Speculations About Seasonal Factors

J. Scott Armstrong, The Wharton School U. of Pennsylvania
Fred Collopy, The Weatherhead School, Case Western Reserve U
Study done in 1993, published at forecastingprincples.com in 1998, updated May 2000

To address the issue of uncertainty in timing, we suggest a spillover model. It would average the seasonal factor with seasonal factors that precede and follow it:

$$S_j^{?} = K * S_j + (1-K)*[(S_{j-1} + S_{j+1})/2].$$

Set $K = 1 - 2/d^2$ initially where d is the number of observations available for estimating the seasonal factors (for example, the number of Aprils.). To do this, you must have at least 2 observations.

Lack of information can be addressed by using a shrinkage and horizon model. This can be defined as follows for multiplicative factors:

$$\underline{S}_{j,h} = M_h + (1-M_h)\ S_{j,h}$$

where j represents the time period (e.g., month), and M_h is the modifier for year h in the forecast horizon (the same modifier applying to all periods in the year). M_h is bounded between 0 and 1. When it is 0, no modifying occurs; that is, the full seasonal factors are used. When it is 1, no seasonal factors are used. The subscript ‘h’ allows you to change the modifier over the years of the forecast horizon. The modifier would typically increase over the forecast horizon. In other words, because of increasing uncertainty, less emphasis should be given to the historically estimated seasonal factors as the forecast horizon increases.

Ideally, the guidelines for modification should be specified so that the procedure can be applied automatically. However, subjective inputs can also be examined to determine whether the analyst can contribute to the estimation of the modifier. The following model is suggested for selecting the modifier, M:

$$M_h = (d+d_s)^{-k} * h^l$$

where d represents the number of cycles (years) of data in the calibration sample, d_s represents the subjective data about the causal factors underlying seasonality, k is the parameter to adjust for errors in the historical estimation, h is the number of periods ahead in the forecast horizon, and l is a parameter to adjust for the deterioration of the factors over the forecast horizon. If the calculated modifier, M_h exceeds 1.0, then use 1.0.

Domain knowledge about causality, represented by d_s, is expressed as a number of years. Thus, when d_s equals 2, this means that the expert believe subjective information on seasonality to be worth about two years of actual data.

The more confidence that the forecaster has in the subjective knowledge, the higher the value for d_s. Use $d_s = 0$ when there is little knowledge about casual factors that would produce seasonality. When excellent information exists on causality, use $d_s = 3$. The subjective adjustment is important only for situations in which the data are limited.

Search routines can be used to obtain good estimates for the parameters k and l. Subjective estimates may also help determine the best value for k. A higher k should be used if the experts in that situation expect substantial changes in the seasonal pattern over the forecast horizon. These changes may be due to increasing uncertainty or to planned changes that affect the seasonal pattern.

Some examples may help to demonstrate the properties of this model. Assume k = 0.5 for the estimation parameter and 1 = 0.1 for the horizon parameter. Consider then the following cases:

Case 1: *Few data (one year) and no basis for assuming that casual factors exist ($d_s = 0$):* In this situation, seasonal factors should not be used. That is, M_h should be equal to one (fully modified, meaning no seasonal adjustment) for all horizons. This is what the model provides.

Case 2: *Few data (one year) and good subjective information that seasonality exists ($d_s = 2$):* Here, the modifier would be .577 for the first year in the forecast horizon. This means that 57.7% of the seasonal variation has been removed from the forecast. The modifier would increase in the future (e.g., in the year 10 it would be .726.)

Case 3: *Much data (ten years):* Given good subjective information ($d_s = 2$), M_h would be .289 for the first year to be forecasted. For year 10, $M_h = .36$.

You can expect seasonal factors to be most effective when the data have been cleaned up to remove errors, and when obvious causes of variation have been accounted for. Prior to calculating seasonal factors, data should be adjusted. To the extent that you cannot control for these relevant factors, you should have less confidence in the historically calculated seasonal factors.

We selected 62 monthly series from the data used in the M-competition (Makridakis, et al. 1982), those whose identification numbers ended in 6. We made forecasts for one-ahead to six-ahead forecasts. We started with x observations and used successive updating for each series.

We used the median absolute percentage error (MdAPE) (Armstrong and Collopy 1992). For one-ahead forecasts, the MdAPE was reduced by 7.2%. For 18-month-ahead forecasts, the MdAPE was reduced by 5.0%. The estimation modifier led to improved accuracy for 66% of the series. This is significant at $p < .01$. The horizon modifier yielded improved accuracy on 56% of the series, which was not significant ($p = .22$).

References

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen & R. Winkler (1982), "The accuracy of extrapolation (time-series) methods: Results of a forecasting competition," *Journal of Forecasting*, 1, 111-153.

Armstrong, J. S. & F. Collopy (1992), "Error measures for generalizing about forecasting methods: Empirical comparisons," (with discussion), *International Journal of Forecasting*, 8, 69-111. Full-text at forecasting principles.com.