Bayesian Forecasting Methods for Short Time Series

by Enrique de Alba and Manuel Mendoza

Preview

This article by Enrique de Alba and Manuel Mendoza extends Foresight's coverage of approaches to forecasting seasonal data from short historical series (less than 2-3 years of data.) The authors describe and illustrate a Bayesian method for modeling seasonal data and show that it can outperform traditional time series methods for short time series.

Key Points

- **When you have seasonal data but too short a history to estimate a seasonal model, a Bayesian approach can be productive.**
- **The Bayesian approach relies on proportions of partial-year to total-year figures and assumes that these proportions are stable from year to year.**
- **We compare the Bayesian approach with that of a traditional model and find that the Bayesian approach is superior for short time series but inferior for lengthy time series.**

BACKGROUND

FORESIGHT, spring 2007 (pp.11-25) includes a series of papers on the subject "Modeling Seasonality in Short Time Series." These articles address the challenges of forecasting a time series known to be seasonal but having only a short history, i.e. few observations.

The first paper, by Hyndman and Kostenko, examines the minimum sample size required to forecast with seasonal models based on regression, exponential smoothing, and ARIMA. The authors mention that Bayesian forecasting methods may be another alternative but they do not pursue it.

In the second paper, Michael Leonard shows how additional sources of information may be used to determine the seasonal properties of a short time series and presents a selection tree that can be used to determine an appropriate modeling approach. He does not mention Bayesian methods.

Dan Williams follows with a shrinkage approach to measuring seasonality in small samples. However, he points out that these procedures require at least two to three years of data and sometimes much more.

Finally, Philip Hans Franses considers changing seasonality. He points out that the simplest models assume constant seasonality and indicates that, where seasonality is evolving, it is necessary to have three or more decades of monthly or quarterly data in order to adequately fit a model. He concludes that you need lots of data to be able to confidently select an appropriate seasonal model.

It comes as a surprise to us that Bayesian methods are barely mentioned in connection with this challenging problem. In his editorial preface to this section, Len Tashman states, "When the seasonal items to be forecast have short histories – certainly when there are less than 3 or 4 years of quarterly or monthly data - it is prudent to explore alternatives to the fitting of seasonal models directly to the data." On this basis, it is worth noticing that Bayesian analysis is not only a possible alternative but a promising one, according to the many papers published on this approach.

BAYESIAN FORECASTING

Bayesian Statistics is not just another inference technique. It is a *statistical theory* with its own methods and techniques derived from a unique strategy for the solution of any inference problem. In fact, this strategy (maximizing expected utility) arises as a consequence of adopting a set of axioms.

A model is most often recognized as Bayesian when a probability distribution is used to describe uncertainty regarding the unknown parameters and when Bayes Theorem is applied. Bayes theorem is used to update a *prior* distribution (probabilities specified prior to data analysis) into a *posterior* distribution (the probabilities following data analysis) by incorporating the information, called likelihoods, provided by the observed data. A full Bayesian analysis can lead to the optimal choice among a set of alternative inferences, taking into account all sources of uncertainty in the problem and the consequences of every possible selection.

For forecasting problems, Bayesian analysis generates point and interval forecasts by combining all the information and sources of uncertainty into a predictive distribution for the future values. It does so with a function that measures the loss to the forecaster that will result from a particular choice of forecasts.

BAYESIAN ANALYSIS FOR A SEASONAL SERIES

When forecasts are required for the year-end total of a short time series, the use of the yearly ratios of part-year totals to whole-year totals for previous years can play an important role. A series can be said to have a stable seasonal pattern when the expected proportion of events that occur in a given fraction of the year is constant over time. Under this assumption, the whole-year total for the variable in question can be forecast, given a few (say two) years of data and the corresponding part-year figures (up to a given month) of the current year.

To determine whether seasonality is stable, you can calculate for each year of history the cumulative proportion of the part-year total occurring through the current month. If seasonality is stable, then these yearly-to-date proportions should all be close to each other.

When a short series is being analyzed, it is important to make use of the simplest possible models. Specifically, the number of unknown parameters must be kept at a minimum.

Binomial, normal and log-normal distributions have been considered in the literature for this purpose.

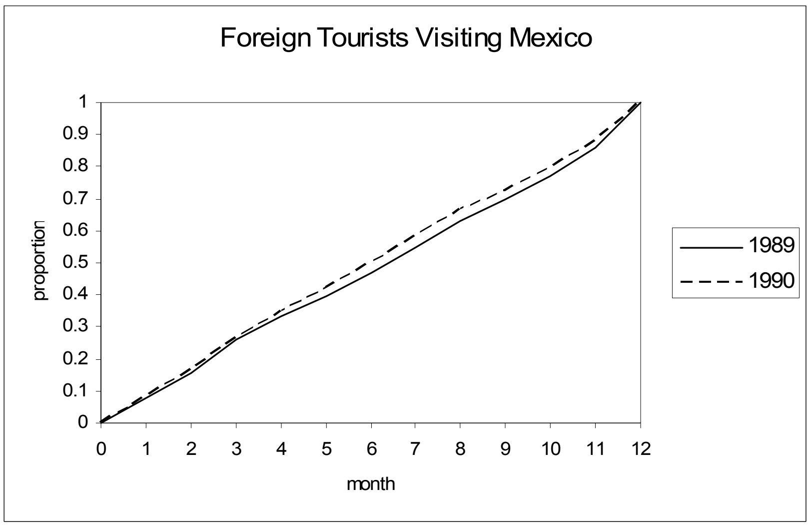
When a Bayesian analysis is conducted, inferences about the unknown parameters are derived from the posterior distribution. This is a probability model which describes the knowledge gained after observing a set of data. The procedure to obtain the posterior distribution is known as Bayes Theorem (see Custer and Miller's article in the Summer 2007 issue of Foresight for an elementary discussion of this theorem).

Bayes Theorem calculates the posterior distribution as proportional to the product of a prior distribution and the likelihood function. The prior distribution is a probability model describing the knowledge about the parameters before observing the currently available data. It can be elicited from past information or expert judgment. Alternatively, it can be chosen to represent a state of relative ignorance; in which case, the prior distribution is said to be neutral or noninformative and the resulting posterior distribution is mostly dependent on the observed data (see Migon and Gamerman, 1999, chapter 3 for an excellent discussion on prior distributions). Likelihood is simply the name given to the model applied to the data. For the problems discussed here, we can refer to Binomial, normal or log-normal models. For short time series, it is quite important to let the scarce data available speak for themselves. As a consequence, prior distributions for the parameters must be noninformative.

In our earlier paper (de Alba & Mendoza, 1996) these ideas were applied to an example based on the number of foreign tourists visiting Mexico. We used a binomial model and noniformative priors to forecast the whole-year total for 1991, given monthly data for 1989 and 1990. For more information on our procedure, see the technical note at the end of the paper.

Table 1 shows the monthly data for the first 2 years while Figure 1 shows the cumulative part-year proportions corresponding to the number of tourists accumulated up to each month. They are close to each other; hence we have empirical evidence of the existence of a stable seasonal pattern.

Figure 1



Given a particular month, say January 1991, we used data from 1989 and 1990 to estimate the proportion p of tourists that visit Mexico every January. *p* is the proportion of tourist volume in January to the total number of tourists of the year. The inverse of this proportion *(1/p)* is used as an expansion factor which, when applied to the count in January 1991, with this model, leads to the forecast for the whole-year total in 1991:

Forecast of 1991 total = January 1991 count / p

By this procedure it is possible to obtain a forecast for the year total in 1991 from every month in that year, as the information becomes available. These forecasts for 1991 appear in Table 1. The true observed total of tourists for 1991 was 6374 (in thousands).

Table 1: Observed and forecast tourist count (figures in thousands)

	1989 Observed	1990 Observed	1991 Total Forecast
Jan	505	522	6168
Feb	494	539	6197
Mar	625	632	6349
Apr	466	532	6359
May	417	464	6489
Jun	455	518	6566
Jul	485	528	6607
Aug	518	527	6605
Sep	424	374	6544
Oct	466	451	6534
Nov	550	530	6511
Dec	893	769	6347
TOT	6298	6386	6347

COMPARISON OF BAYESIAN AND STANDARD PROCEDURES

Here is a second example, one that we used (Mendoza & de Alba, 2006) to compare the Bayesian approach with a standard time series model (ARIMA). The data series Electricity measures the monthly average residential electricity usage in Iowa City from 1971–1978 (Abraham & Ledolter, 1983). We held out 1978 to serve as the test period and used only 2 years, 1976-77 to fit the models. Table 2 shows the forecasts from a Bayesian model for continuous variables which makes use of a lognormal distribution for the likelihood, with a noninformative prior distribution. It also shows the forecasts and from a traditional time series (ARIMA) model.

Table 2. Electricity Use in Iowa City (Kw/h): Comparison of Bayesian and ARIMA models using 2 Years of data

	Forecast of Year Total for 1978 using data up to Current Month	
	Bayesian	ARIMA
Jan	6081	6203
Feb	6025	6001
Mar	6120	5516
Apr	6092	5085
May	6056	4928
Jun	6001	5928
Jul	6008	7275
Aug	5966	6237
Sep	6116	6809
Oct	6178	6004
Nov	6147	6044
Dec	6118	6118
TOTAL	6118	6118
MSE	6183	431861
MAE	62.80	488.28
MAPE	1.03%	7.98%

Each monthly figure in Table 2 is a forecast for 1978 as a whole based on data up through that month. Three accuracy metrics are reported at the bottom of the table – Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Note that the MAPE is slightly above 1%, indicating that the 12 Bayesian forecasts of the 1978 total made from January through December differed from the actual 1978 figure by an average of about 1%.

We also generated forecasts for the 1978 total using an ARIMA model. Two years of data is generally not sufficient for ARIMA to work well, even though the 24 months exceeds the requisite minimum of 16 months (Hyndman & Kostenko, 2007) for being able to obtain a fit. The forecast accuracy metrics in Table 2 reveal that the ARIMA forecast errors were substantially larger than those from the Bayesian model. For example, the MAPE is nearly

8 times greater. We obtained similar results (not shown) using Winters' exponential smoothing method.

We repeated the comparison between the Bayesian and ARIMA models but using a longer time series; in this case 7 years of Electricity data (1971-1977) served as the historical data to forecast 1978. The results in Table 3 show that the traditional ARIMA model was superior for the lengthy time series, with a MAPE about one half that of the Bayesian model. With this lengthy series, ARIMA could adequately detect the underlying autocorrelations in the data.

Table 3. Electricity use in Iowa City (Kw/h): Comparison of Results Using 7 Years of data.

	Forecast of Year Total for 1978 using data up to Current Month	
	BAYESIAN	ARIMA
Jan	6242	6151
Feb	6263	6158
Mar	6283	6174
Apr	6201	6134
May	6099	6091
Jun	6070	6088
Jul	6091	6112
Aug	6001	6028
Sep	6084	6139
Oct	6136	6153
Nov	6116	6115
Dec	6118	6118
TOTAL	6118	6118
MSE	8105	1588
MAE	71.18	32.46
MAPE	1.16%	0.53%

Note, however, that the Bayesian model used here was chosen because it is a very simple one and so suitable for short series. More sophisticated Bayesian models could have been developed for the longer time series.

Our experience with other series bears out the results in Tables 2 and 3: traditional time series models such as ARIMA are better for long series while Bayesian forecasts outperform those in the case of the short series. So Bayesian procedures can be effective if only a small amount of past data is available. Our research also shows that the Bayesian forecasts perform well even in the presence of cycles and trends. A limitation of these models, however, is that they cannot be applied if there are negative values in the series.

CONCLUSIONS

When the time series is short, such as 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 (as through ARIMA models), even though the series is known to be seasonal. In this situation, Bayesian analyses can prove productive, provided we assume that the seasonality is stable. A simple, graphical test can be used to verify this assumption. So the Bayesian approach offers a useful alternative to deal with the problem of modeling seasonality in short time series.

REFERENCES

Abraham, B. & Ledolter, J. (1983). *Statistical Methods for Forecasting*, John Wiley & Sons, Inc., New York.

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

de Alba, E., & 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*, T. B. Fomby & R. C. Hill, (Eds.) Stamford, CT: JAI Press, 267-281.

Guerrero, V.M. & Elizondo, J.A. (1997). Forecasting a cumulative variable using its partially accumulated data, *Management Science,* 43, 879-889.

Mendoza, M. & de Alba, E. (2006). Forecasting an accumulated series based on partial accumulation II: A new Bayesian method for short series with seasonal patterns, *International Journal of Forecasting*, Issue 4, 781-798.

Migon, H. S. and Gamerman, D. (1999), *Statistical Inference: An Integrated Approach*, Hodder Arnold