

Forecasting at an Hourly level

Posted on Oct 19, Posted by [Tom Reilly](#) Category [Forecasting](#)

Moving away from Allocation of Historical Contributions to Multiple Regression Models

Allocation vs. Modeling

Today's approaches to forecasting for example, inbound call center volume, rely upon calculating "contributions" from historical data and then allocating those to create a forecast using Excel or some similarly with programming code in SAS or R. For example, you might take the last year of data and calculate the total and find out that January "contributed" 1% to the total year and then use that to help forecast next January. The guiding assumption is that the distributions are constant and don't change. For the day of the week forecast, using the same type of approach looking at how the 7 days of the week percentages stack up over the last year. If we wanted to forecast 16 semi-hours of the day, we would identify and allocate percentages. We could identify an average impact of holidays on and around and also an average impact of long weekends that are used inappropriately for all holidays. These percentage based allocations don't account for the nuances in the data that need to be addressed in order to have a high quality forecast. These approaches could be enhanced using statistical modeling techniques called as multiple regression that will identify complexities in the data. Multiple regression will attempt to do similar things like we just discussed, but allow for many more complicated issues that occur. For example, outliers occur in data and if ignored will skew the forecast and can be identified and adjusted for using schemes to search and detect. See Ruey Tsay's Work on [Outlier](#) detection for more on this. Additionally, if there are changes in trends(up or down) in the data that need to be factored with a trend variable and yes there could be more than one trend! There could be changes in the mean that are not a trend, but more of a one-time bump (ie level shift) and yes again there could be more than one level shift! Your marketing promotions might be the reason for the level shift, but if you don't actually model that promotion statistically you aren't going to be able to leverage it into the forecast like you can do with regression modeling. There could be changes over time where Mondays were always low and now become high (ie seasonal pulse). You will have single one time "pulse" outliers that need to be adjusted for as random swings occur without cause. Now, you might be aware of some of these outliers like system outages and you should include them

as a dummy causal variable to explain the variability caused by it. Economists call this knowledge “apriori” when building these models. If you don’t have a forecasting approach that identifies these changes then your forecast will ignore all of these impacts and you will be left with a poor man’s modeling approach or what accountants like to call “allocation”.

Some try and model the call volume data and other data on a weekly level. Weekly level data is severely impacted by moving holidays like Easter, Thanksgiving and Ramadan. In addition, using weekly level data means that you are allocating and not modeling your data. You should also know that all of the items listed in the previous paragraph can’t be factored into the forecast using weekly data.

We know:

- Different days in a week have statistically different call patterns
- Different semi-hours in a day have statistically different call patterns based on which day of the week it is

How do we solve this problem? Mixed Frequency Modeling

To solve all of the above, this approach can be best described as “Mixed Frequency”. It is called Mixed Frequency because the seasonality of the two datasets that will be analyzed are different but used together. The daily level data is a seasonality of 7. The semi-hourly data is a seasonality of 48.

STEP 1 :To solve this puzzle, we build a regression model at the daily level using promotional variables(lead and lag impacts) and the future expected promotions, 16 U.S. holidays(lead and lag impacts), days of the week with 6 dummy variables, search for special days of the month,

months of the year, impact on Mondays after holiday, impact on Fridays before a holiday, along with checking for changes in trend, level shifts, outliers and seasonal pulses to get a forecast. We recommend 3 years of historical data so that you can get a good measure on the holidays. For example, in 2010 and 2011 Christmas landed on a weekend so you have no measure of the impact of Christmas on a weekday. You could include a causal that explains the macro direction of the business (ie # of outlets your product is sold in) always be used as a good causal to help explain the overall trend in the call volume. You could provide future sales estimates using a forecast to help guide the direction of the forecast as well as allow the managements input on expected future business success for upcoming months.

STEP 2: If you wanted to forecast the 9 am to 5 pm time frame by half hour periods, you would use the history and forecast from or the daily level data and use it as causal. You would use that causal in each of 16 separate regressions for each of the half hour data sets. The daily causal and the day of the week causal would be forced into the model so as to provide a “guiding hand” for each of the semi-hours. This framework allows the day of the week and intraday nuances to be modeled and forecasted.

The Forecasts from the first step building the “daily level” model and forecast would need to go through a “Forecast Reconciliation” step so that the sum of the 16 semi-hourly forecasts would match the daily level. The opposite could be done if requested.

Try challenging your vendor or your forecaster and ask them how they actually calculate the forecast. Are they allocating or modeling?

Tags: Tagged in: [fixed effects daily effects](#) [forecasting regression time series](#)
[holidays](#)