Methods comparison for forecasting time series with multiple seasonal patterns

by

M.S., Wenxiao Zhou

University of Connecticut, 2020

Final Project

STAT5825 – Applied Time Series

Advisor: Nalini Ravishanker

Fall 2020

Abstract

Many time series have complex seasonal patterns such as multiple seasonal effects in a long term, weekly series with a non-integer period, or series with dual-calendar seasonal effects. However, most existing models are designed to accommodate simple seasonal patterns, some improvements should be included to fit those patterns. In this paper, four methods are included for multiple seasonal patterns: Seasonal and Trend decomposition using Loess, Dynamic harmonic regression model, BATS model, and TBATS model. Four methods focus on different problems: MSTL allows the dynamic seasonal components and is robust to the outliers; Dynamic harmonics regression allows autocorrelated errors with a harmonic approach to fit multiple seasonal patterns; BATS model get better performance than the simple state-space model; TBATS model allows a slowly change on seasonal patterns, a trigonometric method make progress on BATS model for non-integer seasonal effects. Finally, the forecast accuracies among the four models are given by RMSE, we have the TBATS model with the best post-sample forecast performance that RMSE=276.551 among all the models.

Section 1. Introduction

In reality, many data are collected in time order with useful time series plots to see how a given economic variable, a nature index, or a signal change over time. For long-period collected time series, complex seasonal patterns are inevitable: Some weekly series have patterns with non-integer period, some hourly collected series has multiple seasonal patterns like daily seasonal effect as well as weekly effect, while other series may have dual-calendar patterns due to special cultures and social customs like Chinese use both lunar and solar calendars for daily work and special festivals.

Here, we are facing a problem of forecasting the hourly energy consumptions collected from PJM Interconnection LLC basing on the overly 10 years data. PJM is a regional transmission organization (RTO) that coordinates the movement of wholesale electricity in all or parts of 13 states and the District of Columbia. Our data select the hourly energy consumption from Duquesne Light Company, which generates electricity primarily in the area around the city of Pittsburgh, Pennsylvania. The data is obtained from the Kaggle website and is named as “Grad_Wenxiao-Zhou-Data.csv”.[1]

Most methods of learning so far such as seasonal structural decomposition, Holt-Winters’ additive and multiplicative as well as ARIMA models aim at dealing with simple seasonal patterns with a small integer-valued period (such as 12 for monthly or 4 for quarterly data), which are far beyond our needs for complex problems. Luckily, R.B. Cleveland et al. (1990)[2] come up with the STL decomposition method for complex seasonal effects as well as seasonal component changing over time; Harvey et al. (1997)[3] comes up with using the trigonometric approach for single seasonal time series.

Our data is hourly collected, which usually has at least two types of seasonality: a daily pattern, a weekly pattern. To keep data on a successive timeline, we select data that contains complete power consumption hourly through 2005Jan1st. to 2018Aug2nd. Time series plot for 13 years is shown in Figure 1.1:

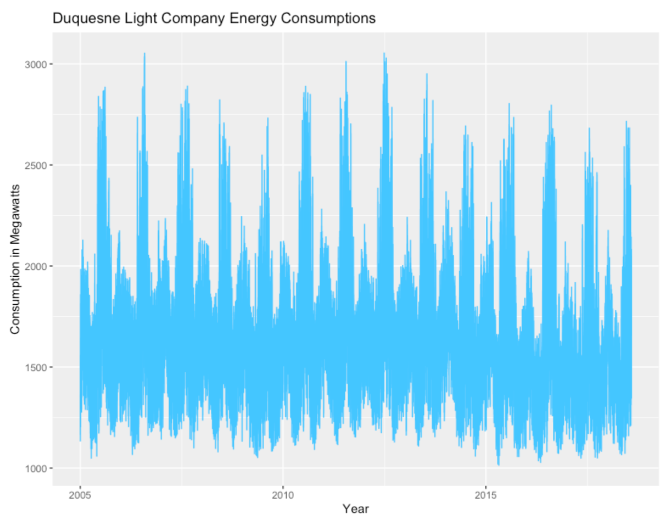

Figure 1.1: Duquesne Light Company Energy consumption for 2005-2018

To detect a seasonal pattern clearly, we consider the monthly, weekly, and hourly effects orderly when randomly take the year 2010 as an example:

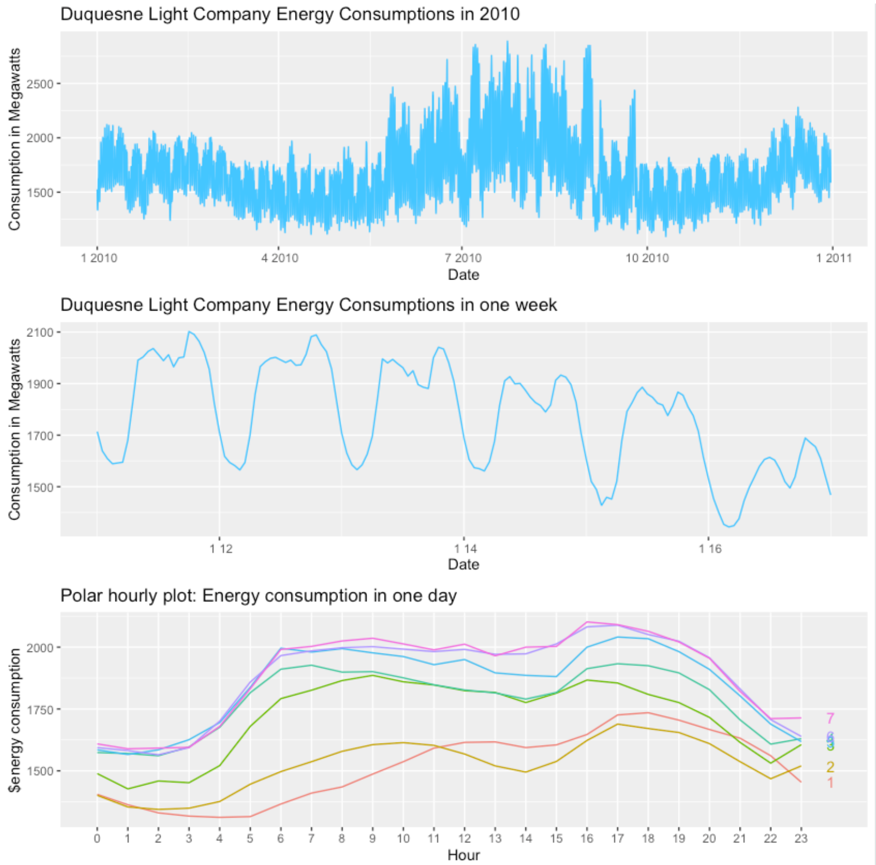

Figure 1.2: Seasonal effects (monthly, weekly, hourly) detection

From the three plots above, first of all, plot 1 records the monthly changes of consumption in 2010: there is an increasing trend of energy consumption from May to August and a similar trend from November to January in the next year. We can consider it as people using electricity for cooling and heating respectively that cause these two trends. Then, plot 2 records the daily changes in consumption in one week in 2010: a daily week from Monday to Sunday is chosen showing that energy consumptions during weekdays are higher than the weekends. If we focus on the hourly effect, plot 3 records the hourly changes of consumption in one day in 2010, we randomly select 7 successive days that recording one week: energy consumptions rise to the peaks at 9 am and 4 pm respectively every day, these results are consistent with people's daily work and study periods.

To predict the future energy consumptions basing on historical time series, it's important to find adequate models dealing with multiple seasonal components.

Section 2. Goal of the Analysis

According to the seasonal detection in Figure 1.2, our energy consumption data may contain hourly, weekly, yearly seasonal effects. The first goal of this paper is to list some traditional methods that can solve with multiple seasonal effects. The basic theories and essential data analysis will be presented in the following sections. Then, the application basing on the Rob J Hyndman et al. (2010)'s innovations state-space modeling framework for forecasting our energy consumption data will also be provided.[4] What's more, we will compare the out-of-sample performance measured by the Root Mean Square Error (RMSE) for all the models we use in detail. A summary of the advantages and shortcomings of each model is included in the results section.

Section 3. Comprehensive Data Analysis

To begin with, we restrict the data from Year 2012 to Year 2018. The data is separated into two parts: data from 2012-08-01 to 2018-06-31 is training group, data from 2018-07-01 to 2018-07-31 is test group. The training group is used for modeling fitting, while test group is used for validation.

3.1 MSTL decomposition

MSTL is an acronym for "Seasonal and Trend decomposition using Loess (STL) with multiple seasonal periods", while Loess is a method for estimating nonlinear relationships. Suppose the data can be decomposed as $X_t = T_t + S_t + R_t$, where T_t, S_t, R_t are respectively denoting trend, seasonal and random component for the time series. Here, S_t contains more than one seasonal effect. STL is a filtering procedure that can handle any type of seasonality, the seasonal component is allowed to change over time along the rate of change can be controlled by the user. The smoothness of a trend-cycle can also be controlled by the user (R. Cleveland et al. 1990) [5].

From figure 3.1 (in Appendix), if we take a subset of the time series and decomposed it into three parts mentioned above with one seasonal pattern frequency=24, there is a rough increase for trend along with two times decrease. This pattern in trend may be caused by some uncaptured extra seasonality from the higher natural period, i.e. it can be considered as multiple nested seasonal data. The last panel shows the error component is influenced by the multi-seasonal effects with a certain pattern.

In that way, we consider time series with three main seasonal components: daily, weekly, and yearly. The MSTL model is used for time series decomposition, the decomposed model is shown below:

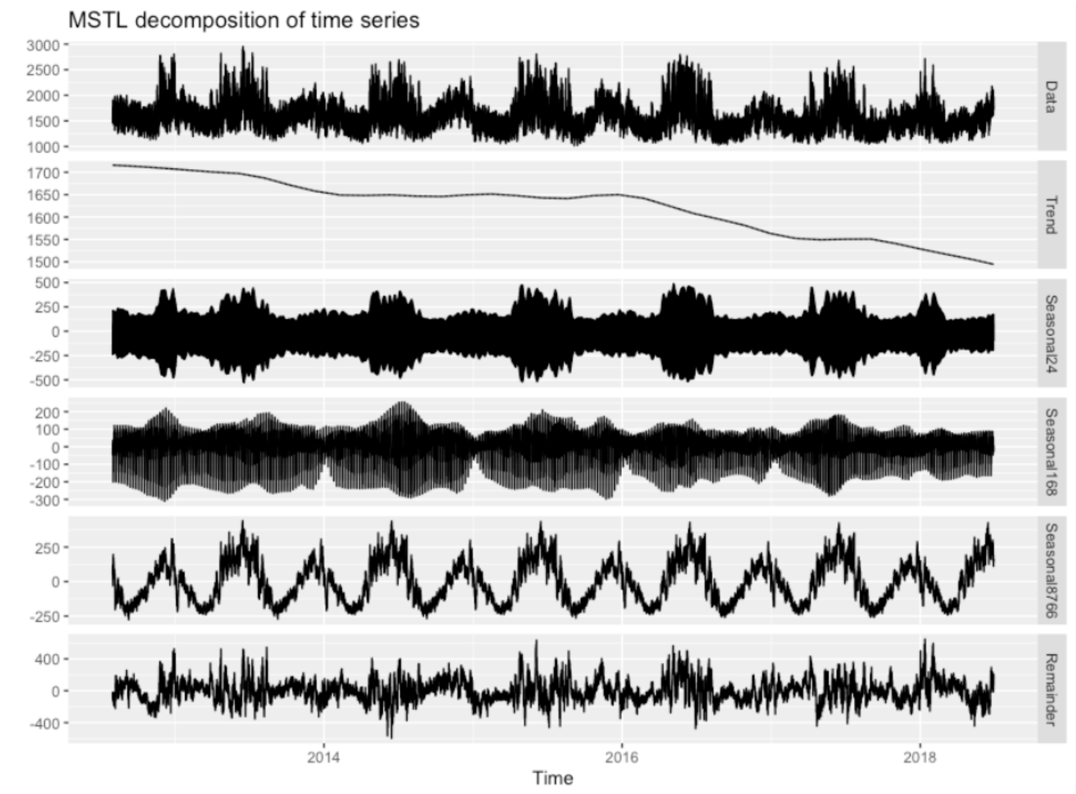

Figure 3.2 MSTL decomposition of training data

Now, a clearly decreased trend after the MSTL decomposition as well as hourly, weekly, and yearly seasonality for our data are seen in the plot.

3.2 Dynamic harmonic regression

Rob J Hyndman[6] mentions that in practice, it is usually run out of memory whenever the seasonal period exceeds 200. For our data, the yearly seasonal effect with a period

Also, seasonal differencing of high order does not make sense. Thus, using harmonic regression where the multiple seasonal patterns are modeled by Fourier terms with short-term time-series dynamics handled by an ARMA error.

For original dynamic regression, we assume the model as $X_t = \beta_0 + \beta_1 z_{1,t} + \cdots + \beta_r z_{r,t} + \varepsilon_t, where\ \varepsilon_t \sim ARIMA(p,d,q)$. However, for dynamic harmonic regression, we have Fourier terms for seasonal periods:

$$s_k(t) = \sin\left(\frac{2\pi kt}{m}\right), c_k(t) = \cos\left(\frac{2\pi kt}{m}\right)$$

$$X_t = \beta_0 + \sum_{k=1}^{K}[a_k s_k(t) + b_k c_k(t)] + \varepsilon_t, where\ \varepsilon_t \sim ARMA(p,q)$$

Here, assumes the seasonal pattern is unchanging.

$m\ represents\ seasonal\ period, m = 24{,}24 \times 7{,}24 \times 365.25\ respectively;$

$K\ is\ number\ of\ Fourier\ terms\ of\ each\ m;\ a_k\ and\ b_k\ are\ regression\ coefficients;$

The first step for this model is detecting the numbers of Fourier terms K. Notice that as K increases the Fourier terms capture a more "wiggly" seasonal pattern. However, based on the Law of parsimony and to avoid the overfitting problem, we select an adequate K basing on the local minimal AICc value. Also, to fit the dynamic model to find the significant harmonics, we finally determine the Fourier numbers for three seasonal components are K=(2,6,6).

First differenced residuals after model fitting are obtained with ACF and PACF plots below. A cut off after lag 2 on PACF plot, ACF with an exponential decay can be approximately determined.

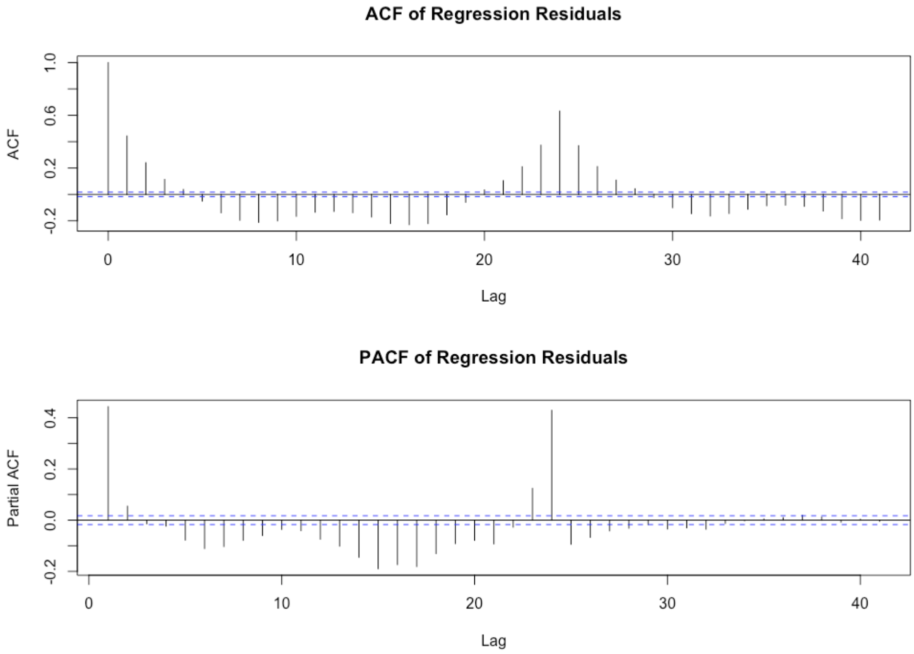

Figure 3.3 ACF and PACF of Diff(Residuals)

After using auto.arima() function in R, the fitted dynamic harmonic regression model is as below, the R output shown in Appendix:

$$X_t = -194.4\sin\left(\frac{2\pi t}{24}\right) - 41.7\cos\left(\frac{2\pi t}{24}\right) - 38.6sin\left(\frac{4\pi t}{24}\right) + 27.6cos\left(\frac{4\pi t}{24}\right)$$
$$-31.4sin\left(\frac{2\pi t}{168}\right) - 54.3cos\left(\frac{2\pi t}{168}\right) - 22.0sin\left(\frac{4\pi t}{168}\right) - 3.4cos\left(\frac{4\pi t}{168}\right)$$
$$-1.9sin\left(\frac{6\pi t}{168}\right) + 3.7cos\left(\frac{6\pi t}{168}\right) - 6.0sin\left(\frac{8\pi t}{168}\right) - 2.3cos\left(\frac{8\pi t}{168}\right) - 15.4sin\left(\frac{10\pi t}{168}\right)$$
$$+1.4cos\left(\frac{10\pi t}{168}\right) - 0.8sin\left(\frac{12\pi t}{168}\right) + 35.5cos\left(\frac{12\pi t}{168}\right) + 7.2sin\left(\frac{2\pi t}{8766}\right)$$
$$-28.7cos\left(\frac{2\pi t}{8766}\right) - 67.8sin\left(\frac{4\pi t}{8766}\right) - 134.7cos\left(\frac{4\pi t}{8766}\right) - 7.7sin\left(\frac{6\pi t}{8766}\right)$$
$$-27.5cos\left(\frac{6\pi t}{8766}\right) + 17.3sin\left(\frac{8\pi t}{8766}\right) - 43.8cos\left(\frac{8\pi t}{8766}\right) - 13.3sin\left(\frac{10\pi t}{8766}\right)$$
$$-4.0cos\left(\frac{10\pi t}{8766}\right) - 16.2sin\left(\frac{12\pi t}{8766}\right) - 12.3cos\left(\frac{12\pi t}{8766}\right)$$
$$where\ \widehat{\varepsilon_t} \sim ARIMA(2,1,0)$$

However, after checking the residuals of the model, we have an autocorrelated series of errors, with rejection of Ljung-Box test. This model is not adequate for data.

3.3 TBATS model

Some traditional exponential smoothing methods improved for complex seasonal effects, however, with some shortcomings: Holt-Winters method has a large number of parameters and seeds to be estimated when the number of seasonal components is large; most non-linear seasonal versions of the state space models can be unstable with infinite forecast variances in a certain horizon (Akram et al. 2009)[7]; exponential smoothing methods require the error is serially uncorrelated. The linear innovations state space model helps to solve the problems above.

Anderson & Moore (1979)[8] mentioned TBATS and BATS models are two special cases of the linear innovations state space model adapting to incorporate Box-Cox transformation to handle nonlinearities. Two methods are different mainly on the way that models the seasonal effects: BATS uses a traditional method allowing multiple integer seasonal periods, while TBATS uses trigonometric representation based on Fourier series to model seasonality. Generally, BATS requires m_i seed states for season i. In our example, there are three seasonal patterns with periods $24, 24 \times 7, 24 \times 365.25$, which leads a huge estimation in this model. In practice, it is time-consuming to get the model. Nevertheless, we have less parameters to estimate in TBATS model when most seasonal components can with small numbers of harmonics.

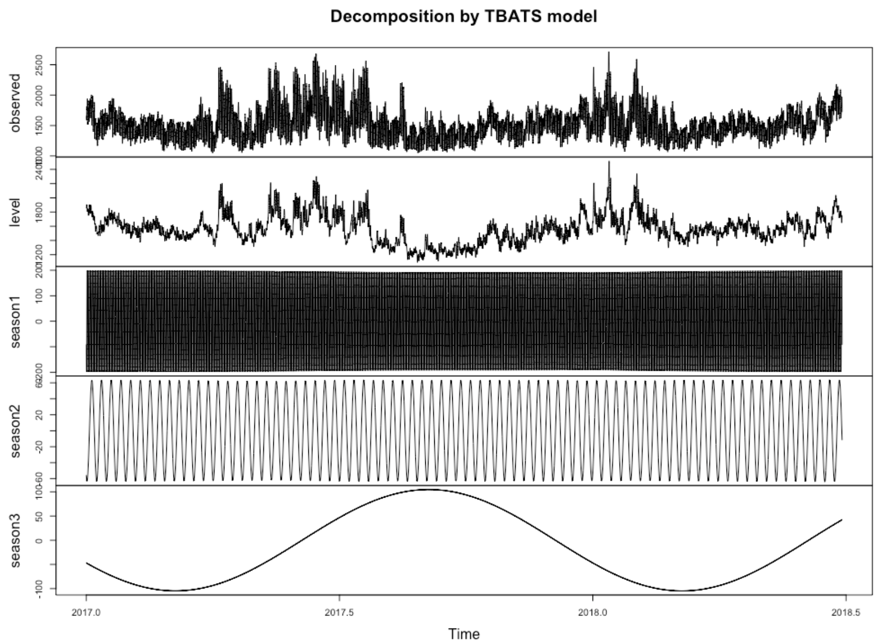

Figure 3.4 TBATS decomposition of training data

In figure 3.4, we show a decomposition of training data obtained by the fitted TBATS model. There are three clearly seasonal patterns obtained from this decomposition. The weekly seasonal pattern and yearly seasonal pattern are relatively stable. The hourly seasonal pattern is hard to detect since the time series is too long. As is seen from the time series plot itself, the trend component is larger in magnitude compared to the seasonal components.

Here, we only provide the fitted TBATS model for Duquesne energy consumption from Year 2017Jan. to 2018Jun. A subset period for prediction performance between BATS and TBATS models will be provided in the forecast accuracy section.

The estimated fitted $TBATS\{\omega, \phi, (p, q), \{m_1, k_1\}, \{m_2, k_2\}, \{m_3, k_3\}\}$ model with parameters is show below:

Table 1: Parameters chosen for application of the TBATS model

Parameters	ω	ϕ	α	β	γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}	γ_{23}
Value		1	0.140		-2.529×10^{-5}	-5.173×10^{-5}	-9.344×10^{-6}	-5.251×10^{-5}	1.203×10^{-4}	-4.221×10^{-5}

Table 1 Continued

Parameters	θ_1	θ_2	θ_3	θ_4	φ_1	φ_2	φ_3
Value	-0.003	-0.489	-0.176	-0.065	1.222	0.181	-0.509

The estimated value of $\phi = 1$ implies a purely deterministic growth rate without damping effect. The model implies that the irregular component of the series is correlated and can be described by an $\mathrm{ARMA}(3,4)$ process, and that a strong transformation is not necessary to handle nonlinearities in the series.

3.4 Forecast accuracy comparison

From the above three sections, we discuss four methods dealing with multiple seasonal patterns, which are MSTL model, Dynamic harmonic regression model, BATS model and TBATS model. Basing on the reference of the paper[4], the out-of-sample performance was measured by the Root Mean Square Error (RMSE), defined as $RMSE_h = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \widehat{x_\iota})}$, $where\ N$ is the length of the test sample, $x_i\ and\ \widehat{x_\iota}$ represent the true value and predicted value in the training group. Here, we have the test sample with hourly energy consumption from Year 2018July 1st. to 2018July 31th.

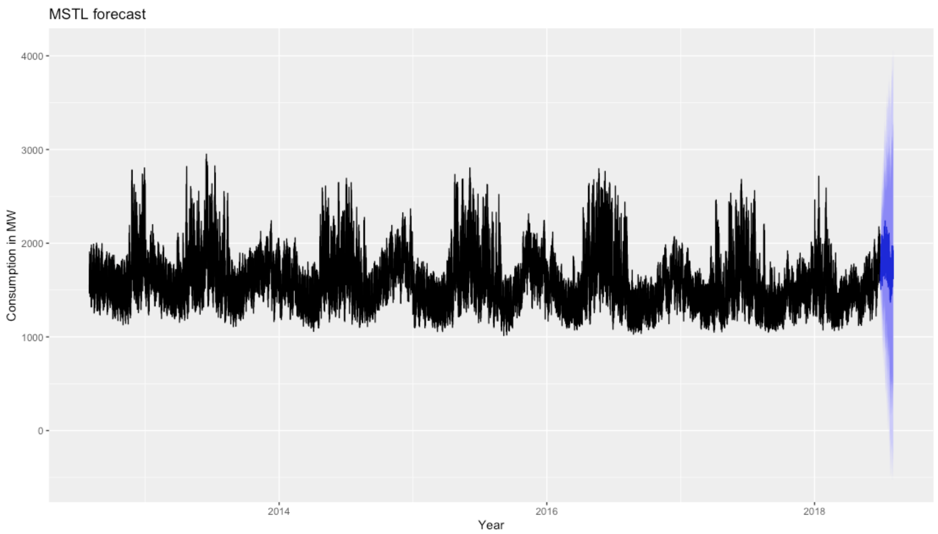

Figure 3.5 MSTL forecast for Duquesne Light Company Energy consumption

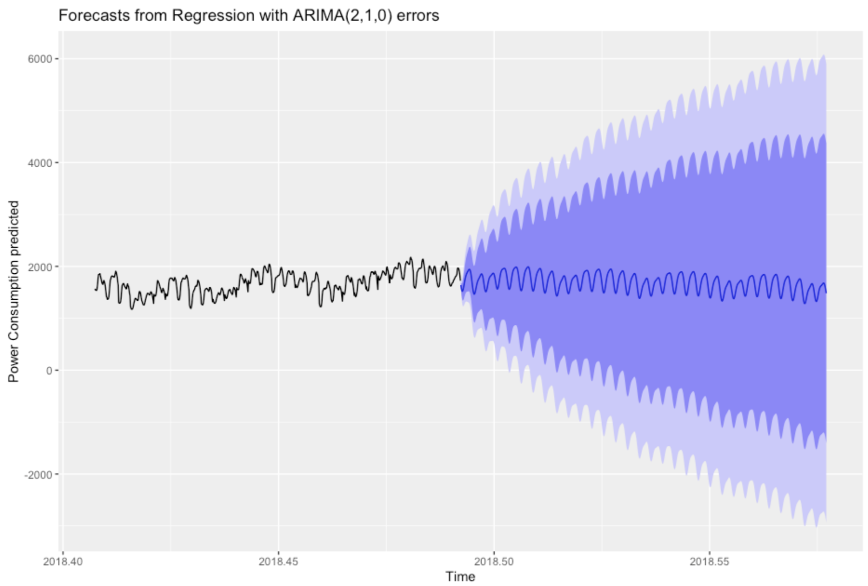

Figure 3.6 Dynamic harmonic regression forecast for Duquesne Light Company Energy consumption

Figure 3.5 shows the MSTL forecast with the RMSE=358.026 for the test-samples. Figure 3.6 shows the Dynamic harmonic regression forecast with the RMSE=339.952

for the test samples. Comparing the two methods, we have Dynamic harmonic regression model that performs better with lower RMSE. Dynamic harmonic regression model is superior to its' smoothness of the seasonal pattern can be controlled by Fourier terms. However, there is one disadvantage for this model: the seasonality is assumed to be fixed, which may be unreasonable for long time series.

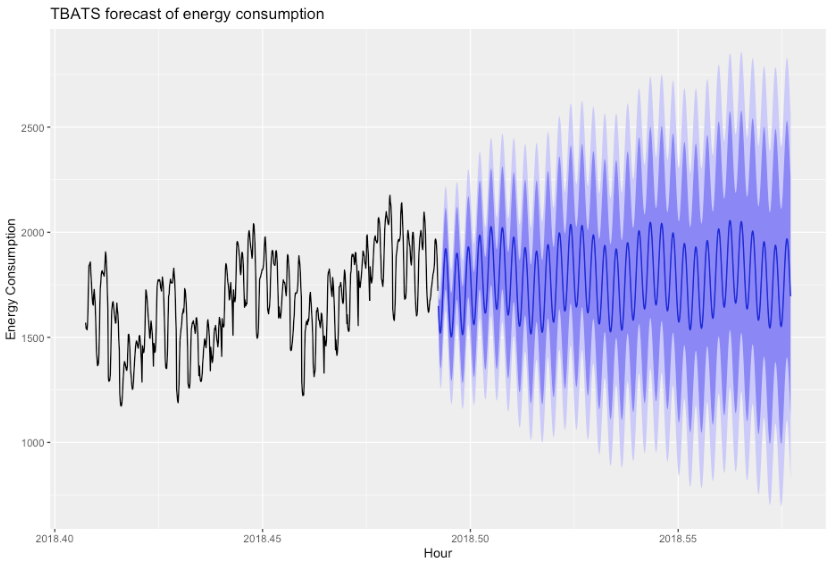

Figure 3.7 TBATS forecast for Duquesne Light Company Energy consumption

A TBATS model differs from dynamic harmonic regression in that the seasonality is allowed to change slowly over time. Figure 3.7 shows the TBATS forecast with the RMSE=276.551 for the test-samples, which presents a better out-of-sample performance comparing to Dynamic harmonic regression.

Since BATS require much more parameters to estimate, especially with long time series, if we consider a comparison between BATS model and TBATS model, we take a much smaller subset of training sample from Year 2018Jan1$^{st.}$ to 2018July1$^{st.}$ to save time. Thus, yearly seasonal pattern does not need to be considered in this time series.

Figure 3.8 shows the BATS forecast in Year 2018. The out-of-sample performance for BATS model is RMSE=323.278, while for TBATS model with RMSE=265.773.

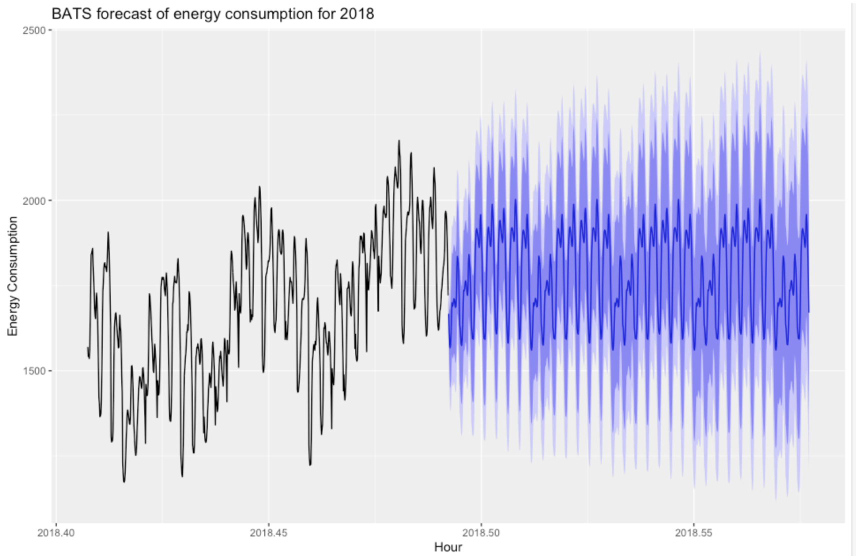

Figure 3.8 BATS forecast for Duquesne Light Company Energy consumption

There is one more feature should notice, TBATS model usually provides with a wider prediction interval than the BATS model, the comparison of the forecast plot is shown in Figure 3.9 (in Appendix). What's more, a comparison of the decomposition of time series basing on these two methods is also shown in Figure 3.10 (in Appendix). From Figure 3.10, BATS model only detects two seasonal patterns, and there is sequence accumulation in the hourly pattern, indicating that seasonal effects overlapped.

Here, the estimated fitted $BATS\{\omega, \phi, (p,q), \{m_1, m_2\}\}$ model with parameters is shown below:

Table 2: Parameters chosen for application of the BATS model

Parameters	ω	ϕ	α	β	γ_1	γ_2	φ_1	φ_2	φ_3
Value	1	0.982	0.043	-0.0005	0.058	-0.002	1.163	-0.130	-0.147

The estimated value of $\omega = 1$ implies a Box-Cox transformation on the time series. The estimated value $\phi = 0.982$ implies a damping effect on the growth rate. The model implies that the irregular component of the series is correlated and can be described by an ARMA$(3,0)$ process.

Discussion of Results and Conclusions

We mainly use four models in this paper to forecast the time series with multiple seasonal patterns and make a comparison among these methods. MSTL model allows the seasonal component to change over time. The rate of change, as well as the smoothness of the trend-cycle, can be controlled by the user. Also, this method is robust to the outliers. Although it has the lowest forecast accuracy, as the model with the simplest form, it is easy to interpret when the time series is short.

To allow errors from regression to contain autocorrelation, the Dynamic harmonic regression model is included to solve the problem in exponential smoothing methods. Also, a harmonic approach helps to fit the time series with multiple seasonal periods. However, this model doesn’t perform well in our example since the training sample has long-term dynamics, which are hardly handled.

In order to fix these problems, the TBATS model uses a combination of Fourier terms with an exponential smoothing state-space model. It allows slow change over time, which makes a progress basing on the dynamic harmonic regression model. What’s more, the trigonometric method that is added to the improvement of the BATS model helps with non-integer seasonality as well as to reduce the parameters of the model when the frequencies of seasonalities are high. However, TBATS still has a weakness that the performance for long-term prediction is not very well. In predicting the hourly energy consumption from Duquesne Light Company, the TBATS model performs the lowest RMSE among all the methods. We also make a comparison of forecast accuracy between the BATS model and the TBATS model, a similar solution that a better post-sample forecasting performance is obtained from the TBATS model gained as in Rob J Hyndman's (2010) study[4].

Reference

[1]Data Source: https://www.kaggle.com/robikscube/hourly-energy-consumption?select=DUQ_hourly.csv
[2] Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–33. http://bit.ly/stl1990
[3] Harvey, A., Koopman, S. J. & Riani, M. (1997), 'The modeling and seasonal adjustment of weekly observations', Journal of Business and Economic Statistics 15, 354–368.
[4] Livera, Alysha & Hyndman, Rob & Snyder, Ralph. (2010). Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. Journal of the American Statistical Association. 106. 1513-1527. 10.1198/jasa.2011.tm09771.
[5] Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion). *Journal of Official Statistics*, 6, 3--73.

[6] "Chapter 9 Dynamic Regression Models." *Forecasting: Principles and Practice*, by Rob J. Hyndman and George Athanasopoulos, OTexts, 2018.

[7] Akram, M., Hyndman, R. & Ord, J. (2009), 'Exponential smoothing and non-negative data', Australian and New Zealand Journal of Statistics 51(4), 415–432.
[8] Anderson, B. D. O. & Moore, J. B. (1979), Optimal filtering, Prentice-Hall, Englewood Cliffs.