Modeling Non-Stationary Seasonal Series

A series z_t with an (approximately) repeating seasonal pattern or tendency is non-stationary since the mean Ez_t varies with the time t.

Examples: For temperature data, July may be consistently hotter than December. For sales data, sales in December (which includes Christmas) may consistently exceed sales in February, etc.

A seasonal pattern might be combined with other types of trends (e.g. a long-term increasing trend), leading to a more complicated non-stationary pattern.

One general approach to modeling non-stationary series which exhibit seasonal patterns or seasonal variation is to:

• Make the series stationary by differencing. This may be ordinary or seasonal differencing (described below) or some combination.

• Then choose an appropriate ARIMA$(p, 0, q)(P, 0, Q)_s$ process to model the differenced series.

Seasonal Differencing

Seasonal differencing is differencing at the seasonal lag s.

Suppose we have a series $\{z_t\}$. Seasonal differencing creates a new series:

$$w_t = z_t - z_{t-s}$$

If we know only the values $z_1, z_2, \ldots, z_n$, then we cannot compute $w_1, w_2, \ldots, w_s$. These values are missing. Differencing at lag s leads to a new series with s missing values at the beginning of the series.

Differencing at lag s is often denoted ∇_s:

$$\nabla_s z_t = z_t - z_{t-s} \, .$$

Seasonal differencing eliminates seasonal patterns, and can sometimes remove long-term trends or other types of non-stationarity.

If a series z_t is periodic with period s (i.e. it repeats forever the same cycle of s values), then $\nabla_s z_t = 0$ for all times t since $z_t = z_{t-s}$ for all t. Seasonal differencing removes the seasonal pattern.

Example: the sequence $4, 3, 2, 1, 4, 3, 2, 1, 4, 3, 2, 1, \ldots$ is periodic with period $s = 4$ and $\nabla_4 z_t = 0$.

Example: If the series z_t is periodic with period s, and we create a new series x_t by adding a linear trend to z_t:

$$x_t = z_t + ct \, ,$$

then

$$\begin{aligned}\nabla_s x_t &= \nabla_s (z_t + ct) = \nabla_s z_t + \nabla_s ct \\ &= (z_t - z_{t-s}) + (ct - c(t-s)) = 0 + cs = cs\end{aligned}$$

for all times t. So seasonal differencing has removed both the seasonal pattern and the linear trend.

A series can be seasonally differenced more than once, but this is rare in practice.

Seasonal differencing may be written in terms of the backshift B as

$$\nabla_s = 1 - B^s$$

since

$$(1 - B^s)z_t = z_t - B^s z_t = z_t - z_{t-s}\,.$$

Seasonal differencing D times is written

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

Some series require both ordinary (lag 1) and seasonal (lag s) differencing to make them stationary.

(Recall that lag 1 differencing is denoted $\nabla = 1 - B$.)

The order in which the differencing is done does not matter. Applying ∇ then ∇_s leads to the same result as applying ∇_s then ∇. (In general, backshift operators commute.)

We can show this as follows.

Suppose we start with the series x_t, then calculate $y_t = \nabla x_t$, and then $z_t = \nabla_s y_t$. This leads to

$$
\begin{aligned}
z_t &= y_t - y_{t-s} \\
&= (x_t - x_{t-1}) - (x_{t-s} - x_{t-s-1}) \\
&= x_t - x_{t-1} - x_{t-s} + x_{t-s-1} \,.
\end{aligned}
$$

Now the other order. Start with x_t, then calculate $u_t = \nabla_s x_t$, and then $w_t = \nabla u_t$. This gives

$$\begin{aligned} w_t &= u_t - u_{t-1} \\ &= (x_t - x_{t-s}) - (x_{t-1} - x_{t-1-s}) \\ &= x_t - x_{t-1} - x_{t-s} + x_{t-s-1} \,. \end{aligned}$$

They are the same!

We can also see that the order doesn't matter by using the backshift operators since

$$\begin{aligned} x_t - x_{t-1} - x_{t-s} + x_{t-s-1} &= (1 - B - B^s + B^{s+1})x_t \\ &= (1 - B)(1 - B^s)x_t = \nabla \nabla_s x_t \\ &= (1 - B^s)(1 - B)x_t = \nabla_s \nabla x_t \,. \end{aligned}$$

If we form a new series w_t by differencing the series x_t d times at lag 1 and then D times at lag s, we may write

$$w_t = \nabla_s^D \nabla^d x_t = (1 - B^s)^D (1 - B)^d x_t \,.$$

Multiplicative Seasonal ARIMA Models

If z_t can be (appropriately) differenced to obtain a stationary seasonal ARMA process, then z_t is a seasonal ARIMA process.

In particular, if $w_t = \nabla_s^D \nabla^d z_t$ is a stationary ARMA$(p, q)(P, Q)_s$ process, then z_t is a ARIMA$(p, d, q)(P, D, Q)_s$ process.

Since w_t can be written in backshift form as

$$\Phi(B^s)\phi(B)w_t = C + \Theta(B^s)\theta(B)a_t ,$$

substituting $w_t = \nabla_s^D \nabla^d z_t$ gives us the general expression for a seasonal ARIMA model in backshift form.

ARIMA$(p, d, q)(P, D, Q)_s$:

$$\Phi(B^s)\phi(B)\nabla_s^D \nabla^d z_t = C + \Theta(B^s)\theta(B)a_t$$

Identifying a Seasonal ARIMA Model

Choose orders of differencing D and d which convert the series into a stationary series. The time series plot of the differenced series should appear stationary, and have a sample ACF which decays to zero reasonably rapidly along both the seasonal lags $s, 2s, 3s, \ldots$ and along the early non-seasonal lags $1, 2, 3, \ldots$

Strong seasonal patterns in the time series plot and/or slow decay of the sample ACF along the seasonal lags may indicate the need for seasonal differencing. A non-stationary mean in the time series plot and/or slow decay of the sample ACF may indicate the need for ordinary differencing (at lag 1).

In practice, the orders of differencing are usually small. Usually $d + D \leq 2$. Often $d + D \leq 1$.

After finding reasonable orders of differencing, study the sample ACF, PACF, and IACF of the differenced series to determine plausible values for p, q, P, Q. Identify p and q using the pattern along the early non-seasonal lags $1, 2, 3, \ldots$, and P and Q from the pattern along the seasonal lags $s, 2s, 3s, \ldots$.

It is difficult to identify higher order processes (those with larger values of p, q, P, Q), particularly mixed processes, from the patterns in ACF and PACF. In practice, low order processes usually suffice.