The Bayesian Approach to Forecasting

An Oracle White Paper
Updated September 2006
The Bayesian Approach to Forecasting

INTRODUCTION

The Bayesian approach uses a combination of *a priori* and *post priori* knowledge to model time series data. That is, we know if we toss a coin we expect a probability of 0.5 for heads or for tails—this is *a priori* knowledge. Therefore, if we take a coin and toss it 10 times, we will expect five heads and five tails. But if the actual result is ten heads, we may lose confidence in our *a priori* knowledge. This may be explained by a change to the coin that was introduced to alter the probability—this is *post priori* knowledge. Another example of *post priori* knowledge is future price change or marketing promotion that is likely to alter the forecast.

The main principle of forecasting is to find the model that will produce the best forecasts, not the best fit to the historical data. The model that explains the historical data best may not be best predictive model for several reasons.

- The future may not be described by the same probability as the past. Perhaps neither the past nor the future is a sample from any probability distribution. The time series could be nothing more than a non-recurrent historical record.
- The model may involve too many parameters. Overfitted models could account for noise or other features in the data that are unlikely to extend into the future.
- The error involved in fitting a large number of parameters may be damaging to forecast accuracy, even when the model is correctly specified.

In any of these cases, the model may fit the historical data very well, yet still forecast poorly, illustrating that there is a vast difference between its internal and external validities.
A forecasting model that includes all parameters poorly predicts historical data. From the graph above, we can see how a regular model with all the parameters cannot correctly predict the historical data. We would like to select a model that minimizes the forecasting error and not the historical data error.

**Forecasting Models Using Classical Statistical Methods**

Classical, or orthodox, statistics selects just the “best” model and rejects all the others, even if they are only marginally worse than the best model. Unfortunately, this limitation is often compounded by the well-known problem of over-fitting, where a model is excessively fine-tuned in order to explain the past, usually to the detriment of its predictive power.

Bayesian analysis, in contrast, allows multiple data models of comparable high quality to be combined by assigning probabilities to each model. In addition to improving the accuracy and robustness of predictive abilities, this approach also adds considerable flexibility to the system.

Causal factor selection is complicated by the fact that the number of candidate factors is often comparable to the total length of the data. Classical statistical methods either fail to work, or reject most of the causal factors. This can be especially frustrating if the user already knows, either through experience or common sense, that the candidate factors put forward are relevant. In that case, what is actually needed from the system is for it simply to measure the effect of the causal factors, rather than attempt to test for their relevance.

Bayesian model averaging, a rapidly developing field of modern statistics, treats the problem in a very natural way. It tries out many small (overlapping) subsets of causal factors and determines that a causal factor should be considered relevant if, roughly speaking, it is found to participate in a sufficient number of the subsets—the bigger the number the higher the relevance. Consequently, the length of available history does not limit the number of causal factors that the user can put forward. Of course the user should avoid introducing causal factors that are, a
priori, totally irrelevant to the series in question as this simply slows down computation time and can sometimes affect forecast accuracy. But if the factor seems to be of some relevance, then it should be included and the data allowed to “speak for itself.” It is also worth mentioning that the same causal factor might be used for modeling at different levels, such as city and region; these would be estimated separately and recombined to improve forecast accuracy at all levels.

**Forecasting Models Using the Bayesian Approach**

The Bayesian approach combines the results of individual models. Each model is evaluated, and each model in turn tests a number of subsets of system and user-supplied causal factors (price is almost invariably a causal factor). All combinations of models and subsets of causal factors are assigned weights indicating their relevance. Every combination contributes to the final forecast according to its weighting. The reconciliation procedure ensures that the results meet the necessary constraint of parent-children relationship.

The following figure depicts various approaches to forecasting. While most forecasting approaches rely on one of several methods, the Bayesian approach uses a statistically sound way of combining the advantages of each approach to generate the best forecast accuracy.

**Forecasting Systems Engine Tree**

- **Subjective** (Spreadsheet, Homegrown)
  - Statistical
    - Extrapolative
      - Single Method (Logistic, Mann-Whitney, Most ERP Modules)
      - Best Fit (Forecast, E-Step)
    - Multiple Methods
      - Expert System (FUTURCAST)
  - Causal (SAS)
- **Objective** (Data driven)
  - Non-Statistical (Demand Solutions, Focus Forecasting®)

Several forecasting approaches and the methods they use to generate forecasts.

The Bayesian technique uses a methodology that can be described by the following equation: \( F = w_1f_1 + w_2f_2 + \cdots + w_nf_n \). Where \( F \) is the final forecast; \( f_1 \) refers to the forecast using model 1; \( f_2 \) refers to the forecast using model 2; \( f_n \) refers to the forecast using model \( n \) and \( w_j \) is a weight given to model \( j \).

The value assigned for weight takes into account the residuals, or the difference between the true data and estimated data. When determining the weight value, a
penalty factor is taken into consideration—the more parameters (causal factors) that need to be estimated, the bigger the penalty. If the number of causal factors is larger than, or about half of, the number of observations, then the estimation in general will be poor. Therefore the weight given to such a model will be small.

The Demantra Demand Planner forecast engine automatically combines different forecast models in the same time series.

**Case Study: Bayesian Versus “Best-Fit” Approaches**

The following graph demonstrates that the Bayesian approach provides a forecast with accuracy as good as, or better than, the “best-fit” approaches common in the market.

The customer data is from a global company with over US$9 billion in annual revenues. The customer provided two years of historical data. The objective was to use only the first 18 months of data to “forecast” the last six months sales by unit, and to compare the forecasted values to the actual values. Alternative forecasting scenarios were developed using models picked as best by competing vendors’ packages, as well as the Bayesian approach.
Comparison of forecasting error values obtained with the Bayesian approach versus several other forecasting methodologies.

The X-axis represents item IDs and the Y-axis represents relative error (calculated as Mean Absolute Percent error). For items 6740, 6759, 6753, 6872, 6873, and 6877, the differences between the Bayesian and the “pick-best” approaches used by other vendors are relatively small (maximum absolute difference in MAPE is 4 percent maximum; relative difference in MAPE is approximately 30 percent). The difference between the approaches is noticeable for other item IDs; it was as large as 25 percent on an absolute and more than 100 percent on a relative basis.

CONCLUSION

Because the behavior of items can differ from item to item, and in some cases from location to location, using one pick-best model to generate forecasts is not recommended. The Bayesian forecasting approach relies on an optimal combination model for each item and location combination. The engine adapts the model weights as new data becomes available so that the users do not have to track product behavior and respecify the model usage manually, as some other commercial packages require the user to do.

The patented Bayesian analytical forecast engine used in Oracle Demantra planning solutions offers the most accurate forecasts possible. Automated algorithms consider 15 industry-standard and proprietary forecasting models, each geared to different demand patterns. The forecast engine automatically combines different forecast models in the same time series. This produces a forecast that accommodates seasonality, promotions, trends, and many other causal factors.