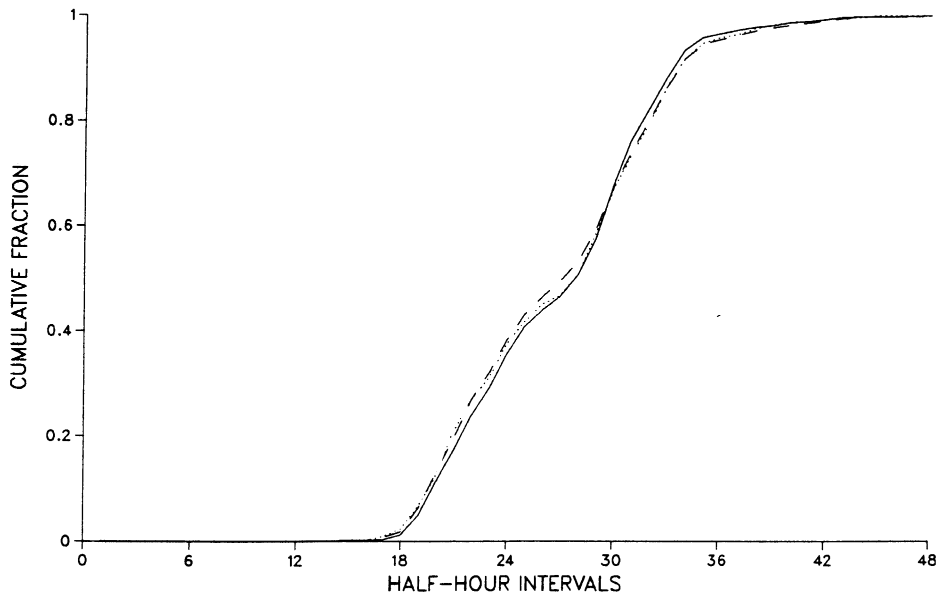

Figure 3. Daily Fraction of Incoming Telephone Calls.

by expressing results, particularly the updating equations, in terms of the normalized posterior variance

$$q_t = p_t/m_0, \qquad t \geq 0. \tag{17}$$

With the "Poisson-like" prior $q_0 = 1$, it is seen that on the first and every succeeding update we have $q_t = 1 - F(t)$ and $m_t - n(t) = m_0(1 - F(t))$, independent of $n(t)$ so that the cumulative number of arrivals expected in a future interval (s, t) is unaffected by the cumulative number of arrivals at time s; the result only depends on the fraction of expected total seasonal demand that has not yet arrived.

9. FORMULATION AS A DYNAMIC LINEAR MODEL

In the discrete time case with a total of p periods in a season, define $\mathbf{f} = (f_1, f_2, \ldots, f_p)^T$ as the seasonal fractions vector and $\mathbf{F}_p^T = (1, 0, 0, \ldots, 0)$. The seasonal fractions can be viewed as the fraction of total demand (throughout the entire season) that occurs in each period; they are dimensionless numbers between 0 and 1. If the length of the forecast horizon remains constant, we have a "rolling" forecast in which the season does not end at a fixed point in time as happens when forecasting the end of a fiscal, calendar, or sales year. If no

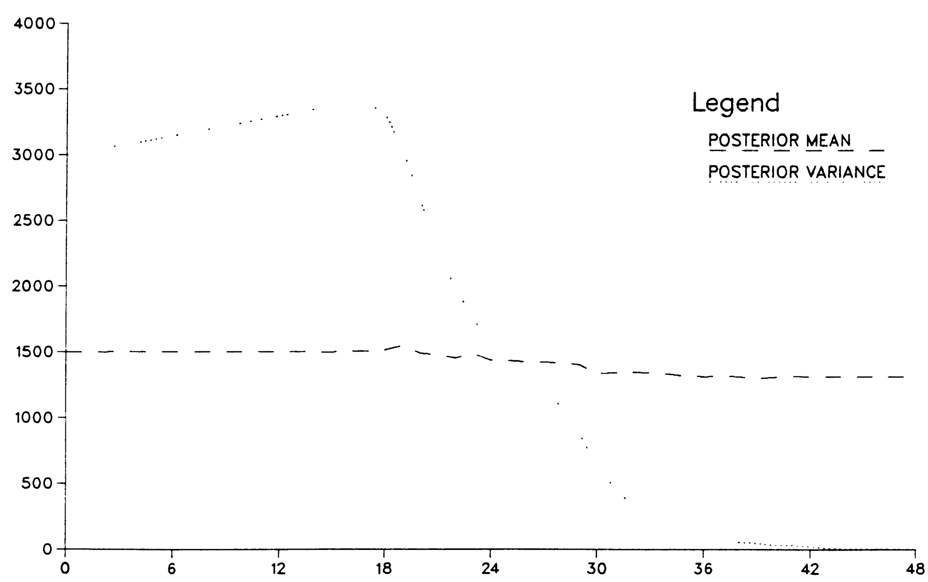

Figure 4. Posterior Mean and Variance of Thursday Total.

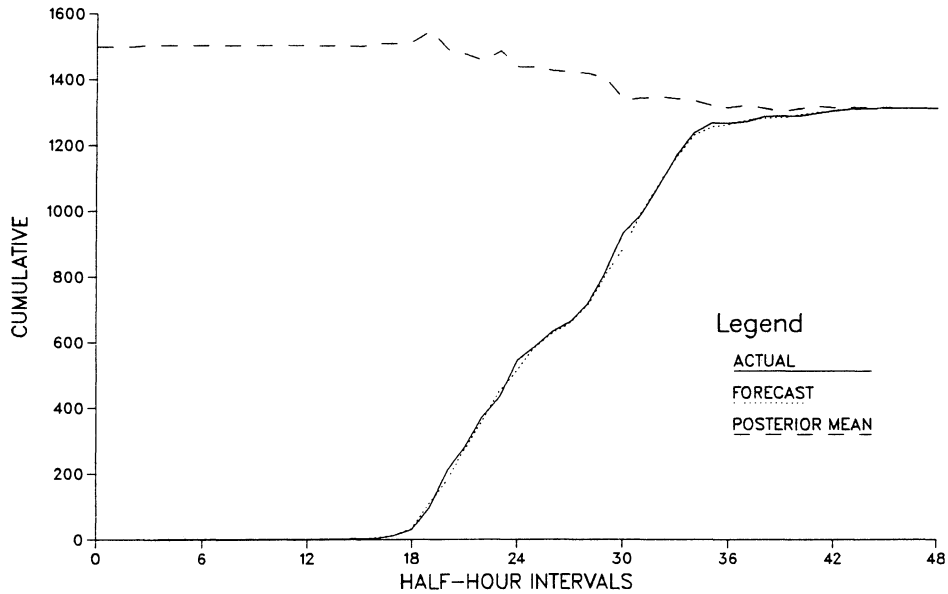

Figure 5. Actual and Half-Hourly Forecasts of Thursday Cumulative.

updating of seasonal fractions is made during the remainder of the season, the discrete multiplicative model can be rewritten as a DLM (West, Harrison, and Migon 1985). With $\mathbf{P}$ the square $(p \times p)$ cyclic permutation matrix, $\mathbf{I}_p$ the $(p \times p)$ diagonal matrix, $\mathbf{1}^T = (1, 1, \ldots, 1)$, and U_t, V_t the corresponding random measurement and system errors, we have

$$\begin{aligned} X_t &= N_t \mathbf{F}_p^T \mathbf{f}_t + U_t \\ N_t &= N_{t-1} + V_t \\ \mathbf{f}_t &= \mathbf{P}\mathbf{f}_{t-1} \\ \mathbf{1}^T \cdot \mathbf{f}_t &= 1. \end{aligned} \tag{18}$$

N_t is time-dependent, because it is allowed to vary randomly from period to period. If one conditions on N, total season demand, then X_t can be viewed as the conditional expectation of the amount to be received in the next period plus an error term. When X_t is binomial, that is, $X_t \sim \text{bin}(N, f_1(t))$, then $\text{var}[U_t \mid N] = Nf_1(t)(1 - f_1(t))$. With large N, it is reasonable to assume that $U_t \sim \mathbf{N}(0, N_t f_1(t)(1 - f_1(t)))$ and that N_t and V_t are also Gaussian with $N_0 \sim \mathbf{N}(m_0, p_0)$ and $V_t \sim \mathbf{N}(0, \sigma_V^2)$. Note, however, that in this formulation N_t has the interpretation of a season total extending forward p periods from now—that is, a season with a "rolling" horizon rather than one with a fixed ending date. As is well known (West et al. 1985), there is no difficulty

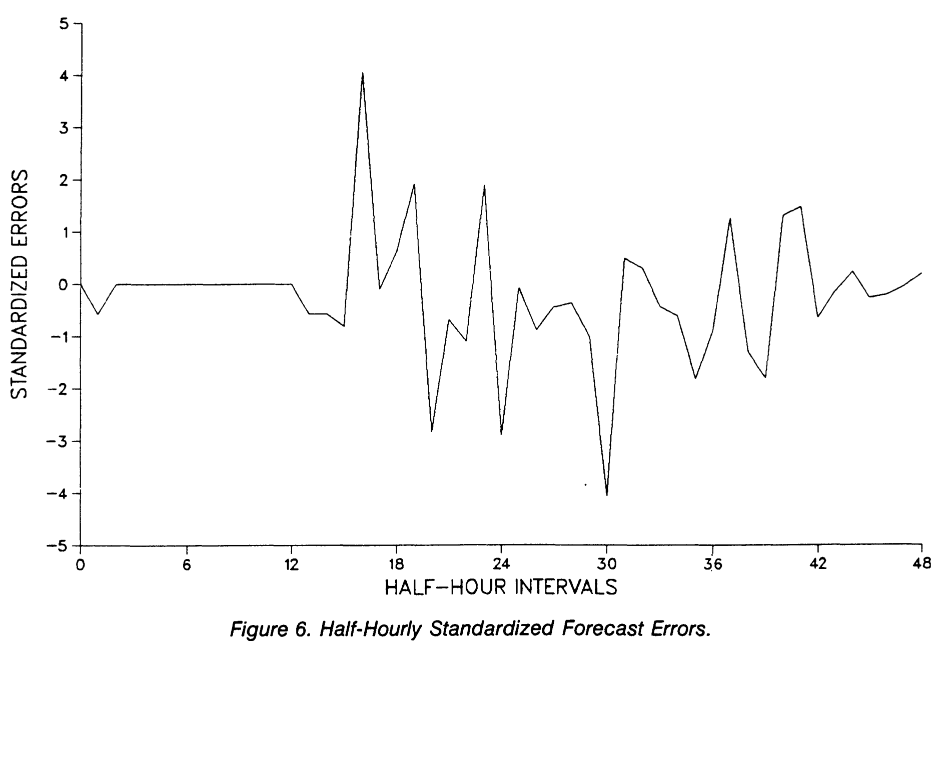

Figure 6. Half-Hourly Standardized Forecast Errors.

in adding trend terms or other linear constraints to the system equations in (18).

10. PREDICTING INCOMING HALF-HOURLY TELEPHONE CALLS: AN ILLUSTRATION

The half-hourly incoming telephone calls at the University of Warwick, England, Monday through Friday, September 6–10, 1982, are graphed in Figure 1 (also see West 1985). The first half hour of each day is between midnight and 12:30 a.m.; the last half hour is between 11:30 p.m. and midnight. The cumulative number of calls on each of the first three days is shown in Figure 2, and the corresponding cumulative fractions of each day are plotted in Figure 3. Approximately 20% of the half-hour periods record no incoming phone calls, particularly in the early morning hours of each weekday. In addition, there is a pronounced dip during each noon hour and an increased number of late afternoon calls, possibly due to the reduced rates for dialing. The shape of the cumulative number of calls and the cumulative fractions are seen to be similar except for scaling and total volume of demand probably associated with different weekday loads. The model of this article seemed particularly appropriate, since it is reasonable to assume that outside calls are made independently of one another, although there may be a general dependence on common external factors such as weather, proximity to holiday periods, and the timing of terms.

Using estimates for $F(t)$ based on the fractional totals of the first three days of the week and the theory and updating algorithms discussed in this article, a summary of results is given in Figures 4 and 5, one comparing the conditional expectation forecasts of cumulatives with actuals, the other comparing the posterior means and variances for the end-of-day totals calculated for Thursday, September 9. We have assumed that the level of the daily total varies randomly over time so that $N(t) = N(s) + V(s, t)$, with $E[V(t)] = 0$ and $\mathrm{var}[V(s, t)] = (t - s)\sigma_V^2$ (see, e.g., the DLM and GDLM models of West et al. 1983). Estimates for our choice of $F(t)$ were based on data obtained from the half-hourly records Monday through Wednesday of that week by simply dividing totals for each half hour by the total number of calls received in the three-day period. The amount of information carried by the early loads reduces the posterior variance by noon to about one-half of its assumed (prior) value of 3,000 and the (deliberately assumed) large prior mean of total calls from 1,500 to a posterior value of approximately 1,360 calls. The large prior mean for calls was purposely chosen to reflect a "bad" guess for each day's totals, even though experience indicated that total daily volume decreased throughout the week. Finally, Figure 6 is a plot of the one-step-ahead standardized forecast errors for each half hour. Naturally, the forecast errors have their largest values during the noontime period; even if the daily total were known precisely by 10:00 a.m., the binomial variance term has its largest value when $F(t)$ is close to $\frac{1}{2}$ and $F(t)(1 - F(t))$ close to $\frac{1}{4}$. This occurs about midday when the noontime peaks and dips occur. Furthermore, the estimates of $F(t)$ undoubtedly have their largest sampling errors during this same period. Except for the two morning outliers discussed by West (1985), standardized forecast errors lie within plus or minus 3.

West (1983) used a generalized linear model (GLM) with a total of seven scale parameters, three parameters for a quadratic model of lunchtime dip, a five-parameter Fourier component, and a daily parameter to make predictions beginning on Monday morning. The assumptions of his model are that phone calls are generated by a Poisson process in which the natural logarithm of the mean contains two components, one a quadratic effect to model lunchtime peaks and dips, the other a Fourier component to model daily seasonalities. Details of their GLM can be found in several references, notably West (1985) and West et al. (1985).

The major lesson to be learned seems to be that in the absence of a large amount of data, when the assumption of a stable, possibly complex, seasonal pattern appears reproducible from season to season, the model of this article is parsimonious in that it has two parameters that reflect prior information on the mean and variance of total end-of-season demand, is not troubled by missing data (witness the increasing posterior variance when no telephone calls occur in the early morning hours), does not require logarithmic or other transformations of the original data, depends on the commonsense notion that cumulative counts of events reflect the superposition of the actions of many individuals acting independently of one another, and has the appealing feature that the weights used to revise previous forecasts are both data- and time-dependent.

ACKNOWLEDGMENTS

I gratefully acknowledge the helpful comments and recommendations made by two referees and by Mike West and Gordon Newell.

[Received July 1985. Revised March 1986.]

REFERENCES

Britto, Mark, and Oliver, R. M. (1986), "Forecasting Donors and Donations to the Berkeley Engineering Fund," *Journal of Forecasting*, 5, 39–55.

Chang, S. H., and Fyffe, D. E. (1971), "Estimation of Forecast Errors for Seasonal Style-Goods Sales," *Management Science*, Ser. B, 16, 89–96.

Fildes, R., and Stevens, Colin (1978), "Look—No Data: Bayesian Forecasting and the Effects of Prior Knowledge," *Forecasting and Planning*, eds. R. Fildes and D. Wood, London: Saxon House and Gower Press, pp. 119–136.

Green, M., and Harrison, P. J. (1973), "Fashion Forecasting for a Mail Order Company Using a Bayesian Approach," *Operations Research Quarterly*, 24, 193–205.

Grinold, R. C., and Marshall, K. T. (1977), *Manpower Planning Models,* New York: Elsevier—North-Holland.

Marshall, K. T., and Oliver, R. M. (1970), "A Constant Work Model for Student Enrollments and Attendance," *Operations Research Journal,* 18, 193–206.

——— (1979), "Estimating Errors in Student Enrollment Forecasting," *Research in Higher Education,* 11, 195–205.

Olson, Jerome A. (1982), "Generalized Least Squares and Maximum Likelihood Estimation of the Logistic Function for Technology Diffusion," *Technological Forecasting and Social Change,* 21, 241–249.

West, Mike (1985), "Generalized Linear Models: Scale Parameters, Outlier Accommodation and Prior Distributions," in *Bayesian Statistics 2,* eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Amsterdam: North-Holland.

West, Mike, Harrison, P. J., and Migon, H. S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting," *Journal of the American Statistical Association,* 80, 73–83.