Unit 16: Seasonal ARIMA Models

Jeffrey Woo

Department of Statistics, University of Virginia

Spring 2016

Readings for Unit 16

Textbook chapter 3.9.

Last Unit

1. ACF, PACF for seasonal ARMA models.
2. Multiplicative Seasonal ARMA Models.

This Unit

1. SARIMA models to account for non-seasonal and seasonal differencing.
2. Building SARIMA models.

Motivation

In Unit 15, we looked at the pure seasonal ARMA model, and the multiplicative seasonal ARMA model. These models assumed stationarity. In this unit, we consider non-stationarity and apply differencing to the non-seasonal and seasonal components.

Seasonal Trend

Consider the following time series:

$$x_t = S_t + y_t$$

where y_t is stationary and S_t is a seasonal trend.

Seasonal Trend

Since S_t is a seasonal trend, we have

$$S_t = S_{t-s} = S_{t+s}$$

where s is the length of the period. For example, for monthly data, a reasonable choice of s is 12. For quarterly data, $s = 4$. Think about

$$S_t = cos(2\pi \frac{t}{s}).$$

Then S_t is a periodic function with period s. How can we get rid of this trend? In the past, we've done a number of things namely regression, smoothing, and differencing.

Seasonal Trend

We usually apply seasonal differencing, where we difference across the period.

$$(1 - B^s)x_t = x_t - x_{t-s} = S_t - S_{t-s} + y_t - y_{t-s} = y_t - y_{t-s}.$$

If $s = 12$, then we have a stationary $\mathrm{MA}(1)_{12}$.

Seasonal Trend

In general, we need seasonal differencing when the ACF of the time series ______________________________. We can represent the differencing using the following notation

$$\nabla_s^D = (1 - B^s)^D.$$

Typically, $D = 1$ is sufficient to obtain seasonal stationarity.

SARIMA Model

In many settings with seasonal data, ____________ components may contribute to the model. For example, in monthly sales of ice cream, sales in the previous month (or two), together with sales from the same month a year ago, may help predict future sales.

SARIMA Model

If a linear trend is also present in the data (along with seasonality), we will probably also need a non-seasonal difference. Therefore, a non-seasonal and seasonal difference will be applied. We end up analyzing

$$(1 - B^{12})(1 - B)x_t =$$

SARIMA Model

This type of differencing leads us to the definition of the full SARIMA model which we denote by

$$\mathrm{ARIMA}(p, d, q) \times (P, D, Q)_s$$

and the model is

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d x_t = \alpha + \Theta_Q(B^s)\theta(B)w_t, \tag{1}$$

SARIMA Model

where

$$
\begin{aligned}
\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \\
\theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q \\
\Phi_P(B^s) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps} \\
\Theta_Q(B^s) &= 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs} \\
\nabla^d x_t &= (1 - B)^d x_t, \quad \nabla_s^D x_t = (1 - B^s)^D x_t
\end{aligned}
$$

SARIMA Model

Question: How do we write an ARIMA$(0, 1, 1) \times (0, 1, 1)_4$ model?

SARIMA Model

Question: Identify the following as a multiplicative seasonal ARIMA model

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} + w_t - 0.5w_{t-1} - 0.5w_{t-12} + 0.25w_{t-13}.$$

SARIMA Model

The rest of the analysis is very similar to what we did for the regular ARIMA models. We will use plotting including ACF and PACF to help discover s and also the amount of differencing that is necessary. Then we will attempt to fit appropriate models using AIC and parameter estimates to guide us. We will use the same diagnostics to evaluate the appropriateness of the fits.

Building SARIMA Models

Just like how we built ARIMA models in unit 14, the main steps in building SARIMA models consist of the following:

- Exploratory data analysis.
- Model estimation.
- Model diagnostics.
- Model selection.

Exploratory Data Analysis

We typically look at

- the time series plot,
- the ACF,
- and the PACF

of the data. This step guides us to our choice for the elements of the SARIMA model, p, d, q, P, D, Q, s.

Exploratory Data Analysis: Time Series Plot

With the time series plot, we examine for trend and seasonality. Typically, understanding what the data describes will help us know what the period is (month, quarterly, etc), and what lags to look out for, so you should have an idea of what value s is.

Exploratory Data Analysis: Differencing

The general guidelines for differencing are:

- Seasonality and no trend: Take a difference of lag s.
- Linear trend and no seasonality: Take first difference like with ARIMA model. If quadratic trend and no seasonality, take second difference.
- Seasonality and trend: Apply non-seasonal and seasonal differences, as successive operations in either order.

Based on these, you should have an idea of what d and D are.

Exploratory Data Analysis: ACF and PACF

Examine the ACF and PACF of the differenced data. Some guidelines:

- Non-seasonal terms: Examine the early lags $(1, 2, 3, \cdots)$ to judge the values of p and q. This will be similar to what was done with ARIMA models (Unit 14).
- Seasonal terms: Examine the pattern of the lags that are multiples of s. Typically you should look at the first two or three multiples. Judge the values of P and Q as in slide 23 from Unit 15.

Model Estimation

After exploratory data analysis, you should have an idea (or ideas) about the values of p, d, q, P, D, Q, s. When using R, I highly suggest using the original data, and then specify the values of the differencing used.

Model Estimation: Significance of Estimates

For the estimated coefficients of the parameters, use the z statistic

$$z = \frac{\text{estimated coefficient}}{\text{s.e. of coefficient}}.$$

If $|z| > 1.96$, the estimated coefficient is significantly different from 0. This is exactly what was done with ARIMA models.

Model Diagnostics

For model diagnostics, we usually check the following (from sarima() function):

- Time series of standardized residuals.
- ACF of residuals.
- Ljung-Box-Pierce statistic.

Use the same criteria with ARIMA models.

Model Diagnostics

If at least one of these plots is unreasonable for a model, the model has to be discarded. If all the candidate models did not have satisfactory diagnostic plots, some questions you have to ask yourself include whether applying a log transform (to stabilize the variance of the data), whether the correct differencing operation(s) was applied, whether you failed to consider other plausible values of p, q, P, Q etc.

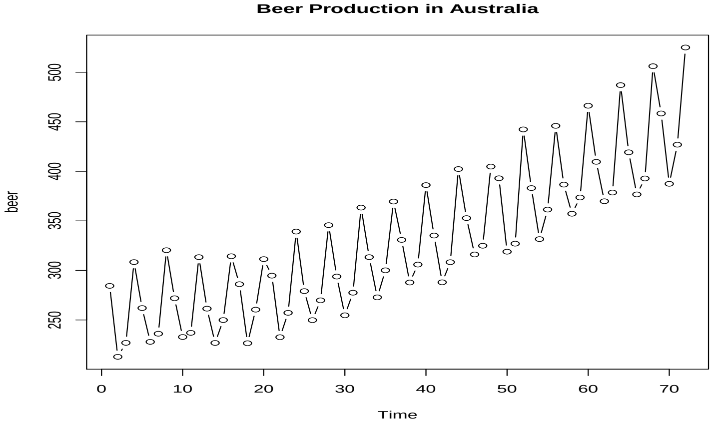
Model Selection

If model diagnostics suggest more than one model works, here are some issues to keep in mind when comparing models:

- Smaller MSE, AIC, etc.
- Simpler model.
- Standard errors of forecasts.

Worked Example: Beer Production

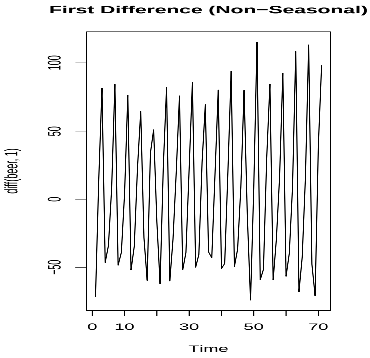
The dataset contains quarterly beer production in Australia.

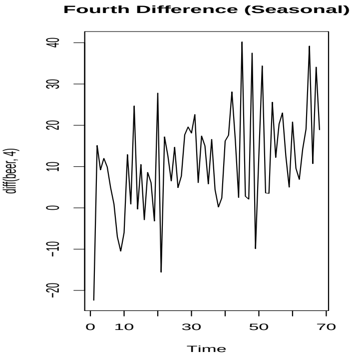


Comments on trend and/or seasonality?

Worked Example: Beer Production

Based on the time series plot, we should apply both a first difference (non-seasonal) and difference with lag 4 (seasonal) to the data. Out of curiosity, we may wish to look at the time series plots of the data when just one of these differences is applied.

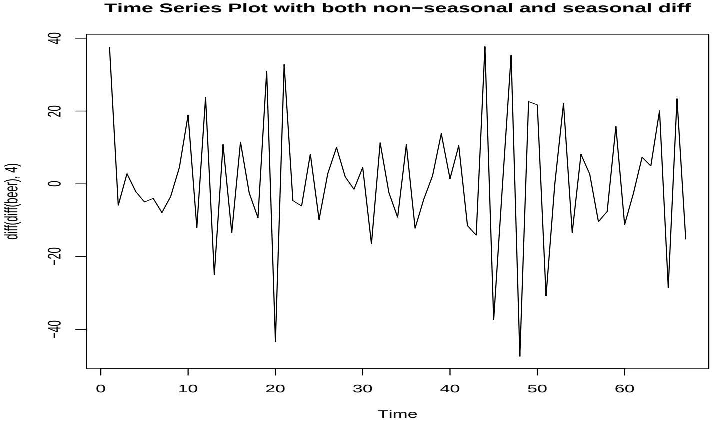




Comments?

Worked Example: Beer Production

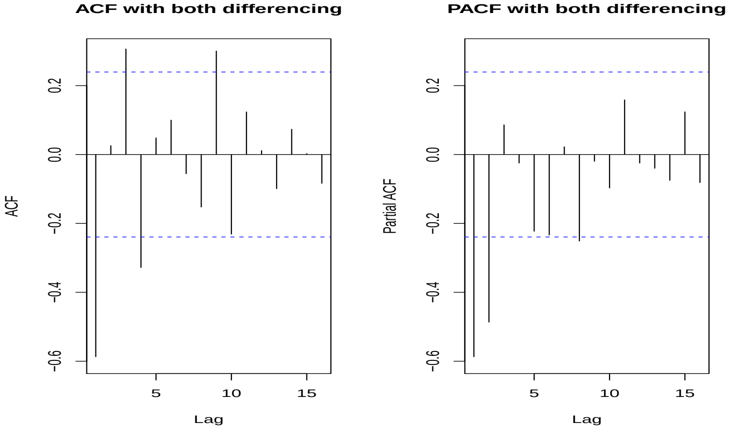
Take a look at the time series of the data after both differencing operations have been applied.



$d = 1, D = 1, s = 4$ appear to be reasonable.

Worked Example: Beer Production

Examine the ACF and PACF of the data after both differencing operations.



Worked Example: Beer Production

Question: Thoughts on non-seasonal components, p, q? How about seasonal components, P, Q?

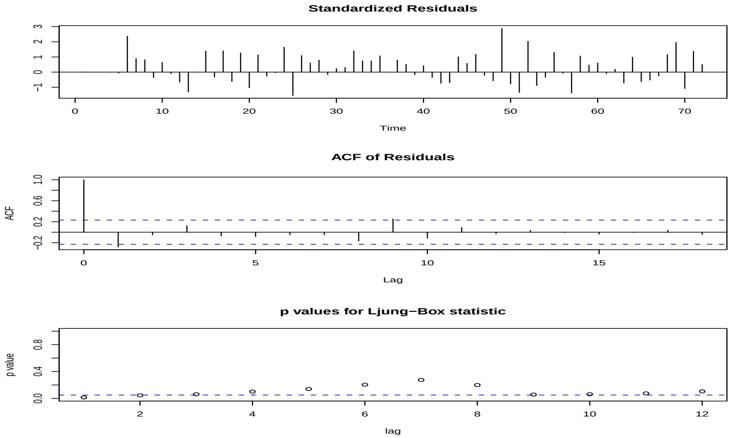
Worked Example: Beer Production

Consider the following models, based on ACF and PACF.

1. ARIMA$(2, 1, 0) \times (0, 1, 2)_4$
2. ARIMA$(0, 1, 1) \times (0, 1, 2)_4$
3. ARIMA$(1, 1, 1) \times (0, 1, 2)_4$

Worked Example: Beer Production

Check diagnostic plots. Model 2 had diagnostics indicating model is not a good fit, so we remove model 2 from consideration.



Worked Example: Beer Production

Let's check the AIC, AIC_c, and BIC for the remaining two models.

```
> f1.cri
[1] 5.812033 5.852437 4.938514
> f3.cri
[1] 5.786946 5.827350 4.913428
```

Model 3, ARIMA$(1,1,1) \times (0,1,2)_4$, has the smallest AIC, AIC_c, and BIC.

Worked Example: Beer Production

Check the significance of the parameters for
ARIMA$(1,1,1) \times (0,1,2)_4$.

```
Coefficients:
          ar1      ma1     sma1    sma2
      -0.3311  -0.7182  -0.5798  0.0042
s.e.   0.1368   0.0958   0.1541  0.1374
```

Notice that Θ_2 is __________. So we may decide to drop that term and try to fit ARIMA$(1,1,1) \times (0,1,1)_4$, which is not one of the models we originally considered.

Worked Example: Beer Production

Model diagnostics for ARIMA$(1,1,1) \times (0,1,1)_4$ suggest the model fits adequately. So we compare the AIC, AIC_c, and BIC for ARIMA$(1,1,1) \times (0,1,2)_4$ and ARIMA$(1,1,1) \times (0,1,1)_4$.

```
> f3.cri
[1] 5.786946 5.827350 4.913428
> f4.cri
[1] 5.759288 5.795358 4.854149
```

So what is your decision (or are there other things to consider)?