Forecasting an Accumulated Series Based on Partial Accumulation: A Bayesian Method for Short Series With Seasonal Patterns

Enrique DE ALBA
Departamento de Actuaría y Seguros, Instituto Tecnológico Autónomo de México (ITAM), México 01000 D.F., México (*dealba@itam.mx*)

Manuel MENDOZA
Departamento de Estadística, Instituto Tecnológico Autónomo de México (ITAM), México 01000 D.F., México (*mendoza@itam.mx*)

We present a Bayesian solution to forecasting a time series when few observations are available. The quantity to predict is the accumulated value of a positive, continuous variable when partially accumulated data are observed. These conditions appear naturally in predicting sales of style goods and coupon redemption. A simple model describes the relation between partial and total values, assuming stable seasonality. Exact analytic results are obtained for point forecasts and the posterior predictive distribution. Noninformative priors allow automatic implementation. The procedure works well when standard methods cannot be applied due to the reduced number of observations. Examples are provided.

KEY WORDS: Bayesian inference; Prediction; Stable seasonality; Time series.

In certain practical situations the problem of forecasting a future value of a time series must be faced with only a very small amount of data. This situation arises, for example, when a drastic change has occurred with respect to the observational conditions so that most past data become irrelevant. Of course, it also appears when a rather new series is being recorded, as when a new product or service is being launched. In any case, under such circumstances, most of the usual forecasting techniques are no longer applicable. In addition, sometimes the series $\{X_i\}$ to be considered is such that (a) $\{X_i\}$ represents the accumulated value of a positive and continuous variable X over the complete ith period (the ith year, say); (b) for a given, fixed, subinterval of each period it is possible to observe Y_i, a partial accumulation of X (the accumulated value of X for the first quarter of each year, for example); (c) for each period, we have $Y_i = W_i X_i$, where $\{W_i; i = 1, 2 \ldots\}$ are iid random variables over (0,1). We shall then say that the process has a stable seasonal pattern. A comprehensive review of situations in which this kind of structure arises can be found in Oliver (1987). It is clear that the seasonal pattern may be used in the forecasting process, no matter the amount of past data. However, it seems particularly valuable for the case in which only a few pieces of information are available. We present a Bayesian procedure for forecasting an accumulated value of a positive and continuous-valued time series, based on observation of a partial accumulation. The procedure is directed at obtaining forecasts given a small number of the partially accumulated observations, taking advantage of a seasonal pattern that may be known to exist in the series. The procedure is extremely simple to apply. We provide two examples. One of them is taken from Abraham and Ledolter (1983)—Monthly Average Residential Electricity Usage in Iowa City (Series 3). It is known to be a seasonal series. The second series is Administration Expenses of the Mexican Bank System (Guerrero and Elizondo 1997). As indicated by these authors, this series clearly has a trend.

It is important to emphasize that the procedure is especially useful when both conditions are present—seasonality and a small number of observations. If a large number of observations are available, then standard time series methods can be used to forecast the portion of the period that has not been observed yet. It can then be added to the partial accumulation already available. If few observations are available, then it is very unlikely that standard methods will yield good forecasts. If, in addition to having a small number of observations, it is known that the series is seasonal, then traditional methods such as autoregressive integrated moving average (ARIMA) are at a disadvantage. With monthly data, a typical seasonal autoregressive moving average (ARMA) model will include seasonal lags of order 12, which reduces the effective sample size by 12 (Abraham and Ledolter 1983, p. 283–291; Box and Jenkins 1970). If only 24 observations are available then the effective sample size would be 12. If in addition the series is nonstationary so that seasonal differencing is required (an ARIMA model), this will again reduce the effective sample size by 12: With a seasonal ARIMA model, the number of observations that can effectively be used in the estimation process is reduced by 24.

Previous Bayesian analyses for this problem were provided by Oliver (1987), de Alba (1988), and Lenk (1992), among others. The results in these works are based on approximations and/or somewhat involved models. These difficulties have already been pointed out by Guerrero and Elizondo (1997), who dealt with a similar problem using a frequentist approach and a large number of observations.

If a forecast is produced on the basis of a small amount of data, we feel that the modeling phase must be kept as simple

Journal of Business & Economic Statistics
January 2001, Vol. 19, No. 1

as possible. A Bayesian treatment for the case of a discrete variable was reported by de Alba and Mendoza (1996). Here, we present a Bayesian analysis of this problem in the case of a continuous-valued variable with a very simple model that does not require any kind of approximation.

1. THE MODEL

We now describe the problem in detail. On the basis of the two observed series $\{X_1, \ldots, X_T\}$ (the total observations series), $\{Y_1^r, \ldots, Y_T^r\}$ (the partial observations series up to subperiod r), and the partial observation Y_{T+1}^r, we must forecast the value of X_{T+1}. If, for example, X_i represents the accumulation of the underlying variable over the ith year and Y_i^r describes the partial accumulation corresponding to the first r months ($r \in \{1, \ldots, 11\}$) of that year we assume, following condition (c), that $Y_i^r = W_i^r X_i$, where $\{W_i^r; i = 1, 2 \ldots\}$ are independent random variables following the same distribution over (0,1).

To proceed with the analysis of this problem, it suffices to specify a model for the distribution of the random quantity W_i^r. There are many possible choices for $f(W_i^r|\theta_r)$. However, to obtain a result that may be easily applied in practice, we follow the simplest approach. The convenience of this choice will be illustrated later through a number of examples. Hence let us suppose that

$$f(W_i^r|\theta_r) = \theta_r (W_i^r)^{\theta_r - 1};\ W_i^r \in (0, 1), \tag{1}$$

where $\theta_r > 0$ is an unknown parameter. We are assuming the same subinterval, r, of each period and thus this parameter depends on this particular partial accumulation scheme. Under this specification, the expected proportion that the partial observations represent with respect to the total observations is constant for $i = 1, 2, \ldots, E(W_i^r) = \theta_r(\theta_r + 1)$.

From (1) we obtain the conditional pdf of Y_i^r as

$$P(Y_i^r|X_i, \theta_r) = \theta_r \left(\frac{Y_i^r}{X_i}\right)^{\theta_r - 1} \frac{1}{X_i}, \qquad 0 < Y_i^r < X_i. \tag{2}$$

Moreover, $\{Y_i^r, \ldots, Y_T^r\}$ are conditionally independent, given $\{X_1, \ldots, X_T\}$ and θ_r. Thus the likelihood function for θ_r, given the past data, is simply given by

$$L(\theta_r|\mathbf{Y}_T^r, \mathbf{X}_T) \propto \theta_r^T \left[\prod_{i=1}^T \left(\frac{Y_i^r}{X_i}\right)\right]^{\theta_r - 1},$$

where $\mathbf{Y}_T^r = (Y_1^r, \ldots, Y_T^r)'$ and $\mathbf{X}_T = (X_1, \ldots, X_T)'$. It is interesting to note that, if we define $V_i^r = -\ln W_i^r$, then, for a fixed value of r, $\{V_i^r; i = 1, 2, \ldots\}$ are iid variables following an exponential distribution with parameter θ_r. Thus, the log-ratios for a given accumulation subinterval are independent and identically distributed according to an exponential model across the years. It must be noted, however, that log-ratios corresponding to different months of the same year, say V_i^r and V_i^s with $r \neq s$, although exponentially distributed, are not independent and do not have the same distribution.

In connection with this result, it can be shown that, if we let $\{X_{ij}; i = 1, \ldots, T; j = 1, \ldots, 12\}$ be the series of monthly observations of the underlying variable so that $Y_i^r = \sum_{j=1}^r X_{ij}$, then

$$E(X_{ij}|\theta) = \left[\frac{\theta_j}{1+\theta_j} - \frac{\theta_{j-1}}{1+\theta_{j-1}}\right] E(X_i) \tag{3}$$

and

$$E(Y_i^r|\theta_r) = [\theta_r/(1+\theta_r)] \cdot E(X_i) = E(W_i^r) \cdot E(X_i), \tag{4}$$

where $\theta = (\theta_1, \ldots, \theta_{12})'$ with $0 = \theta_0 < \theta_1 < \theta_2 < \cdots < \theta_{12} = \infty$. From Equation (3) it follows that, if we adopt the model proposed in this article, the underlying level of the series may vary between years as well as within years. The expected value of the partial accumulation is a function of the (stable) proportion and of the expected total accumulation. One possible source of variation in the series between years can, of course, be the existence of a trend. Furthermore, since standard time series methods cannot be applied in situations in which our procedure can, it does not make much sense to try to find a formulation of one in terms of the other.

For ease of exposition the subscript/superscript r will be omitted in the following discussion. In Section 2, we produce a posterior distribution for the parameter θ using the past information $(\mathbf{Y}_T, \mathbf{X}_T)$ and then use this to this obtain a posterior predictive distribution for the quantity of interest, X_{T+1}.

2. PRIOR–POSTERIOR ANALYSIS

To describe the available knowledge about θ, it suffices to assess a prior distribution $P(\theta)$ for this parameter to obtain, via Bayes theorem, the corresponding posterior as $P(\theta|\mathbf{Y}_T, \mathbf{X}_T) \propto P(\theta)L(\theta|\mathbf{Y}_T, \mathbf{X}_T)$.

The prior may be any probability function describing the initial state of knowledge of the researcher with respect to θ or, equivalently, with respect to $E(W)$, which may be easier to elicit. However, if we are dealing with a situation in which only a small amount of past information is available, an informative prior may be crucial in the final results.

If the researcher is unwilling to involve his/her prior beliefs concerning θ or any other unknown quantity, he/she might prefer to obtain a posterior mainly based on the data and thus to use a noninformative prior. In this case we have a very simple regular model and only one parameter involved. Thus it is straightforward to verify that, if we adopt Jeffreys's rule (Box and Tiao 1973), we get the noninformative prior

$$P(\theta) \propto \frac{1}{\theta},$$

so

$$P(\theta|\mathbf{Y}_T, \mathbf{X}_T) \propto \theta^{T-1} \left[\prod_{i=1}^T \left(\frac{\mathbf{Y}_i}{\mathbf{X}_i}\right)\right]^{\theta - 1}.$$

Equivalently, $P(\theta|\mathbf{Y}_T, \mathbf{X}_T) \propto \theta^{T-1} \exp(-\lambda_T \theta)$, where $\lambda_T = \sum_{i=1}^T (\ln X_i - \ln Y_i) > 0$. Thus, $P(\theta|\mathbf{Y}_T, \mathbf{X}_T)$ is a *gamma* distribution $\mathrm{Ga}(\theta|T, \lambda_T)$.

3. THE FORECAST PROBLEM

If we recall that the basic objective is to forecast X_{T+1}, we may refer to the corresponding likelihood function $L(X_{T+1}, \theta|\mathbf{X}_T, \mathbf{Y}_T, Y_{T+1}) = P(\mathbf{Y}_T, Y_{T+1}|\mathbf{X}_T, X_{T+1}, \theta)$

$$\propto \theta^{T+1}\left[\prod_{i=1}^{T+1}\left(\frac{Y_i}{X_i}\right)\right]^{\theta-1}\frac{1}{X_{T+1}}, \tag{5}$$

where both X_{T+1} and θ are unknown and the quantity of interest is precisely X_{T+1}. On this basis, we must produce the joint posterior distribution $P(X_{T+1,\theta|\mathbf{Y}_T, \mathbf{X}_T, Y_{T+1}}) \propto P(\mathbf{Y}_T, Y_{T+1}|\mathbf{X}_T, X_{T+1}, \theta)P(X_{T+1}, \theta|\mathbf{X}_T)$.

To this purpose, if we assume that X_i (for $i = 1, 2, \ldots, T+1$) and θ are independent (a sensible hypothesis if X_i describes the total accumulated value for the ith period and θ refers to the distribution of this total over the period), we get

$$\begin{aligned} P(X_{T+1}, \theta|\mathbf{X}_T) &= P(\theta|\mathbf{X}_T)P(X_{T+1}|\mathbf{X}_T) \\ &= P(\theta)P(X_{T+1}|\mathbf{X}_T). \end{aligned}$$

Hence, we have

$$\begin{aligned} &P(X_{T+1}, \theta|\mathbf{Y}_T, \mathbf{X}_T, Y_{T+1}) \\ &\quad\propto \theta^{T+1}\left[\prod_{i=1}^{T+1}\left(\frac{Y_i}{X_i}\right)\right]^{\theta-1}\left(\frac{1}{X_{T+1}}\right)\left(\frac{1}{\theta}\right)P(X_{T+1}|\mathbf{X}_T) \\ &\quad\propto \theta^{T}\exp(-\lambda_{T+1}\theta)P(X_{T+1}|\mathbf{X}_T) \\ &\quad\propto \lambda_{T+1}^{-(T+1)}\mathrm{Ga}(\theta|T+1, \lambda_{T+1})P(X_{T+1}|\mathbf{X}_T). \end{aligned}$$

Thus, for any prior distribution $P(X_{T+1}|\mathbf{X}_T)$, the nuisance parameter θ can be integrated to obtain $P(X_{T+1}|\mathbf{Y}_T, \mathbf{X}_T, Y_{T+1}) \propto \lambda_{T+1}^{-(T+1)}P(X_{T+1}|\mathbf{X}_T)$. At this point, it must be noticed that $\lambda_{T+1}^{-(T+!)} = [\lambda_T + \ln(X_{T+1}) - \ln(Y_{T+1})]^{-(T+1)}$ so that, as a function of X_{T+1}, this factor of the posterior distribution is bounded and decreases to 0 as X_{T+1} goes to infinity, although it is not integrable over $[Y_{T+1}, \infty)$. In any case, and to obtain the posterior distribution of the quantity of interest, we must elicit the prior $P(X_{T+1}|\mathbf{X}_T)$.

It is interesting to note that this prior must reflect the knowledge we have with respect to X_{T+1} if we ignore the partial values Y_i for $i = 1, 2, \ldots, T+1$. It might, however, take into account the information provided by the total observations X_i (for $i = 1, 2, \ldots, T$). Thus, if we are able to forecast the quantity of interest using only the total observations series, the resulting predictive distribution could be used for this purpose. In any case, the procedure proposed in this article is intended to deal with a situation in which only a small amount of data is available (in fact, in our examples $T = 2$). Under such circumstances, the idea of having a reasonable model for the total observation series is rather unrealistic.

As an alternative, we can use the fact that, in our model, X_{T+1} only appears in the density function of Y_{T+1} and plays the role of a scale parameter. Thus, if we restrict our attention to that structure, a (noninformative) prior distribution for X_{T+1} is given by

$$P(X_{T+1}) \propto \frac{1}{X_{T+1}}.$$

This prior distribution describes the absence of information regarding this quantity when $\mathbf{X}_T, \mathbf{Y}_T$ and Y_{T+1} are ignored. Hence, we can finally get the *exact* expression

$$P(X_{T+1}|\mathbf{X}_T, \mathbf{Y}_T, Y_{T+1}) = \frac{T\lambda_T^T}{X_{T+1}}\left[\lambda_T + \ln\left(\frac{X_{T+1}}{Y_{T+1}}\right)\right]^{-(T+1)},$$

where $Y_{T+1} < X_{T+1} < \infty$. This posterior has a number of interesting properties. It is always proper, and its density is strictly monotone decreasing. Therefore, the posterior mode for X_{T+1} is precisely Y_{T+1}. The moments of this distribution do not exist, but the quantities can be easily calculated as $X_{T+1}^{(q)} = Y_{T+1}(\dot{P}_T)^{-T[(1-q)^{-1/T}-1]}$ for $0 < q < 1$, where

$$\dot{P}_T = \left[\prod_{i=1}^{T}\left(\frac{Y_i}{X_i}\right)\right]^{1/T} \tag{6}$$

is the geometric mean of the observed proportions. In particular, the posterior median is given by

$$X_{T+1}^{(.5)} = Y_{T+1}(\dot{P}_T)^{-T(2)^{1/T}-1}.$$

In this later case, since $\ln 2 \le T[2^{1/T} - 1] \le 1$, we have that, roughly speaking, the posterior median X_{T+1}^B must satisfy the condition

$$Y_{T+1} = \dot{P}_T(X_{T+1}^B). \tag{7}$$

Clearly, this quantity can be used as a forecast of X_{T+1}. It is interesting to note that, if we proceed this way, the pointwise forecast for the $(T+1)$th total can be obtained from the observation on the $(T+1)$th partial accumulation and a point estimate of the average proportion over the past periods. The idea is simple but it is interesting to note that the estimate for the average proportion is not the rather intuitive arithmetic mean but the geometric mean of the observed proportions (more in general, a power of this geometric mean). As an additional remark it can be shown that, if $T \to \infty$, we get $X_{T+1}^{(q)} = Y_{T+1}(\dot{P}_T)^{\ln(1-q)}$.

Hence, the limiting posterior is not a degenerate distribution. This fact is reassuring if we recall that we are not estimating an unknown parameter but forecasting a future value of a random variable. A more general approach to pointwise forecasting should include the specification of a loss function reflecting the researcher's preferences. It must be remarked that, since the posterior distribution for X_{T+1} has no moments, this loss function must be carefully selected to guarantee the existence of the integrals involved. One possible choice is the following (de Alba and Van Ryzin 1980):

$$\ell(X_{T+1}, \widehat{X}_{T+1}) = \begin{cases} a\left(\frac{1}{X_{T+1}} - \frac{1}{\widehat{X}_{T+1}}\right) & X_{T+1} < \widehat{X}_{T+1} \\ \left(\frac{1}{\widehat{X}_{T+1}} - \frac{1}{X_{T+1}}\right) & X_{T+1} \ge \widehat{X}_{T+1}. \end{cases} \tag{8}$$

This loss function penalizes underestimation differently from overestimation and the discrepancy between forecast and true value is measured on a reciprocal scale. It is easy to show that under this loss function, the optimal choice for the forecast is $X_{T+1}^{(q)}$, where $q = 1/(a+1)$.

In Section 4, we discuss some relevant issues concerning the implementation of the procedure proposed herein.

4. IMPLEMENTATION

In this section we establish some criteria that are necessary to make our method operational. On the one hand, we need to have a practical criterion to decide how stable the seasonal pattern is. One way to do it is, of course, to look at the autocorrelation function, or at the partial autocorrelation function, or at the periodogram, or at the spectrum, and so forth of the series. However, with only 12 observations in a series of monthly data, they will not be of much help. One way is to look directly at the plot of the series overlaying different years. We found that looking at the plot of the cumulative proportion of the series up to a given period for each one of the years being considered may be useful. If the lines lie very close together, then we have evidence in favor of a stable pattern. On the other hand, if they tend to wander in separate directions, then this is an indication of nonstability. Let us see in more detail.

As in Section 1, let $\{X_1, X_2, \dots, X_T\}$ be the observed series of totals for the T years; that is, X_t is the total in the tth year, $t = 1, \dots, T$. Also let $\{X_{i1}, X_{i2}, \dots, X_{i12}\}$ be the series of partial observations in each year; that is, X_{ij} is the total observation in month j of the ith year, thus $Y_i^r = \sum_{j=1}^r X_{ij}$ is the cumulative total observed up to the rth month in the ith year, $r = 1, \dots, 12$. Then the cumulative proportion in year t and month r is $P_{ir} = Y_i^r / X_i$. Clearly $P_{i12} = 1$, since $Y_i^{12} = X_i$. We plot P_{ir} versus r for each i. This plot allows us to determine graphically if there is a stable seasonal pattern in the series. In this case the plots of the P_{ir} for each i should all be very close to each other.

In each one of the examples, we will compare the performance of our procedure with that of using an ARIMA model based on the complete (seasonal) time series. We will also compare it with the results of fitting an ARIMA only to the last two available years of data, since this is the kind of situation for which our proposal is intended. Several criteria are normally used to compare the performance of alternative forecasting approaches—for example, mean absolute deviation, mean squared error (MSE), or others (Abraham and Ledolter 1983). However these measures cannot be computed in the kind of situations we are considering because we are using different information sets—that is, the partial accumulations—to forecast one quantity, the $(T+1)$st total. In addition, the expected value of the predictive distribution does not exist. We use the following measure $\text{MSE}_m = \sum_{r=1}^{11} (X_{T+1,r}^{(m)} - X_{T+1})^2 / 11$ instead. This is the MSE between the 11 estimators of X_{T+1}, one for the partial accumulation up to each month.

This expression for MSE corresponds to the more general definition as the average squared deviation from the true value, which we do have (Mood, Graybill, and Boes 1974, p. 291). The subscript m indicates which method was used; that is, $m =$ Bayes, ARIMA, TS/2YR depending on whether we estimate using our Bayesian method, an ARIMA model based on the complete time series, or only the last two available years of data. Evidently, in the examples $X_T + 1$ is a known value.

The value of q (or $a = q/(1-q)$) in the loss function (8) must be specified by the statistician. This determines which quantile yields the best Bayesian estimator for the total in year $T+1$. In addition, the value for q will be an indication of the relative costs of underestimation versus overestimation with this loss function. Evidently, the choice of q need not be the same in all problems. In fact the choice of q may vary when using different partial accumulations to estimate the same accumulated total. We analyze the behavior of the forecasts for alternative values of q, but using the same q at different partial accumulations with each series. We actually have 11 estimates of the total in year $T+1$, one obtained from each one of the partially accumulated proportions. Let $X_{T+1,r}^{(q)}$ be the qth quantile of the predictive distribution of X_{T+1} obtained from the partial information up to the nth month. We proceed as follows, based on the 11 estimators $X_{T+1,r}^{(q)}$, $r = 1, \dots, 11$, which we can get for X_{T+1}. We obtain the loss function (8) for each value of q and r as

$$\ell(q,r) = \ell(X_{T+1}, X_{T+1,r}^{(q)}) = \begin{cases} a\left(\frac{1}{X_{T+1}} - \frac{1}{X_{T+1,r}^{(q)}}\right) & X_{T+1} < X_{T+1,r}^{(q)} \\ \left(\frac{1}{X_{T+1,r}^{(q)}} - \frac{1}{X_{T+1}}\right) & X_{T+1} \geq X_{T+1,r}^{(q)} \end{cases}$$

and from there compute the "total loss" for each q:

$$L(q) = \sum_{r=1}^{11} \ell(q,r) = \sum_{r=1}^{11} \ell(X_{T+1}, X_{T+1,r}^{(q)}). \tag{9}$$

This total loss will be used to obtain some idea of which would have been a good choice of q in our examples.

5. APPLICATIONS

In this section we present examples of application of the methods developed previously for two datasets. One of the series considered is clearly seasonal. It is Monthly Average Residential Electricity Usage in Iowa City (in kilowatt-hours), January 1971 to October 1979. It is series number 3 of Abraham and Ledolter (1983). These authors presented this series, along with others, as an example of seasonal ARIMA time series and analyzed it in detail in chapter 6 of their book. Hence the seasonality of this series has been formally verified by standard time series procedures and is known to be seasonal. We also include an application to the series Administration Expenses of the Mexican Bank System (Guerrero and Elizondo 1997). As indicated by these authors, this series clearly has a trend. We use it to illustrate the performance of our procedure for a series with trend.

In both cases we have a number of observations large enough so that each series is first analyzed using standard procedures. However, for the purpose of comparison we then assume that only two years of data are available and apply our procedure. It is also clear, as will be seen from the results, that, whenever there is enough information to forecast using standard procedures and the series has a clear, stable (through time) seasonal pattern, they will yield better results than our method. It is also clear that when only two years of monthly data are available it is usually impossible to formally detect seasonal behavior in a given series. There are simply not enough observations to formally verify the seasonality, even though it is known to be seasonal. Let us recall that with monthly data a seasonal ARMA model will include seasonal

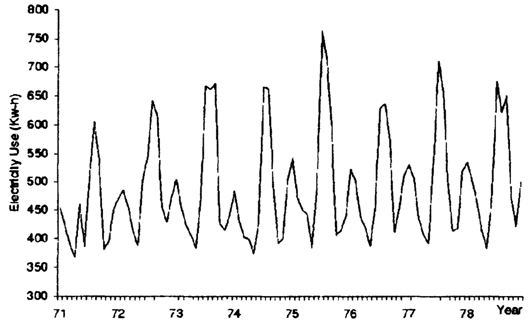

Figure 1. Monthly Average Residential Electricity Usage in Iowa City (in kilowatt-hours), January 1971 to October 1979 (Abraham and Ledolter 1983).

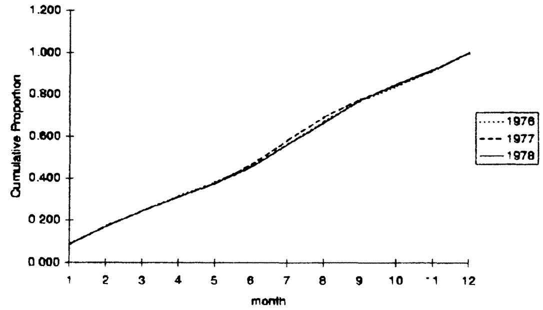

Figure 3. Plot of the Monthly Cumulative Proportions of P_{ik}, $k = 1, \ldots, 11$ for Monthly Average Residential Electricity Usage in Iowa City (1976–1978). They all practically coincide, indicating the existence of a stable seasonal pattern.

lags of order 12, reducing, then, the effective sample size by 12 (Abraham and Ledolter 1983; Box and Jenkins 1970). If in addition seasonal differencing is required, this will again reduce the effective sample size by 12. Moreover, if the seasonal pattern has been changing over time, our method proves definitely superior, and these are situations that are not uncommon in practice.

Figure 1 shows the complete series for Monthly Average Residential Electricity Usage in Iowa City ("Electricity in Iowa" for short) from January 1971 to December 1978. In Figure 2 the data for the last three years, 1976 to 1978, have been overlaid. Clearly, by visual inspection, the series is seasonal with peaks in the summer and winter.

The analysis carried out by Abraham and Ledolter (1983) for this series using all the observations yields a seasonal ARIMA model. They identified and fitted the following seasonal model:

$$(1 - B^{12})z_t = \theta_0 + (1 - \theta_1 B)(1 - \Theta B^{12})a_t. \qquad (10)$$

We will use this model when comparing this standard procedure with ours. It will be considered as the best model, and we will not question their results. To exemplify the application of our proposed method to forecast the total for the year 1978, we assume that only data for two years, 1976 and 1977, are available. According to the process we indicated in Section 4, we first look at the plot of $P_{ik}, k = 1, \ldots, 11$. It is given in Figure 3 for 1976 and 1977. It is clear that they all practically coincide, indicating the existence of a stable seasonal pattern.

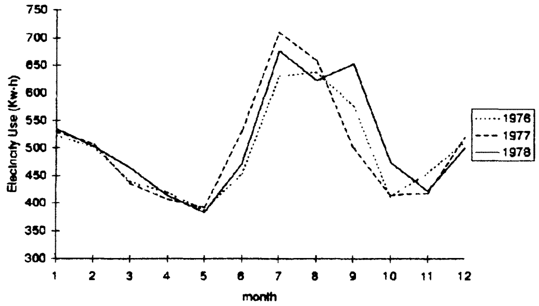

Figure 2. The Last Three Years of Data (1976–1978) for Monthly Average Residential Electricity Usage in Iowa City Are Overlaid, Month by Month. The series is seasonal with peaks in the summer and winter.

Now, to analyze the effect of using alternative values of q, we compute and plot $L(q)$ for $q = .275(.025).8$. From Figure 4, it can be seen that the minimum of $L(q)$ is attained for $q = .557$. It is interesting to note that this particular value of q is the one that makes the exponent in (6) practically equal to 1 (1.005) so that if we had used this value of q we would have estimated Total Electricity in Iowa for the year 1978 using the approximate estimate given by (7).

The corresponding column in Table 1 for the estimated total in 1978 shows the estimates obtained using the Bayesian procedure with this value of q.

We now compare the forecasts of the total for 1978 under different alternatives. The first alternative is to use all the data and Model (10). The results of this exercise are given in Table 1, in the column labeled ARIMA. The second option is to use our procedure assuming we only have data for 1976 and 1977. According to the result in the previous paragraph, we use Formula (7). The estimates are shown in the column labeled Bayes. Finally, for the sake of completeness we use only the data for 1976 and 1977 and apply standard time series methods. We follow the usual procedure of identification, estimation, diagnostic, and forecast (Box and Jenkins 1970). With such a small number of observations it was impossible to identify a seasonal component, even though it is well known that the use of electricity has a very strong seasonal component. The model identified with two years of data was a

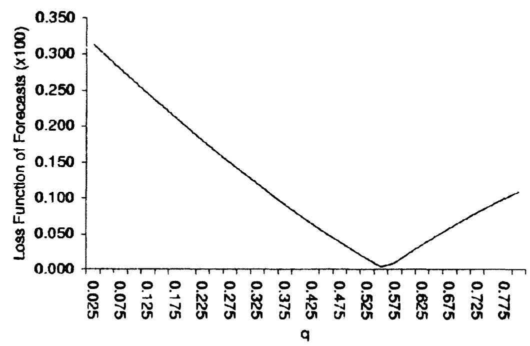

Figure 4. Total Loss Function L(q) for q = .275(.025).8 Computed for Monthly Average Residential Electricity Usage in Iowa City in 1978. The minimum of L(q) is attained at q = .557. This particular value of q makes the exponent in Equation (6) practically equal to 1 (1.005), and the total can be estimated using the geometric mean.

Table 1. Electricity Use in Iowa (Kw-h): Comparison of Results Using Three Different Forecasting Methods

	Observed			*1978 estimated total*		
	1976	*1977*	*1978*	*Bayes*	*ARIMA*	*TS/2 YR*
Jan	523	530	535	6,081	6,151	6,203
Feb	502	507	503	6,025	6,158	6,001
Mar	439	436	464	6,120	6,174	5,516
Apr	420	407	414	6,092	6,134	5,085
May	387	392	383	6,056	6,091	4,928
Jun	453	531	472	6,001	6,088	5,928
Jul	630	710	676	6,009	6,112	7,275
Aug	637	658	622	5,968	6,028	6,237
Sep	576	500	652	6,116	6,139	6,809
Oct	411	414	474	6,179	6,153	6,004
Nov	455	418	422	6,147	6,115	6,044
Dec	512	520	501	6,118	6,118	6,118
Total	5,945	6,023	6,118			
MSE				6,123	1,588	431,861
Ratio				1.00	.26	70.53

nonseasonal ARMA (0,1):

$$z_t = \theta_0 + (1 - \theta_1 B)a_t. \tag{11}$$

This shows that with a limited number of observations it is not necessarily possible to detect seasonality, even if we do know a series is seasonal and even though the seasonality can be verified when enough observations are available. The forecasts obtained from (11) are labeled TS/2YR in Table 1.

Table 1 shows the observed values of the series for the last three years, together with the estimate for the year total obtained for 1978 using partial accumulation up to a given month. This is shown for each one of the three approaches. We include the corresponding value of MSE_m: m = Bayes, ARIMA, TS/2YR. As would be expected, the ARIMA model fitted to the 1971–1977 data yields the lowest MSE, which is .26 of the one obtained using the Bayesian procedure.

The MSE from the forecasts generated by identifying and fitting an ARIMA model only to the data for 1976 and 1977 is much larger that any of the other two. It should be stressed that, under our assumption that only data for these years are available, the Bayesian procedure is much better than trying to use standard methods. All the standard time series analyses were done using MINITAB (Minitab, Inc. 1994). Similar results were obtained when applying our method to series number 4 of Abraham and Ledolter (1983), Monthly Car Sales in Quebec.

The second example is carried out with the series Administration Expenses of the Mexican Bank System (Guerrero and Elizondo 1997). We use it to illustrate the performance of our procedure for a series with trend. This fact is clearly indicated by the authors. Figure 5 shows the complete series (in thousands of new pesos at 1980 value), January 1987 to October 1994 ("Bank Expenses" for short). In Figure 6, the data for the last three years, 1992 to 1994, have been overlaid. Note that the last year is not complete. Visual inspection shows that the series has a trend.

Guerrero and Elizondo (1997) obtained a seasonal ARIMA model and used it to forecast. Using all the data for the years 1987 to 1993, they identified and fitted the following seasonal model:

$$(1 - B^{12})z_t = \theta_0 + (1 - \theta_1 B)(1 - \Theta B^{12})a_t. \tag{12}$$

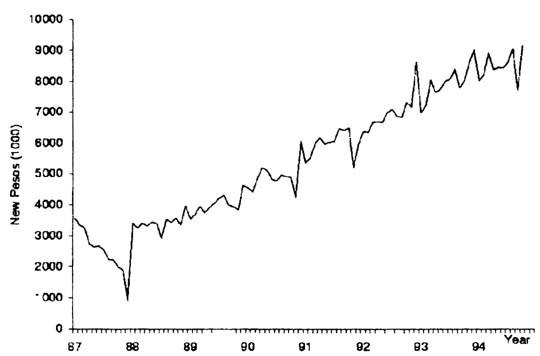

Figure 5. Real Monthly Administration Expenses of the Mexican Bank System (in thousands of new pesos at 1980 prices), January 1987 to October 1994 (Guerrero and Elizondo 1997). This series has a trend.

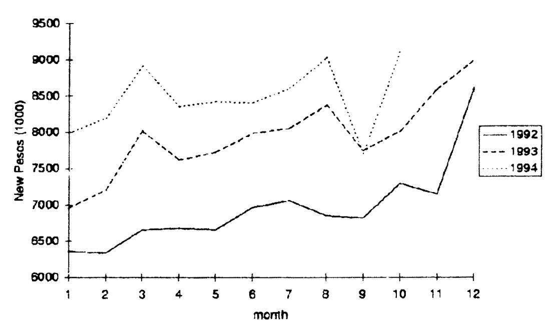

Figure 6. The Last Three Years of Data (1992–1994) for Monthly Administration Expenses of the Mexican Bank System Are Overlaid, Month by Month. The series has a slight seasonality with peaks in March and August. The last two months of 1994 were not available.

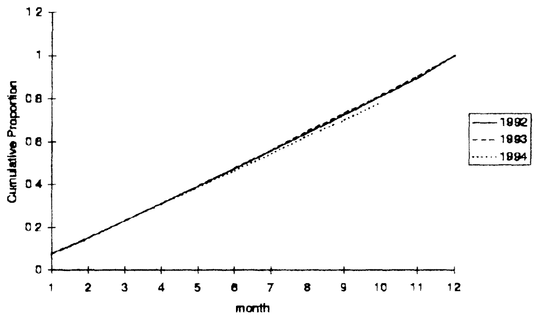

Figure 7. Plot of the Monthly Cumulative Proportions of P_{ik}, $k = 1, \ldots, 11$, for Monthly Administration Expenses of the Mexican Bank System (1992 and 1993). The data only allow computing of P_{ik}, $k = 1, \ldots, 10$, for 1994. They all practically coincide up to October, indicating the existence of a stable seasonal pattern.

They used this model to obtain forecasts for Total Bank Expenses in 1994 and provided the resulting forecasts. We will use their forecasts directly when comparing this standard procedure with ours. As in the previous cases, we do not question their results.

Now let us assume that only data for two years, 1992 and 1993 are available. The plot of $P_{ik}, k = 1, \ldots, 10$ (the partial accumulation for $k = 11$ is not available in this case) is given in Figure 7 for these years. They practically coincide, due to the existence of a stable seasonal pattern. The values of $L(q)$ for $q = .325(.025).8$ are plotted in Figure 8, where it can be seen that the minimum of $L(q)$ is also attained for $q = .557$. This value of q makes the exponent in (6) practically equal to 1. In fact it is essentially the same value as in the previous example. Again, we would have used (7). The corresponding column in Table 2 for the estimated total in 1994 shows the estimates obtained using the Bayesian procedure with this value of q.

We now compare the forecasts of the total for 1994 under different alternatives. The first alternative is to use all the data (Guerrero and Elizondo 1997). The results of this exercise are given in Table 2, in the column labeled ARIMA. The second option is our procedure using only data for 1992 and 1993. In this case we also use Formula (7). The estimates are presented in the column labeled Bayes.

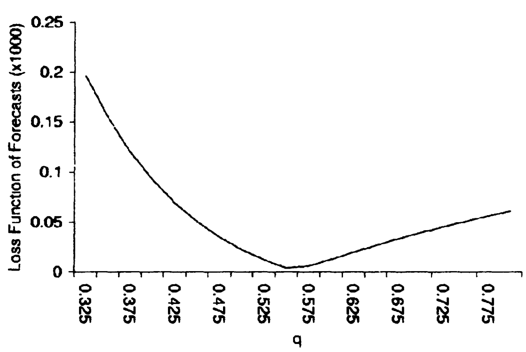

Figure 8. Total Loss Function L(q) for $q = .325(.025).8$ Computed for Monthly Administration Expenses of the Mexican Bank System in 1994. The minimum of L(q) is attained at $q = .557$. This particular value of q makes the exponent in Equation (6) practically equal to 1(1.005), and the total can be estimated using the geometric mean.

Finally, for the sake of completeness we assume that only the data for 1992 and 1993 are available and apply standard time series methods. With such a small number of observations it was impossible to identify a seasonal component. The trend is evident in the fact that differences need to be taken. The model identified with two years of data was a nonseasonal ARIMA (0,1,1):

$$(1 - B)z_t = \theta_0 + (1 - \theta_1 B)a_t. \tag{13}$$

Again, with a limited number of observations it is not possible to detect seasonality, even if we do know the series is seasonal and even though the seasonality can be verified when enough observations are available. However, this model does reflect the existence of a trend in the series, as is evidenced by the presence of a first difference. The forecasts obtained from (13) are labeled TS/2YR in Table 2.

Table 2 shows the observed values of the series for the last three years, together with the estimate for the year total obtained using partial accumulation up to a given month. This is shown for each one of the three approaches. We include the corresponding value of MSE_m: m = Bayes, ARIMA, TS/2YR. To compute the MSE we add a piece of information—the observed year total for 1994. It was obtained directly from A. Elizondo. We do not include the data for November and December of that year so that the results in their article may not be affected. Thus the MSE's used for comparison are based only on the estimates of the total accumulation for 1994, obtained with the partial accumulations from January through October. Once more, the ARIMA model fitted to the 1987–1993 data yields the lowest MSE, which is .78 of the one obtained using the Bayesian procedure. The MSE from the forecasts generated by identifying and fitting an ARIMA model only to the data for 1992 and 1993 is much larger that any of the other two. In this case, again assuming that data are available only for two years and given the fact that the series has a trend, the Bayesian procedure is better.

6. FINAL REMARKS

In this article, a very simple model is proposed to deal with the problem of forecasting a cumulative variable using its partially accumulated data. The basic assumptions are stability of the seasonal pattern and that only a few observations are available. We adopt a Bayesian approach, and exact results are provided for pointwise forecasts as well as for the entire predictive distribution.

It must be noticed that the notion of accumulated observation is only properly defined in the case of positive random variables. It does not make sense to talk about proportions if there are negative values in the series. Thus, the proposed procedure is not intended to be generalized to the case of real valued data.

Our procedure can be considered highly competitive if only a small amount of past data is available. A number of applications with very different series show that, in this case, this proposal outperforms other well-known techniques. Specific results are obtained under a Bayesian framework assuming noninformative prior distributions so that it is not necessary to

Table 2. Expenses of the Mexican Bank System (thousands of new pesos): Comparison of Results Using Three Different Forecasting Methods

	Observed			1994 estimated total		
	1992	1993	1994	Bayes	ARIMA	TS/2 YR
Jan	6,357	6,954	7,988	107,106	107,107	101,950
Feb	6,332	7,206	8,196	107,637	107,348	102,598
Mar	6,657	8,020	8,913	108,011	107,158	104,235
Apr	6,676	7,618	9,356	107,101	107,027	103,999
May	6,657	7,727	8,426	106,616	106,779	104,387
Jun	6,964	7,988	8,407	105,540	106,259	105,167
Jul	7,057	8,051	8,604	104,971	105,869	106,431
Aug	6,849	8,373	9,026	105,134	105,278	106,897
Sep	6,816	7,747	7,703	103,942	104,367	106,135
Oct	7,291	8,016	9,113	104,185	104,562	106,135
Nov	7,143	8,589				
Dec	8,610	8,997				
Total	83,409	95,286	108,152*			
MSE				6.4354E+.06	5.0147E+06	13766378
Ratio				1.00	.78	2.75

NOTE: The total for 1994 was obtained directly from A. Elizondo, and it was used to compute the MSE.

"elicit" elaborate ones. The small amount of sample information available is used to its full extent and the results can be applied automatically. It must be stressed, however, that other *informative* priors might be used although analytic expressions will depend on the specific form of those priors. In particular, if some information is available from past data concerning the level of X, it must be described through the corresponding prior. As for the stable seasonal pattern hypothesis, it must be noted that many real series show this behavior, and, more importantly, this assumption does not prevent the series from having a trend.

We have illustrated our proposal with real examples in which the seasonal pattern can be easily verified. In fact the procedure can be extended to other types of series, such as inflation. At first sight it may appear that this is not a series in which the accumulated year total can be obtained from partial (monthly) accumulations. However, if CPI_i is the Consumer Price Index in the ith period and inflation between periods i and j is defined as $Y_{ij} = \ln(\text{CPI}_i) - \ln(\text{CPI}_j)$, it is easy to see that the inflation rate between periods i and $i+s$, $s>0$, is $Y_{i,i+s} = \ln(\text{CPI}_{i+s}) - \ln(\text{CPI}_i) = [\ln(\text{CPI}_{i+s}) - \ln(\text{CPI}_{i+s-1})] + [\ln(\text{CPI}_{i+s-1}) - \ln(\text{CPI}_{i+s-2})] + \cdots + [\ln(\text{CPI}_{i+1}) - \ln(\text{CPI}_i)] = Y_{i+s,i+s-1} + Y_{i+s-1,i+s-2} + \ldots + Y_{i+1,i}$. Clearly, our model can be applied, as long as the resulting inflation series is positive. This is frequently the case in many countries. In fact, in their article, Guerrero and Elizondo (1997) used a series on Quarterly Inflation Rate in Mexico defined this way.

In the Bayesian framework the value of q in the loss function must be fixed. However, the results obtained in the three examples give some indication that a sensible choice in general might be to use $q = .556$; that is, use the geometric mean of the partial accumulation ratios to forecast the total accumulation.

While this article was in the process of being refereed, Chen and Fomby (1999) published an article on the same topic. However, their approach requires long series. In addition, even though they mentioned the Bayesian approach, it was only marginally. In fact, they mentioned the possibility of using a noninformative prior distribution for the period total, but at the same time they ignored the other parameters involved. Their analysis is, in any case, incomplete.

ACKNOWLEDGMENTS

Research for this article was partially carried out while Enrique de Alba was Visiting Professor in the Department of Statistics and Actuarial Science at the University of Waterloo, Ontario, Canada N2L3G1. Manuel Mendoza was supported by the Alberto Baillères Chair on Insurance and International Finance of ITAM. Both authors acknowledge support from Asociación Mexicana de Cultura, A.C., and the Sistema Nacional de Investigadores in Mexico.

[Received September 1997. Revised March 2000.]

REFERENCES

Abraham, B., and Ledolter, J. (1983), *Statistical Methods for Forecasting*, New York: Wiley.

Box, G. E. P., and Jenkins, G. M. (1970), *Time Series Analysis, Forecasting and Control of Discrete Time Series*, San Francisco: Holden-Day.

Box, G. E. P., and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley.

Chen, R., and Fomby, B. T. (1999), "Forecasting With Stable Seasonal Pattern Models With Application to Hawaiian Tourism Data," *Journal of Business & Economic Statistics*, 17, 497–504.

de Alba, E. (1988), "Pronóstico Bayesiano de Agregados en Procesos Estacinales Estables," *Revista de Estadística*, 2, 1–6.

de Alba, E., and Mendoza, M. (1996), "A Discrete Model for Bayesian Forecasting With Stable Seasonal Patterns," in *Advances in Econometrics Vol. 11: Bayesian Methods Applied to Time Series Data*, eds. T. B. Tomby and R. C. Hill, Stamford, CT: JAI Press, pp. 267–281.

de Alba, E., and Van Ryzin, J. R. (1980), "An Empirical Bayes Approach to Outliers," *Journal of Statistical Planning and Inference*, 4, 217–236.

Guerrero, V. M., and Elizondo, J. A. (1997), "Forecasting a Cumulative Variable Using its Partially Accumulated Data," *Management Science*, 43, 879–889.

Lenk, P. J. (1992), "Hierarchical Bayes Forecasts of Multinomial Dirichlet Data Applied to Coupon Redemptions," *Journal of Forecasting*, 603–619.

Minitab, Inc. (1994), *MINITAB*, State College, PA: Author.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics* (3rd ed.), New York. McGraw-Hill.

Oliver, R. M. (1987), "Bayesian Forecasting With Stable Seasonal Patterns," *Journal of Business & Economic Statistics*, 5, 77–86.