



Forecasting high frequency data with complex multi-layer seasonality using covariates

Shubhabrata Das*

Indian Institute of Management Bangalore, Bangalore, India – shubho@iimb.ernet.in

Increasingly in many contexts, e.g. in energy load or demand, data is recorded at every minute or every 5 minutes. Such data exhibits not only annual and weekly seasonality but also within day and within hour seasonality which are often entangled with each other. Various weather covariates also have significant, at times nonlinear (quadratic) impact on the dependent variable. Such forecasting problems face methodological challenge, as ARIMA type models are not directly suitable for data with large seasonal periods as well as multiple-level seasonality, while relatively newer approaches like TBATS has difficulties in accommodating covariate. On top of that, there are computational challenges to implement advanced models on such high frequency data. In this work, we first adopt traditional time series and regression based models for daily average data. Subsequently the within hour and within day seasonality is analyzed thoroughly and the estimated seasonal components are integrated with projected daily average the final forecast. The methodology is illustrated on energy load data from New York. Results show differential Sunday as well Saturday effect while difference across the different weekdays is not significant. Hourly and within hour seasonality varies in more complicated way and incorporating their estimates improves the forecast substantially.

Keywords: time series; regression; decomposition; big data analytics.