ECO374 - Winter 2021

Forecasting Time Series Values

Eun Lim - 1005280839

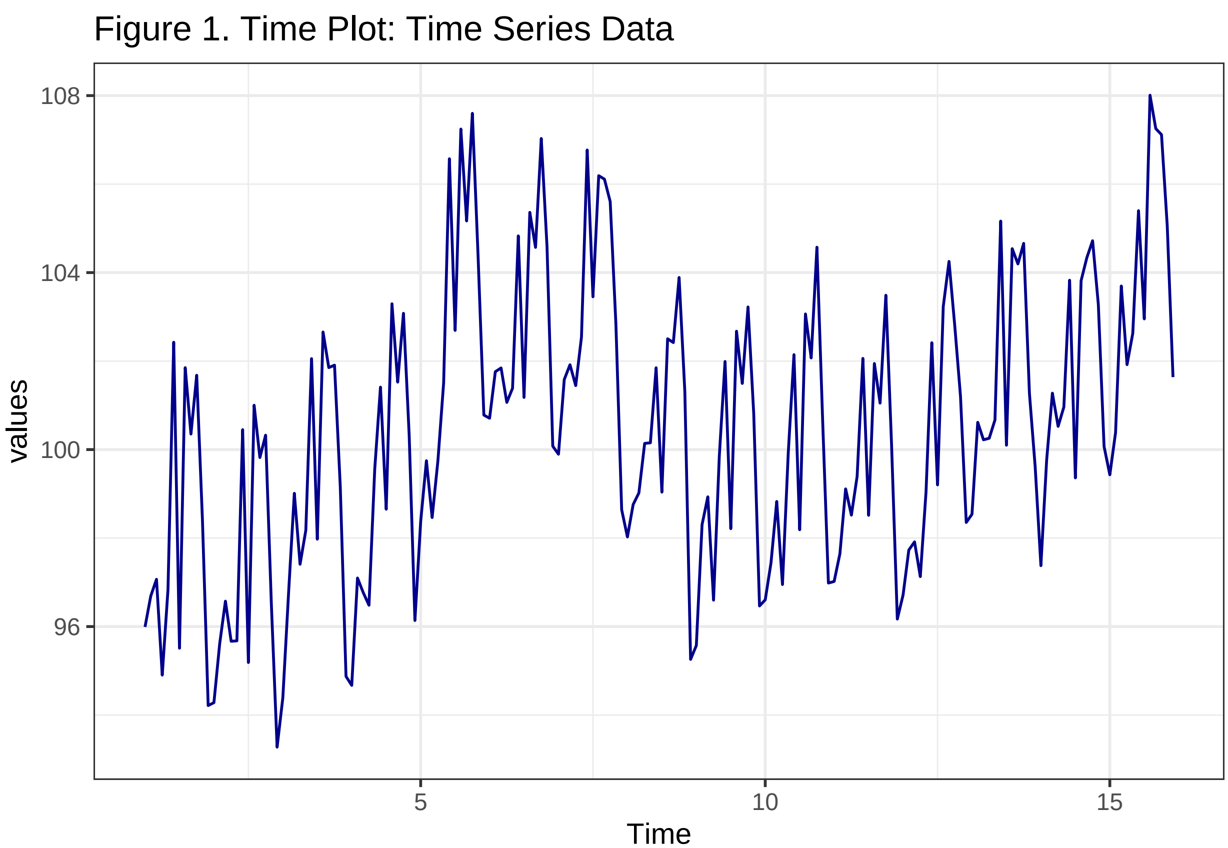
February 12, 2021

Introduction

In this paper, a part of a time series data that contains monthly data from period 1 to T is given. Using the given data, the 12 periods ahead data is predicted. The 12 periods ahead forecasts are driven as a result of applying regression method and first differencing method, assuming that the equation of the time series is,$Y_t = T_t + S_t + \epsilon_t$, where Y_t denotes the time series value (actual data) at period t. T_t denotes the trend component at time t. T_t denotes the seasonal component at time t, and ϵ_t denotes the random error at time t. Equivalently, the time series value can be decomposed into the 3 components, trend, seasonality and random errors.

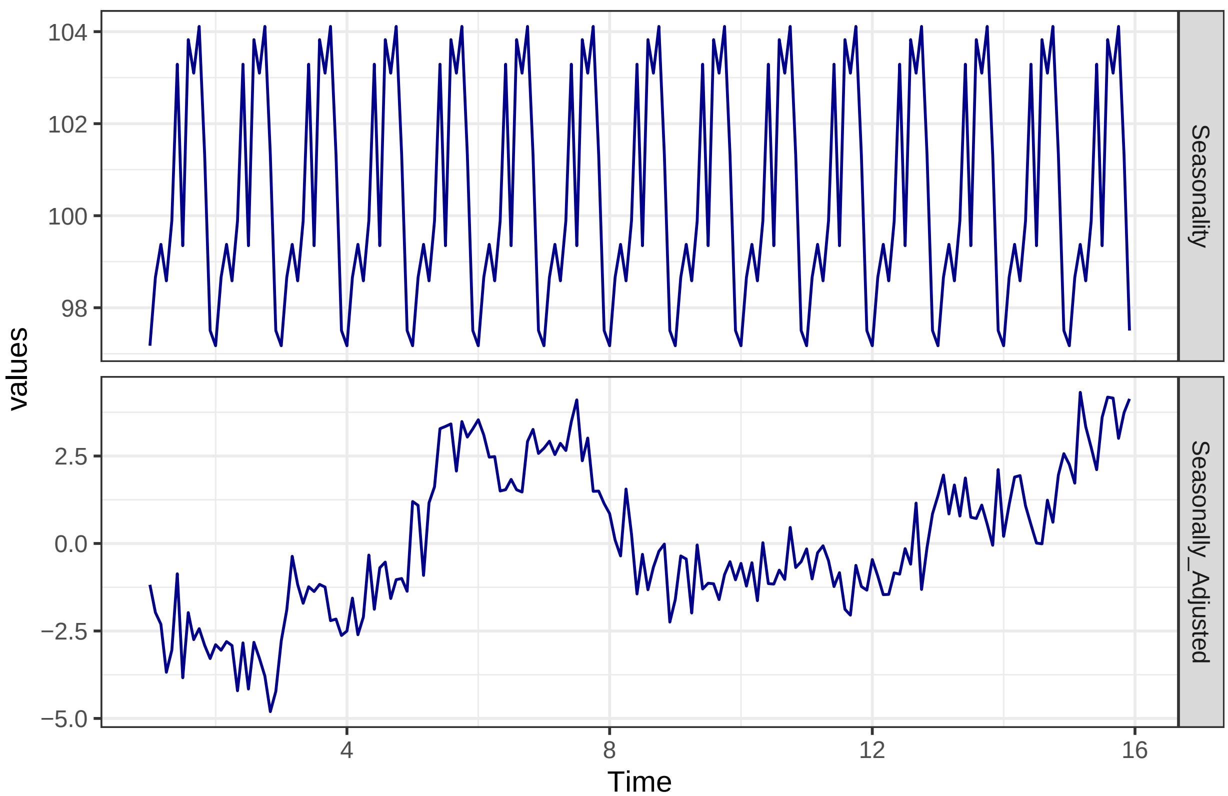
To compute the 12 periods ahead forecasts, the time series is deseasonalized by regression method because silmilar seasonal fluctuations are detected. In contrast to the deterministic seasonality, the seasonally adjusted series has an inconsistent and random upward trend rather a constant upward trend. Thus, it is concluded that there is stochastic trend in the series. Consequently, the seasonally adjusted series is detrended by the first differencing method. Then, the Ljung-Box test is conducted to test if there is autocorrelation in the residual series to determine if further modelling is required. The the stationarity of the residual series is tested by the augmented Dickey-Fuller test. The ACF and PACF plots are used to determine the maximum values of p and q for the ARIMA and seasonal ARIMA model. Then, the 12 period ahead forecasting is made.

Forecasting



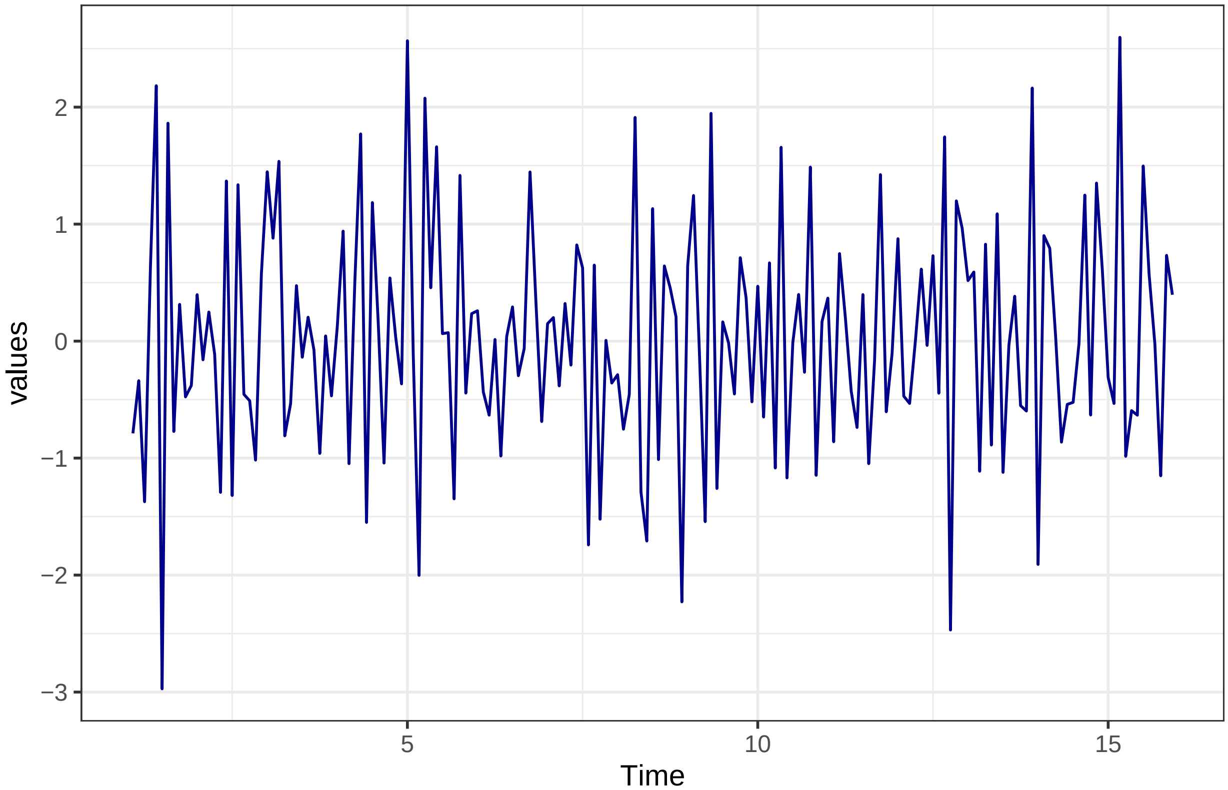
The Figure 1. shows that the time series is clearly non-stationary, with some seasonality and trend. Thus, the time series values are regressed on seasonal dummy variables to deal with the seasonality. The seasonally adjusted data and the seasonal component is shown in Figure 2.

Figure 2. Time Plot: Seasonality vs. Seasonally Adjusted



The seasonally adjusted series still appears non-stationary due to the stochastic trend. Accordingly, the seasonally adjustes series is detrended by first differencing with the lag of 1. This series is the residual series and is shown in Figure 3.

Figure 3. Residual Series



The the first differenced series looks stationary. Then, the Ljung-Box test up to lag 12 is conducted for autocorrelation to determine whether further modelling is needed under the hypotheses,

$$\begin{aligned} H_0 &: \text{ Futher modelling is not required.} \\ H_1 &: \text{ Further modelling is required.} \end{aligned}$$

The p-values of the five Ljung-Box tests, from lag of 1 to lag of 12, are $2.107e^{-8}$, $1.522e^{-7}$, $4.934e^{-7}$, $1.428e^{-6}$, $4.475e^{-6}$, $5.337e^{-6}$, and $2.637e^{-5}$ respectively. Accordingly, the null hypothesis is rejected.

The stationarity is also formally tested by the augmented Dickey-Fuller test under the hypotheses,

$$\begin{aligned} H_0 &: \text{ The residual has an unit root.} \\ H_1 &: \text{ The residual does not have an unit root.} \end{aligned}$$

Similar to the Ljung-Box test, the p-values of all three types of models, no drift no trend, with drift no trend and with drift and trend, are reported to be less or equal to 0.01, so the null hypothesis is rejected. Consequently, the residual is stationary and can be used.

The ACF and PACF plots, shown in figure 4. and figure 5., are used to determine the `max.p, max.q, max.P, max.Q` arguments of the `auto.arima()` command in `R`.

Figure 4. Residual: ACF

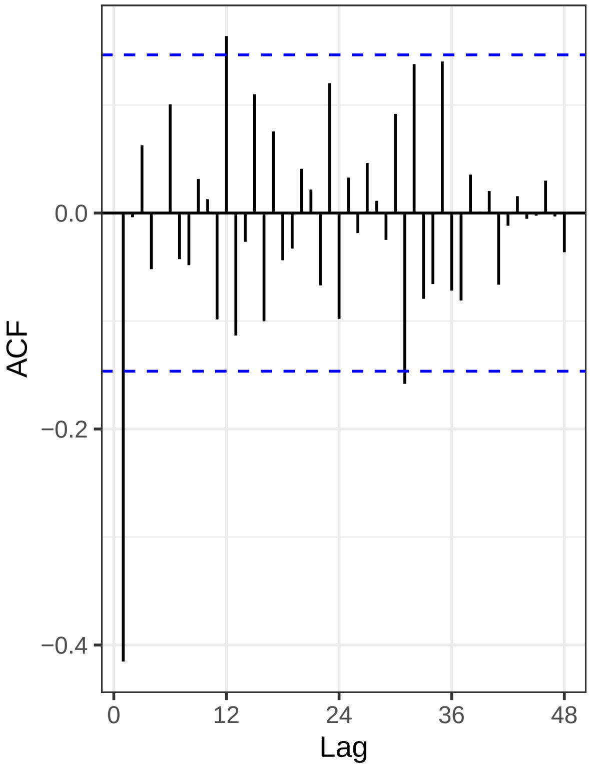
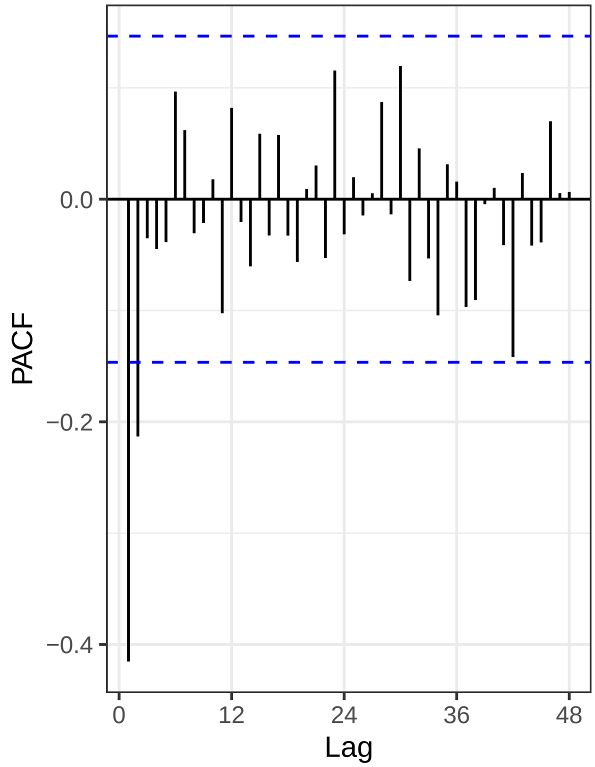


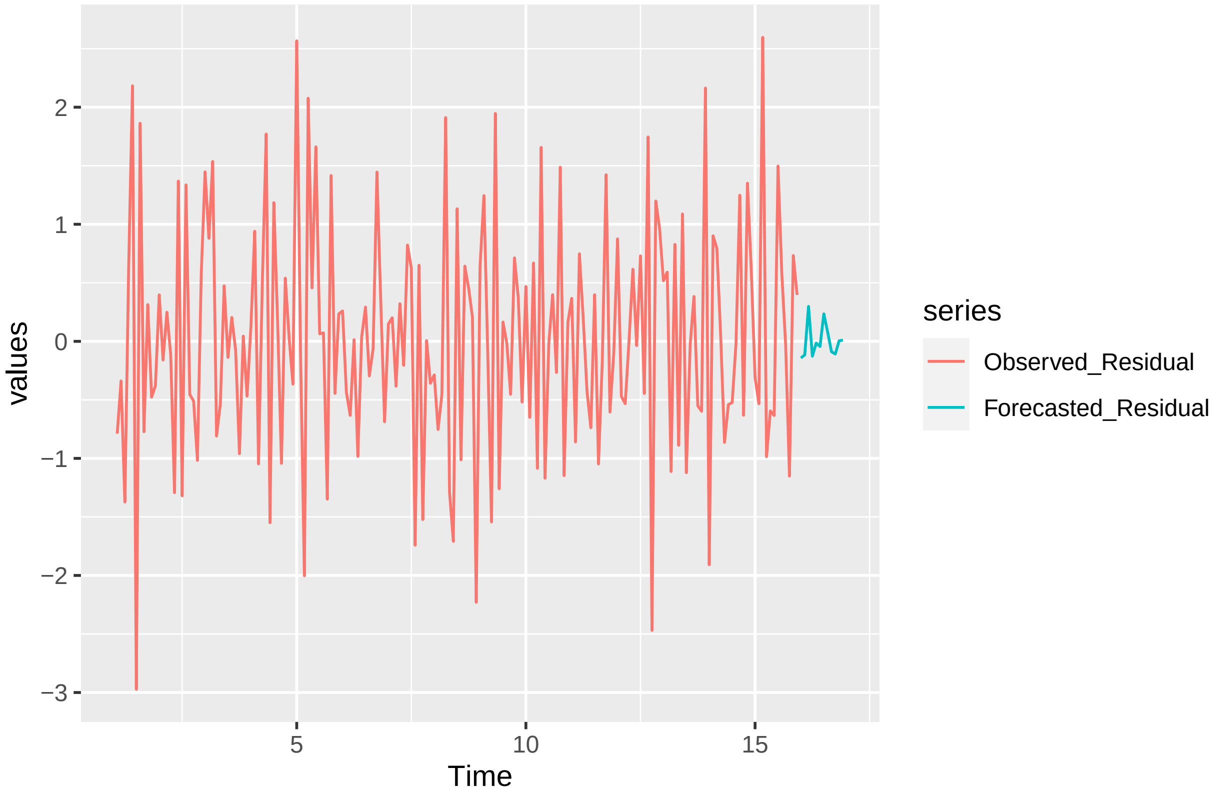
Figure 5. Residual: PACF

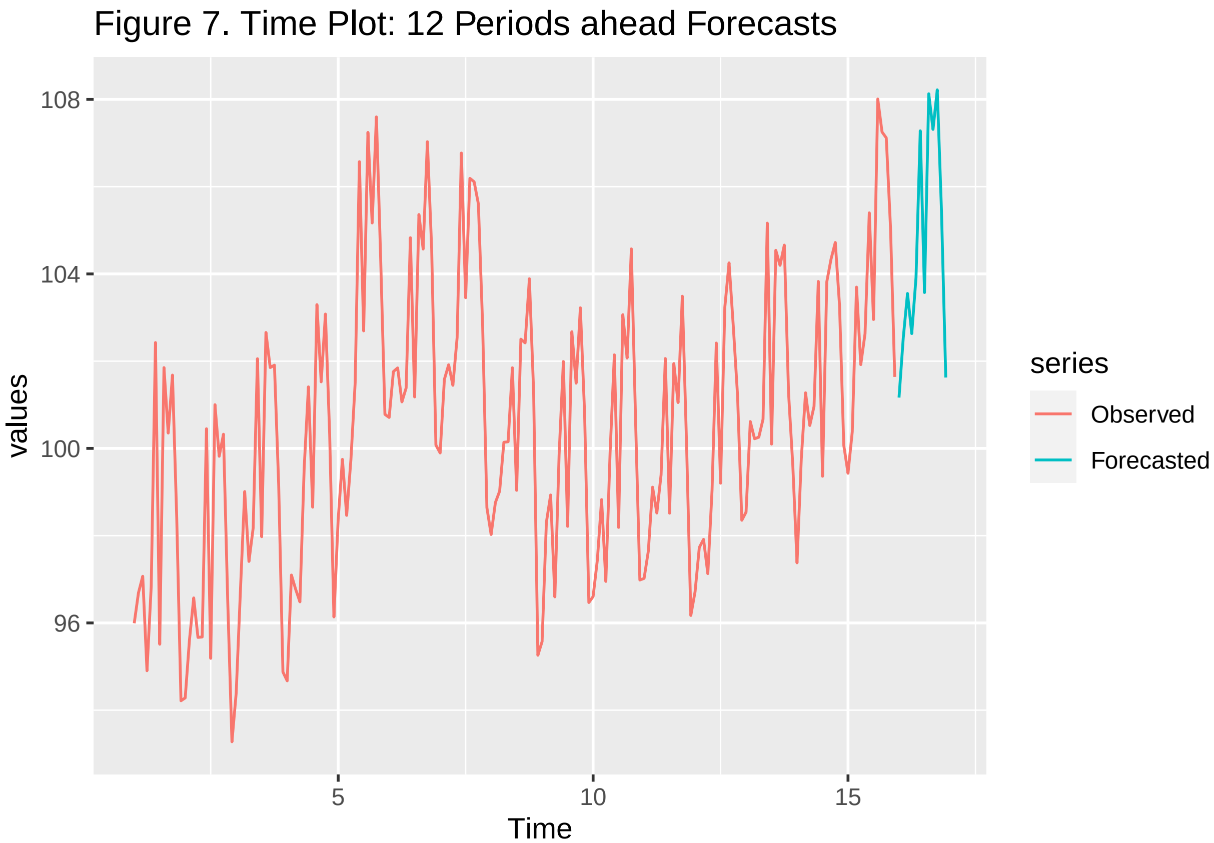


The ACF is no longer significant after lag of 31 and the PACF is no longer significant after lag of 2. The seasonal p and q are not detected, so the maximum values of p and q for `auto.arima()` command are 2 and 31, where the maximum values of P and Q are not specified.

As a result, the ARIMA(1,0,2) with the seasonal ARIMA(2,0,0) is computed as the best model for the residual. So, the 12 periods ahead residual series is predicted based on the model. Then, the 12 periods ahead time series is predicted, as shown in figure 3 and figure 4.

Figure 6. Time Plot: 12 Periods ahead Residual Forecasts





The forecasted residual series, figure 6., shows the random error. In Figure 7., the seasonality and trend is added to the show the forecasted series. The 12 periods ahead series is well predicted as the shape of the fluctuation captures the shape of the repeated monthly fluctuation in the observed series and the "inconstant but generally moving upward" trend is well illustrated.

Conclusion

Given an unknown monthly time series data, the 12 periods ahead forecast of the time series value is computed by the ARIMA(1,0,2) with the seasonal ARIMA(2,0,0) model.

Due to the seasonality and the trend, the given time series is not stationary. Accordingly, deseasolization and detrending are performed by the regression method and first differencing method.

To formally test the autocorrelation and the stationarity of the residual series, the Ljung-Box tests and augmented Dickey-Fuller tests are conducted.

The illustration of the 12 periods ahead forecasts of the time series value captures the seasonality, trend and the random error well.