Chapter 10: Seasonal Models

- Many time series exhibit *seasonal behavior*, with basic patterns that repeat over time according to the season.
- In Chapter 3, we saw deterministic seasonal models such as the *seasonal means model* and the *harmonic regression* model.
- In some cases, the deterministic seasonal models are not flexible enough to accurately capture the patterns in the series.
- We now introduce *stochastic seasonal models* that can work well for more complicated seasonal time series.

When the Deterministic Seasonal Model Fails

- Consider the `co2` data set in the `TSA` package, which measures carbon dioxide levels at a Canadian site over time.
- The time series plot shows clear seasonality, with higher co2 levels each winter and lower levels each summer (see plot).
- The deterministic seasonal means model and harmonic regression model could be attempted.
- However, the residuals from these fits show significant autocorrelations at many lags.
- Clearly, the deterministic models are not able to capture some more subtle correlation patterns in the data.

Seasonal ARIMA Models

- We start by considering stationary seasonal models.
- We assume the period s of the seasonality is known: For monthly series, $s = 12$ and for quarterly series $s = 4$.
- For daily series, $s = 7$ if the same pattern repeats each week (example: daily newspaper sales data).
- For hourly series, $s = 24$ if the same pattern repeats each day (example: hourly temperature data).
- Consider a simple time series following the model $Y_t = e_t - \Theta e_{t-12}$.
- Clearly, for this model, $cov(Y_t, Y_{t-1}) = cov(e_t - \Theta e_{t-12}, e_{t-1} - \Theta e_{t-13}) = 0$.
- But $cov(Y_t, Y_{t-12}) = cov(e_t - \Theta e_{t-12}, e_{t-12} - \Theta e_{t-24}) = -\Theta \sigma_e^2$.
- Such a series is stationary, and based on this pattern, we see that this series has nonzero autocorrelations only at lag 12.

Seasonal $MA(Q)$ Model

- In general, a *seasonal $MA(Q)$ model* of order Q with seasonal period s is:

$$Y_t = e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s} - \cdots - \Theta_Q e_{t-Qs}$$

- This is a stationary process with an autocorrelation function that is nonzero only at the seasonal lags $s, 2s, \ldots, Qs$.
- The ACF is a function of the Θ's.
- Note that this seasonal $MA(Q)$ model is a special case of an MA model of order $q = Qs$ that has all its θ coefficients equal to zero, except at the seasonal lags $s, 2s, \ldots, Qs$.

A Seasonal AR Model

- A seasonal model can be defined with an autoregressive process as well.
- Consider a monthly seasonal time series following the model $Y_t = \Phi Y_{t-12} + e_t$, with $|\Phi| < 1$ and e_t independent of $Y_{t-1}, Y_{t-2}, \ldots$.
- It can be shown that $corr(Y_t, Y_{t-k}) = \rho_k = \Phi \rho_{k-12}$ for $k \geq 1$.
- Since $\rho_0 = 1$ trivially, we have, letting $k = 12$, $\rho_{12} = \Phi \rho_0 = \Phi$.
- Similarly, $\rho_{24} = \Phi \rho_{12} = \Phi^2$.
- In general, $\rho_{12k} = \Phi^k$ for $k = 1, 2, \ldots$.
- The autocorrelations are nonzero at the seasonal lags $12, 24, 36, \ldots$ and we see that these autocorrelations decay exponentially toward zero, just like in an ordinary AR model.

The Zero Correlations in the Seasonal AR Model

- The autocorrelations at the other lags are zero in this model, which can be seen as follows.
- Note that since the series is stationary, $\rho_k = corr(Y_t, Y_{t-k}) = corr(Y_{t-k}, Y_t) = corr(Y_t, Y_{t+k}) = \rho_{-k}$.
- Recall that $\rho_k = \Phi \rho_{k-12}$ for $k \geq 1$.
- Letting $k = 1$, we have $\rho_1 = \Phi \rho_{-11} = \Phi \rho_{11}$.
- And letting $k = 11$, we have $\rho_{11} = \Phi \rho_{-1} = \Phi \rho_1$.
- Thus ρ_1 and ρ_{11} must both be 0.
- A similar approach will show that every autocorrelation is 0 except at the seasonal lags $12, 24, \ldots$.

Seasonal $AR(P)$ Model

- In general, a *seasonal* $AR(P)$ *model* of order P with seasonal period s is:

$$Y_t = \Phi_1 Y_{t-s} + \Phi_2 Y_{t-2s} + \cdots + \Phi_P Y_{t-Ps} + e_t$$

  with e_t independent of $Y_{t-1}, Y_{t-2}, \ldots$.
- This is a stationary process if the solutions of the *seasonal characteristic equation* exceed 1 in absolute value.
- Note that this seasonal $AR(P)$ model is a special case of an AR model of order $p = Ps$ that has all its ϕ coefficients equal to zero, except at the seasonal lags $s, 2s, \ldots, Ps$.
- The ACF values are nonzero only at the seasonal lags $s, 2s, \ldots$, and for these lags the ACF resembles a mix of exponential decay and damped sine functions.
- Specifically, we have $\rho_{ks} = \Phi^k$ for $k = 1, 2, \ldots$ and zero at other lags.

More Flexible Seasonal Models

- Often in reality, seasonal time series have nonzero correlation not only at the seasonal lags, but also at neighboring lags.
- Consider the special case of an MA model that is

$$Y_t = e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta\Theta e_{t-13}$$

- This model has MA characteristic polynomial $(1 - \theta x)(1 - \Theta x^{12})$ and hence is called a *multiplicative* seasonal model.
- It can be shown that the ACF of this process is nonzero only at lags 1,11,12, and 13.
- See the R examples for plots of the ACF for $\theta = -0.5, \Theta = -0.8$ and for $\theta = 0.5, \Theta = -0.8$.

Multiplicative Seasonal ARMA Models

- A very similar model to the previous one would be an MA model of order 12 in which the only nonzero coefficients were θ_1 and θ_{12}.
- In general, a *multiplicative seasonal ARMA*$(p, q) \times (P, Q)_s$ model with seasonal period s is one with a multiplicative AR polynomial and a multiplicative MA polynomial.
- This is a special case of an ARMA model with AR order $p + Ps$ and MA order $q + Qs$, however with only $p + P + q + Q$ of the coefficients being nonzero.
- The model can also include a constant term θ_0.
- Note that the MA model of order 12 in which the only nonzero coefficients are θ_1 and θ_{12} is this multiplicative ARMA model with $s = 12$, and with $q = Q = 1$ and $p = P = 0$.

Another Example Multiplicative Seasonal ARMA Model

- Consider the model

$$Y_t = \Phi Y_{t-12} + e_t - \theta e_{t-1}$$

- This model (where $s = 12$) contains a seasonal AR term and a nonseasonal MA term.
- So this is a multiplicative ARMA model with $s = 12$, and with $P = q = 1$ and $p = Q = 0$.
- This model has exponentially decaying autocorrelations at the seasonal lags $12, 24, \ldots$, and also nonzero autocorrelations at lag 1 and at the neighbors of the seasonal lags, and zero autocorrelations elsewhere.
- See the R examples for plots of these ACFs for $\Phi = 0.75, \theta = -0.4$, and for $\Phi = 0.75, \theta = 0.4$.
- Sample ACFs resembling these patterns are commonly seen in seasonal data (or differenced seasonal data), and such sample ACF patterns should guide the analyst to choose a multiplicative ARMA model.

Specifying Seasonal ARMA Models

- Consider the seasonal $AR(P = 1)$ model with $s = 12$, $Y_t = \Phi Y_{t-12} + e_t$.
- We simulate 3 years of these data, where $\Phi = 0.9$.
- We can plot the true ACF and PACF for such a model.
- Then we plot the sample ACF and sample PACF for the simulated data and see that the significant autocorrelations tend to follow the same pattern.
- In general, we can often specify seasonal AR, seasonal MA, and seasonal ARMA models with the help of the sample ACF and PACF.

Rules for Interpreting ACFs and PACFs for Seasonal ARMA Models

- For seasonal $AR(P)$ models, the ACF tends to tail off (decay toward zero) at lags ks, for $k = 1, 2, \ldots$.
- For seasonal $AR(P)$ models, the PACF tends to cut off (become zero) after lag Ps.
- For seasonal $MA(Q)$ models, the ACF tends to cut off after lag Qs.
- For seasonal $MA(Q)$ models, the PACF tends to tail off at lags ks, for $k = 1, 2, \ldots$.
- For seasonal $ARMA(P, Q)$ models, both the ACF and the PACF tend to to tail off at lags ks, for $k = 1, 2, \ldots$, so the ACF and PACF are not so useful for specifying the seasonal orders of the full SARMA model.
- Look again at the sample ACF and the sample PACF of the simulated seasonal $AR(P = 1)$ data.

Specifying a Real Seasonal Data Set

- See R example on the monthly U.S. birth data.
- We work with the logged data, and we take first differences to remove the obvious nonstationarity.
- The differenced logged series appears as if it may be stationary.
- The ACF tails off, but the PACF cuts off after 1 or 2 periods.
- This suggests a seasonal $AR(P = 1)$ or seasonal $AR(P = 2)$ model for the differenced logged data.

Seasonal Differencing

- We have studied *differencing* as a valuable tool in the analysis in some time series.
- With seasonal data, the concept of the *seasonal difference* (of period s) for the series $\{Y_t\}$ is important.
- The seasonal difference (of period s) for $\{Y_t\}$ is denoted $\nabla_s Y_t$ and is

$$\nabla_s Y_t = Y_t - Y_{t-s}$$

- For a monthly series, the seasonal differences give the changes from January to January, February to February, etc.
- For a quarterly series, the seasonal differences give the changes from Quarter 1 to Quarter 1, etc.
- For a series of length n, the seasonal difference series will contain $n - s$ values.

An Example of Seasonal Differencing

- Consider a process defined as

$$Y_t = S_t + e_t$$

  where $S_t = S_{t-s} + \epsilon_t$, with $\{e_t\}$ and $\{\epsilon_t\}$ being independent white noise processes.
- Then $\{S_t\}$ represents a *seasonal random walk*, a slowly changing (if σ_ϵ^2 is small) seasonal effect.
- For, say, monthly data, the seasonal effect for January 2016 would be the seasonal effect for January 2015, plus some random mean-zero perturbation.

Example of Seasonal Differencing

- Since $\{S_t\}$ is nonstationary (being a random walk), then $\{Y_t\}$ is nonstationary.
- But if we take the seasonal difference of $\{Y_t\}$, we get:

$$\nabla_s Y_t = S_t - S_{t-s} + e_t - e_{t-s} = \epsilon_t + e_t - e_{t-s}$$

- This process is stationary and has the autocorrelation function of an $ARMA(0,0) \times (0,1)_s$ model, i.e., a seasonal $MA(Q=1)$ model.

A General Model with Seasonal Differencing

- We can generalize the previous process to include a nonseasonal, slowly changing stochastic trend M_t:

$$Y_t = M_t + S_t + e_t$$

where $S_t = S_{t-s} + \epsilon_t$, and $M_t = M_{t-1} + \xi_t$ with $\{e_t\}$, $\{\epsilon_t\}$, and $\{\xi_t\}$ being independent white noise processes.

- Then $\{M_t\}$ is a regular random walk, which represents a nonseasonal trend that could be removed by ordinary differencing.
- In fact, if we take the seasonal difference and then the first difference of $\{Y_t\}$, i.e., $\nabla\nabla_s Y_t$, we get a process that is stationary and has nonzero autocorrelation only at lags 1, $s-1$, s, and $s+1$.
- This process has the autocorrelation function of an $ARMA(0,1) \times (0,1)_s$ model.

SARIMA Models

- We have seen that some seasonal processes can be converted to stationary seasonal ARMA models by taking seasonal differences and/or ordinary differences.
- This leads us to formally define the *multiplicative seasonal ARIMA model*, or SARIMA model for short.
- A process $\{Y_t\}$ is a SARIMA process with regular orders p, d, q and seasonal orders P, D, Q and seasonal period s if the process

$$W_t = \nabla^d \nabla_s^D Y_t$$

  is an $ARMA(p, q) \times (P, Q)_s$ model with seasonal period s.
- Notation: We say that $\{Y_t\}$ is $ARIMA(p, d, q) \times (P, D, Q)_s$ model with seasonal period s.
- This is a very flexible class of models, and many real seasonal time series can be described with SARIMA models of relatively low orders.

Example: The co2 Time Series

- Recall the co2 series of monthly carbon dioxide levels at a site in Canada.
- A plot of the original time series shows an upward trend, and we could try to remove this nonstationarity through differencing.
- The ACF of the original time series shows notable autocorrelations at lags 12, 24, 36, . . . , which is to be expected for this monthly series.
- If we take first differences and plot the differenced series, we still see seasonality clearly evident (see plot).
- The ACF plot for the first-differenced series shows the seasonality as well.

Seasonal Differences of the co2 Time Series

- If we take both a seasonal difference (here, a lag-12 difference) of the co2 series, *and* an ordinary first difference, we see the seasonality and nonstationarity is removed (see plot).
- After both differences are taken, the ACF plot shows significant autocorrelation only at lags 1 and 12 (and perhaps at lags 11 and 13).
- This leads us to the SARIMA model

$$\nabla_{12}\nabla Y_t = e_t - \theta e_{t-1} - \Theta e_{t-12} - \theta\Theta e_{t-13}$$

  which is an $ARIMA(0,1,1) \times (0,1,1)_{12}$ model.
- Note that in this model, the coefficient of the e_{t-13} is not a freely varying parameter but is forced to be the product of the other two coefficients.

Fitting the SARIMA Model for the co2 Series

- Since seasonal ARIMA models are simply special cases of ARIMA models, the parameter estimation is carried out similarly as in Chapter 7.
- We can implement the estimation using the `arima` function in the `TSA` package or the `sarima` function in the `astsa` package.
- For the co2 data, the ML estimate of θ is 0.5792 and the ML estimate of Θ is 0.8206, with estimated noise variance 0.5446.
- The R output also provides standard errors for the estimated coefficients, and the estimates of θ and Θ are both highly significant in this example.

Model Diagnostics

- We can also diagnose the model fit using our usual tools.
- A plot of the standardized residuals from our SARIMA fit to the co2 data shows no pattern, except for a notable outlier in September 1998.
- The sample ACF of the residuals shows no significant autocorrelations to speak of (just one marginally significant value at lag 22).
- The Ljung-Box test is nonsignificant at any reasonable value of K, indicating that the residuals' autocorrelations are not larger than we would expect if the model is correctly specified.

More Model Diagnostics

- The Q-Q plot of the residuals shows approximate normality, although the one outlier is noticeable in the Q-Q plot, and a Shapiro-Wilk test on the residuals indicates marginal nonnormality.
- Overfitting with an $ARIMA(0, 1, 2) \times (0, 1, 1)_{12}$ model turned out to confirm the preference for the $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ model (see R example).

Another Example of the $ARIMA(0,1,1) \times (0,1,1)_{12}$ Model

- This $ARIMA(0,1,1) \times (0,1,1)_{12}$ model that we used on the co2 data is a very popular model for monthly seasonal nonstationary time series.
- It was famously used in the textbook of Box and Jenkins to analyze logged monthly airline passenger data, and the $ARIMA(0,1,1) \times (0,1,1)_{12}$ model has come to be known as the "airline model."
- See the R example for a full analysis of that airline passenger data with this model.

Forecasting with Seasonal Models

- Since seasonal ARIMA models are special cases of ARIMA models, forecasting and constructing prediction intervals for future values is done in the usual way.
- Formulas for the forecast $\hat{Y}_t(\ell)$ are most easily written using recursive difference equation forms.
- If the noise components follow a normal distribution, then a prediction interval can be found in the usual way:

$$\hat{Y}_t(\ell) \pm z_{\alpha/2}\sqrt{var[e_t(\ell)]}$$

- Section 10.5 of the book (p. 241-244) gives formulas for $\hat{Y}_t(\ell)$ and $var[e_t(\ell)]$ for a variety of specific seasonal ARIMA models.
- For these seasonal models, the forecast error variance increases as the lead time ℓ increases, so that predictions become less certain farther into the future.

Examples of Forecasting with Seasonal Models

- In practice, forecasts and prediction intervals can be obtained easily in R using the `arima` function in the `TSA` package or the `sarima.for` function in the `astsa` package.
- See the R examples of forecasting the co2 data, the airline data, and the U.S. births data.