MATH 7339 - Machine Learning and Statistical Learning Theory 2

Section ARIMA

1. ARIMA
2. Seasonal ARMA
3. Multiplicative seasonal ARMA model.
4. Seasonal Trend
5. SARIMA Model

Motivation

Thus far, we have assumed stationary ARMA models.

We next consider ARIMA models, which basically consider differences of the observations to coerce stationarity. We then see how to go about building an ARIMA model using techniques we've looked at previously (and some new ones).

These techniques are namely:

- exploratory data analysis (plots, transformations, identifying potential models),
- model estimation,
- model diagnostics, and
- Model selection.

Integrated Models for Nonstationary Data

Recall when we discussed exploratory data analysis and smoothing that we sometimes began with a model such as

$$X_t = \mu_t + Y_t$$

where μ_t was a non-stationary component (trend) and Y_t was a stationary zero-mean process.

If we assume μ_t was of the form $\beta_0 + \beta_1 t$, then the differencing yields

$$\nabla X_t = \beta_1 + \nabla Y_t$$

which is a stationary time series.

In general if μ_t is a polynomial in t, such as $\mu_0 + \mu_1 t + \cdots + \mu_d t^d$, then

$$\nabla^d X_t = d!\,\beta_d + \nabla^d Y_t$$

Definition: A process is ARIMA(p, d, q) if

$$\nabla^d X_t = (1 - B)^d X_t$$

is $\mathrm{ARMA}(p, q)$.

Note: ARIMA models are sometimes called Box-Jenkins models.

Example: Recall the random walk: $X_t = X_{t-1} + W_t$

X_t is not stationary, but $Y_t = (1 - B)X_t = W_t$ is a stationary process. In this case, it is white, so $\{X_t\}$ is an ARIMA(0,1,0).

Also, if X_t contains a trend component plus a stationary process, its first difference is stationary.

If the mean of $\nabla^d X_t$ is zero then we can write the ARIMA model as

$$\phi(B)(1-B)^d X_t = \theta(B) W_t$$

If a mean does exist then we can write

$$\phi(B)(1-B)^d X_t = \alpha + \theta(B) W_t$$

where $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$

ARIMA models example

We usually use a first difference to account for a linear trend in the data. A second difference may be used to account for a quadratic trend in the data.

Recall that differencing can sometimes introduce more dependence in the data. Let's assume that our data has the following model

$$X_t = \beta_0 + \beta_1 t + Y_t$$

where

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + W_t$$

Consider $\nabla X_t = \beta_1 + \nabla Y_t$

Denote $\nabla Y_t = Z_t$, we have

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + W_t - W_{t-1}$$

which is ARMA(2,1). Notice that the original noise term Y_t is an AR(2).

ARIMA models example

Suppose $\{X_t\}$ is an ARIMA(0,1,1):

$$X_t = X_{t-1} + W_t - \theta_1 W_{t-1}$$

If $|\theta_1| < 1$, then we can show

$$X_t = \sum_{j=1}^{\infty} (1-\theta_1)\theta_1^{j-1} X_{t-j} + W_t$$

So,

$$\tilde{X}_{n+1} = \sum_{j=1}^{\infty} (1-\theta_1)\theta_1^{j-1} X_{n+1-j}$$

$$= (1-\theta_1)X_n + \sum_{j=2}^{\infty} (1-\theta_1)\theta_1^{j-1} X_{n+1-j}$$

$$= (1-\theta_1)X_n + \theta_1 \tilde{X}_n$$

Exponentially weighted moving average.

❑ Building ARIMA models

We'll explore some techniques for identifying and estimating non-seasonal ARIMA models, as well as how to analyze the residuals after a model is estimated. Recall that ARIMA models are specified as ARIMA(p, d, q).

The main steps in building an ARIMA model are

1. Exploratory data analysis (Plot the time series and look for trends, seasonal components, step changes, outliers.)
2. Nonlinearly **transform** data, if necessary.
3. **Identify** preliminary values of d, p, and q.
4. **Estimate** parameters.
5. Use **diagnostics** to confirm residuals are white/iid/normal.
6. 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 ARIMA model, $p; d; q$.

Exploratory Data Analysis: Time series plot

We usually check for stationarity in a time series plot. Recall for stationarity, the plot should suggest the mean and the variance are constant. What we look out for in a time series plot:

- Trend (increasing, decreasing, quadratic).
- Increasing variability.
- Seasonality (discuss next).
- Outliers.

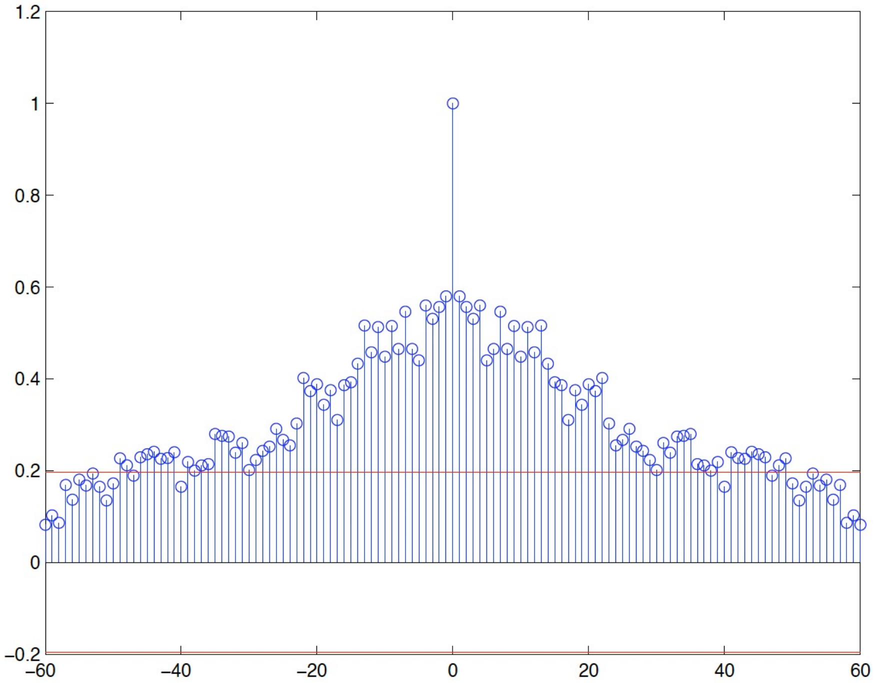
We usually try to stabilize the variance first. If the variance appears to be increasing, we can try transforming the data, usually a log transform.

If there's a linear trend, we consider a first difference. A quadratic trend suggests a second difference. We rarely go beyond $d = 2$, unless there is a contextual or scientific reason to do so.

Over-differencing can introduce unnecessary **dependency** in the model. **Smoothing** techniques may be used, where we analyze the smoothed data instead.

Exploratory Data Analysis: ACF and PACF

For identifying preliminary values of **d**, a time plot can also help.
Too little differencing: not stationary.
Too much differencing: extra dependence introduced.



Trends lead to slowly decaying sample ACF

For identifying p, q, look at sample ACF, PACF of $(1-B)^d X_t$

The ACF and PACF should be used together. Recall that

- AR(p) models have theoretical PACF with non-zero values for $h \leq p$, and zero values for $h > p$. The ACF should decay exponentially to zero.
- MA(q) models have theoretical ACF with non-zero values for $h \leq p$, and zero values for $h > q$. The PACF should decay exponentially to zero.
- ARMA(p,q) models have ACF and PACF that both decay exponentially to zero. The order will not be obvious. In such a case, we may just start o with $p; q = 1\, or\, 2$, and see what happens during model estimation and diagnostics.

Model:	**ACF:**	**PACF:**
AR(p)	decays	zero for $h > p$
MA(q)	zero for $h > q$	decays
ARMA(p,q)	decays	decays

- If the ACF and PACF decay slowly (do not decay exponentially), then the time series is likely to be not stationary (was differencing performed earlier?).
- If all the autocorrelations are insignificant for $h \geq 1$, than the series is (possibly-shifted) white noise.
- If all the autocorrelations are insignificant for $h \geq 1$ for a first difference, then we may have a random walk.

Model Estimation

After exploratory data analysis, you should have an idea (or ideas) about the values of p; d; q.

We use computer software (e.g. MATLAB , R, Python,) to estimate the parameters. Maximum likelihood estimation is usually used.

For the estimated coefficients of the parameters, use the **t-**statistic to show the significance of the estimates.

Model Diagnostics

For model diagnostics, we usually check the following:

- ACF of residuals.
- Ljung-Box-Pierce statistic.
- Plot of residuals against fitted values or time series of residuals.

1. ACF of residuals.

If our model diagnostics are acceptable, we expect our **residuals** to behave like white noise. If something appears unreasonable, you might have to revise your thought at what the model might be.

If you have a good model, all estimated ACFs of residuals should be **insignificant**. If this isn't the case, you probably need to explore a different model.

2. Ljung-Box-Pierce Statistic:

Recall that for white noise, the sample autocorrelations are approximately independent and normally distributed with mean 0 and variance $\frac{1}{n}$.

The Ljung-Box-Pierce Q-statistic takes into account the magnitudes of the sample autocorrelations as a whole.

The Ljung-Box-Pierce statistic is a function of accumulated sample autocorrelations, $\hat{\rho}(h)$, up to a specified time lag H. The Ljung-Box-Pierce Q-statistic is given by

$$Q(H) = n(n+2)\sum_{h=1}^{H}\frac{\hat{\rho}(h)^2}{n-h}$$

The choice of H is somewhat arbitrary. E.g. H = 20.

Under the null hypothesis that the model fits the data adequately, $Q \sim \chi^2_{H-p-q}$ as n large. A large Q-statistic leads to the rejection of the null hypothesis, i.e. model is not an adequate fit for the data.

3. Residual Plot

Using either a plot of residuals against fitted values, or a time series plot of the residuals, we check if the variance is constant. If the variance is not constant, you may need to transform the data.

Model Selection

Sometimes you may have more than one set of values for $p;\ d;\ q$ from exploratory data analysis. To be thorough, you may want to investigate the model estimation and diagnostics for more than one model. If model diagnostics suggest more than one model works, here are some issues to keep in mind when comparing models:

- Simpler model.
- Standard errors of forecasts.
- AIC, AICc, BIC etc.

Two different ARIMA models can be nearly equivalent when converted to an infinite order MA model using the causal representation.

❑ Seasonal ARMA models for seasonal time series.

So far, we've avoided seasonal data. The ARIMA models that we've discussed do not account for seasonality. However, we may wish to have a model for monthly observations which depends on both the previous month and the same month one year ago. SARMA models will allow us to do that.

We can write the **pure seasonal ARMA** model, **ARMA(P;Q)$_s$**, using backshift operators in the following way.

$$\Phi_P(B^s)X_t = \Theta_Q(B^s)W_t$$

where

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$$ *seasonal autoregressive operator*

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}$$ *seasonal moving average operator.*

It is **causal** iff the roots of $\Phi_P(z^s)$ are outside the unit circle.

It is **invertible** iff the roots of $\Theta_Q(z^s)$ are outside the unit circle

Example: Suppose you have **quarterly** data and want to think about an $ARMA(1,1)_4$. This would be

$$(1 - \Phi_1 B^4)X_t = (1 + \Theta_1 B^4)W_t$$

or

$$X_t = \Phi_1 X_{t-4} + W_t + \Theta_1 W_{t+4}$$

This is essentially an ARMA model, except lags between zero and four are omitted.

Let's consider monthly data and look at a seasonal MA(1). The model would be written as

Example: $P = 0, Q = 1, s = 12$. $X_t = W_t + \Theta_1 W_{t-12}$.

$$\gamma(0) = (1 + \Theta_1^2)\sigma_w^2,$$

$$\gamma(12) = \Theta_1 \sigma_w^2,$$

$$\gamma(h) = 0 \qquad \text{for } h = 1, 2, \ldots, 11, 13, 14, \ldots.$$

Example: $P = 1, Q = 0, s = 12$. $X_t = \Phi_1 X_{t-12} + W_t$.

$$\gamma(0) = \frac{\sigma_w^2}{1 - \Phi_1^2},$$

$$\gamma(12i) = \frac{\sigma_w^2 \Phi_1^i}{1 - \Phi_1^2},$$

$$\gamma(h) = 0 \qquad \text{for other } h.$$

ACF and PACF of Pure seasonal ARMA Models

When looking at ACF and PACF plots, we are going to use the same criteria as before but look only at the lags that are a multiple of the period.

A pure seasonal MA(1) should have a significant value for the ACF at the lag of the period and roughly zero otherwise. A pure seasonal AR(1) should tail off exponentially at the lag of the period and be roughly zero otherwise.

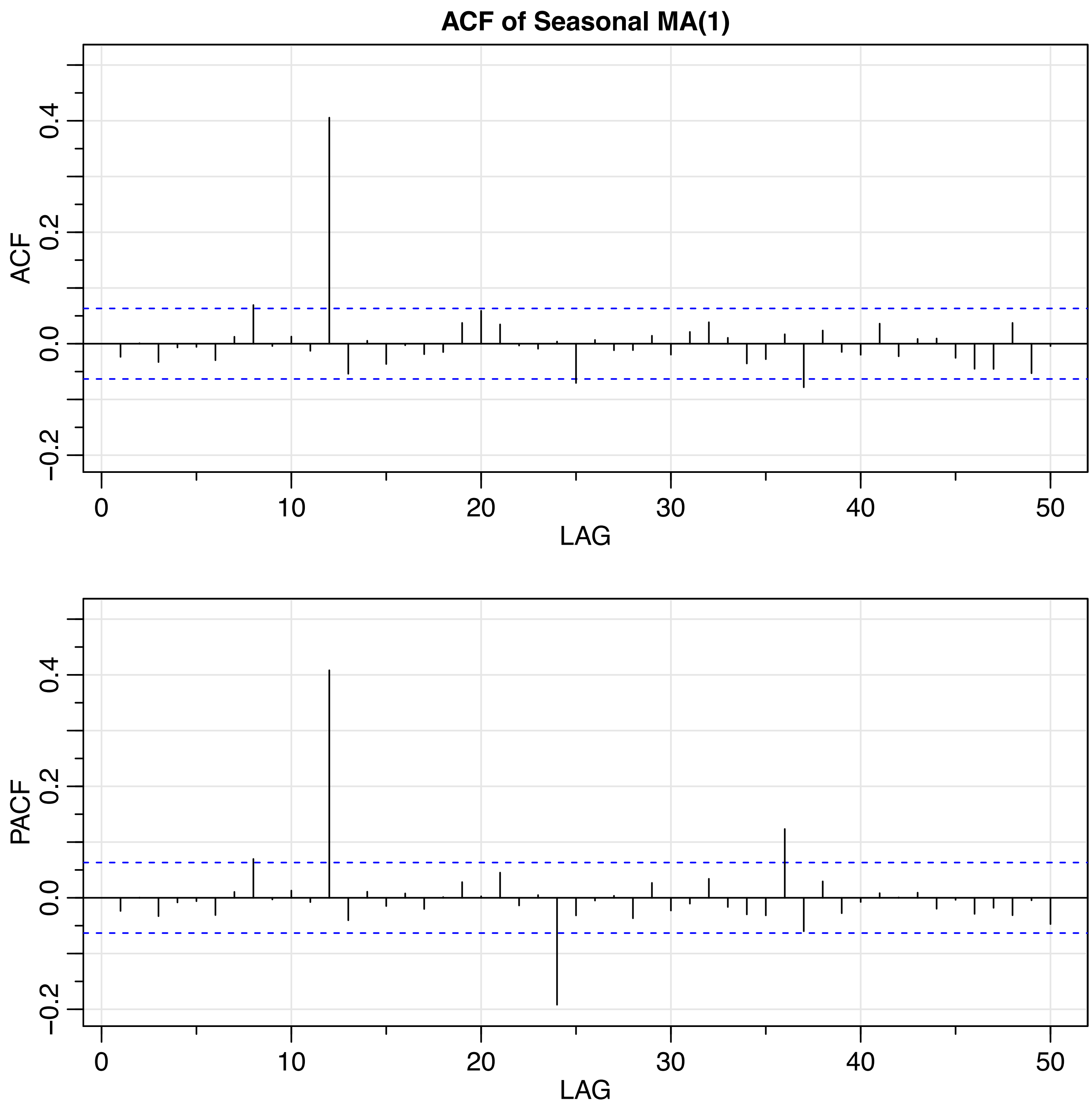
The PACF of a pure seasonal MA(1) should decay exponentially at multiples of the period and be zero otherwise. The PACF of a pure seasonal AR(1) should cut off after the lag of one period and should be zero for all other values.

For $ARMA(P;Q)_s$, both ACF and PACF tail off exponentially at multiples of the period.

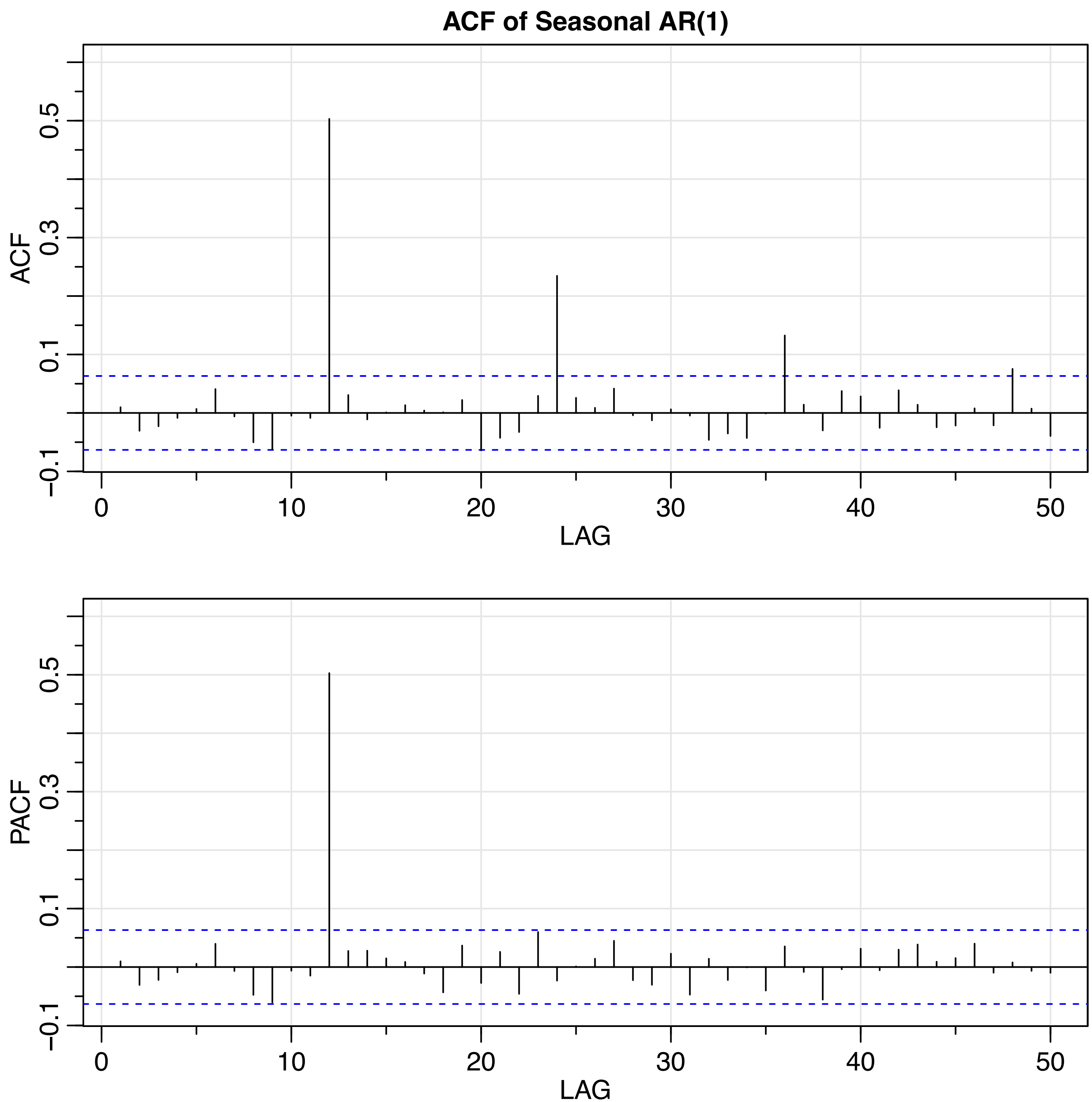
ACF and PACF for Seasonal ARMA

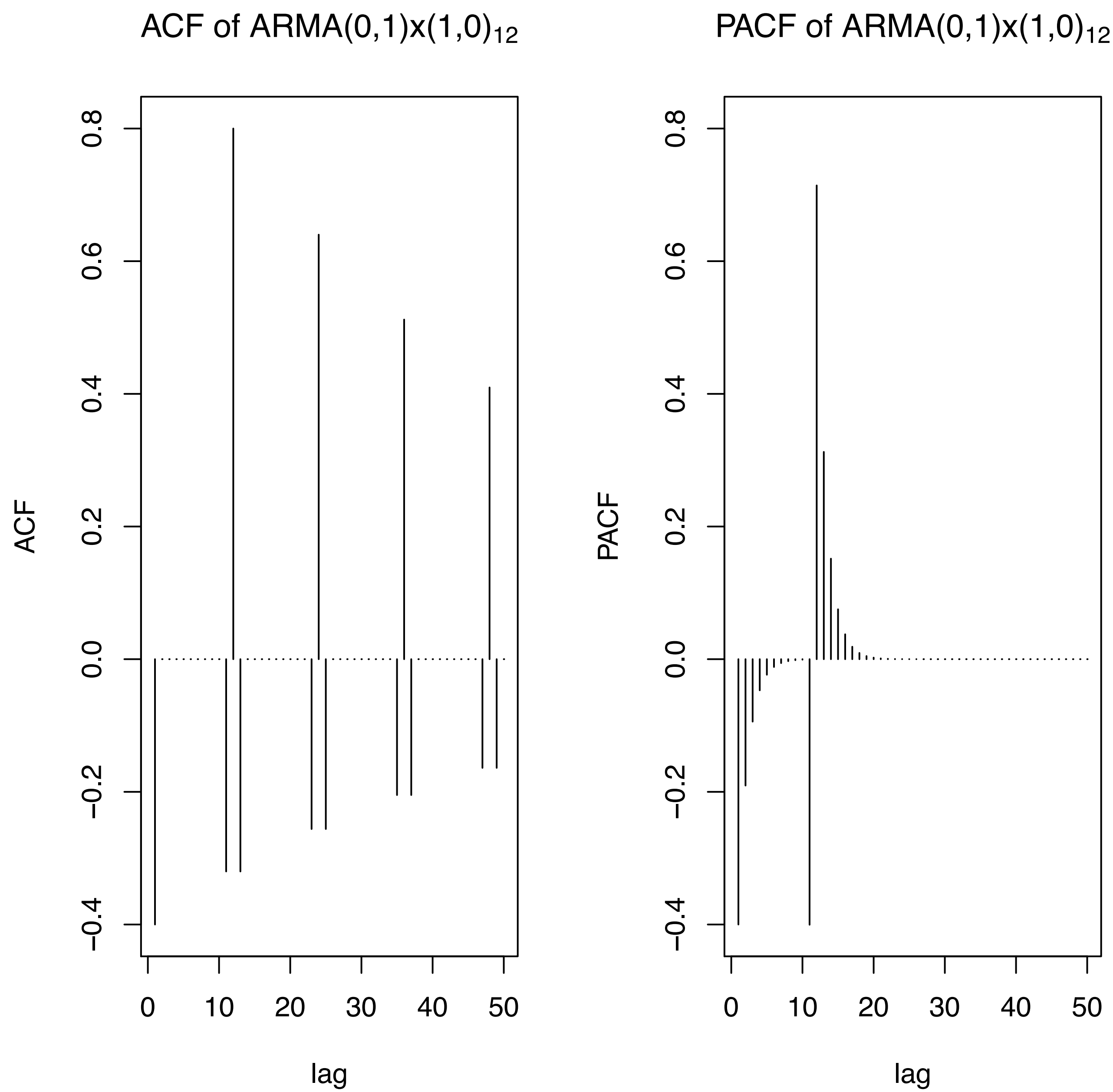
The ACF and PACF for a seasonal $ARMA(P,Q)_s$ are zero for $h \neq si$. For $h = si$, they are analogous to the patterns for ARMA(p,q):

Model:	**ACF:**	**PACF:**
$\mathrm{AR(P)}_s$	decays	zero for $i > P$
$\mathrm{MA(Q)}_s$	zero for $i > Q$	decays
$\mathrm{ARMA(P,Q)}_s$	decays	decays



ACF of Seasonal MA(1)
ACF
0.4
0.2
0.0
−0.2
LAG
0
10
20
30
40
50
PACF
0.4
0.2
0.0
−0.2
LAG
0
10
20
30
40
50


ACF of Seasonal AR(1)
ACF
0.5
0.3
0.1
−0.1
0
10
20
30
40
50
LAG
PACF
0.5
0.3
0.1
−0.1
0
10
20
30
40
50
LAG

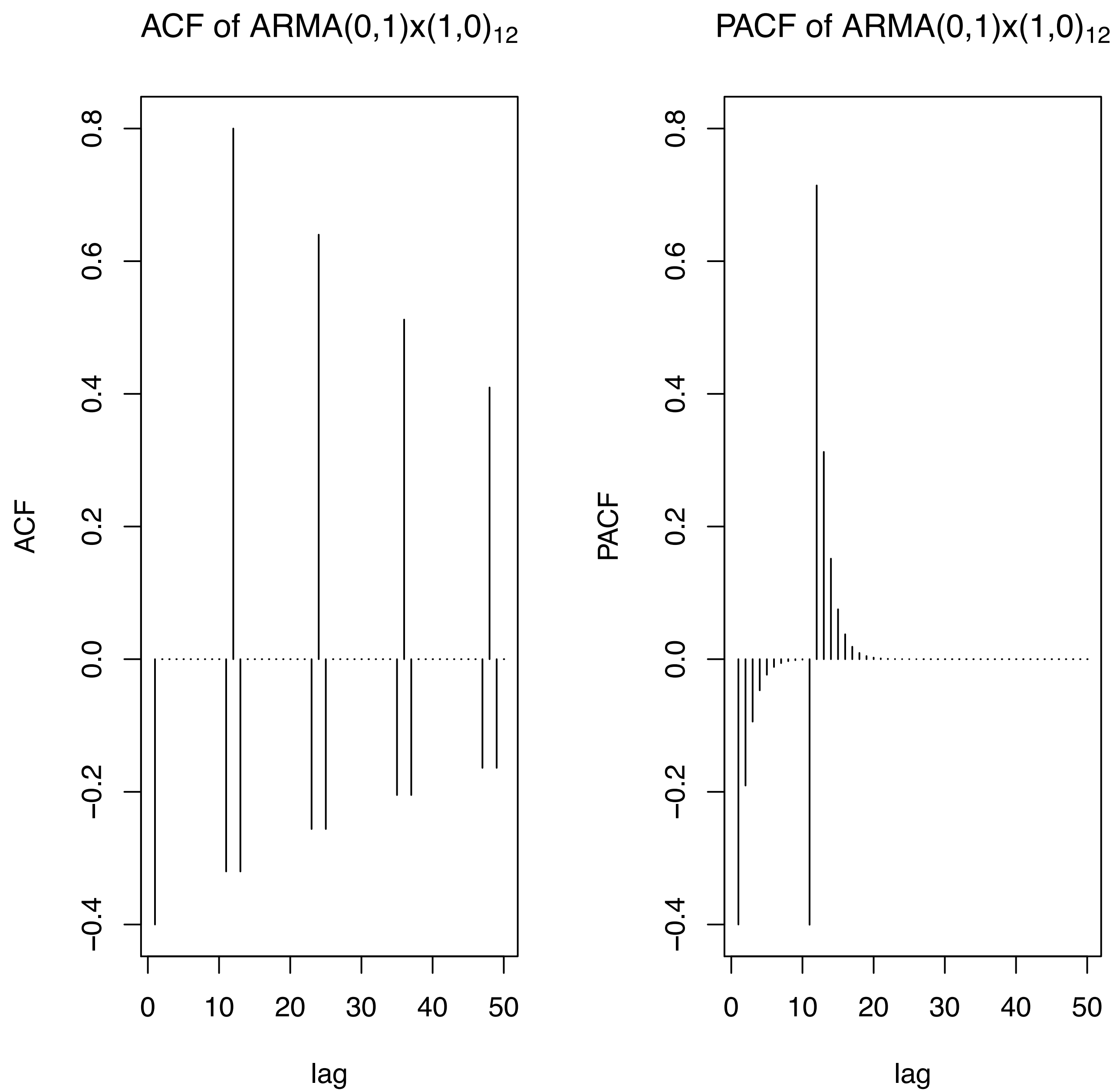

ACF of ARMA(0,1)x(1,0)12
PACF of ARMA(0,1)x(1,0)12
ACF
PACF
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
10
20
30
40
50
lag
lag

Multiplicative Seasonal ARMA Models

We can also combine the seasonal aspects and the regular ARMA models to get multiplicative seasonal autoregressive moving average models denoted $ARMA(p;\ q)\times(P;Q)_s$ *. We may write the model as*

$$\Phi_P(B^s)\phi(B)X_t = \Theta_Q(B^s)\theta(B)W_t$$

For the multiplicative models, we should expect to see a mix of patterns that we observe in non-seasonal and pure seasonal ARMA models.



ACF of ARMA(0,1)x(1,0)$_{12}$
PACF of ARMA(0,1)x(1,0)$_{12}$
ACF
PACF
lag
lag
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
10
20
30
40
50

Seasonal Trend

The pure seasonal ARMA model, and the multiplicative seasonal ARMA model assumed **stationarity**. We can also consider **non-stationarity** and apply **differencing** to the non-seasonal and seasonal components.

Consider the following time series:

$$X_t = S_t + Y_t$$

where Y_t is stationary and S_t is a 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 $s = 12$. For quarterly data, $s = 4$.

How can we get rid of this trend? In the past, we've done a number of things namely regression, smoothing, and differencing.

SARIMA Model

In many settings with seasonal data, shorter-term 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.

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

$$ARIMA(p; d; q) \times (P;D;Q)_s$$

The model is

$$\Phi_P(B^s)\phi(B)\nabla_s^D \nabla^d X_t = \alpha + \Theta_Q(B^s)\theta(B)W_t$$

where the seasonal difference operator of order D is defined by

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

$$\nabla^d X_t = (1 - B\ )^d X_t$$

$$\alpha = \mu\big(1 - \phi_1 - \cdots - \phi_p\big)(1 - \Phi_1 - \cdots - \Phi_P)$$

Building SARIMA Models

Just like how we built ARIMA models, 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, $p;\ d;\ q;\ P; D; Q;\ s$.

- "Time Series Analysis and Its Applications", 4th ed. 2017, by Shumway and Stoffer.

  Sections 3.6, 3.7.

Select ARIMA Model for Time Series:

https://www.mathworks.com/help/econ/box-jenkins-model-selection.html