SEASONAL DIFFERENCING
IN THE BOX-JENKINS APPROACH

I. Using the Autocorrelation Function to determine if there is possible seasonality in your data

Case 1: Data are Flat. Consider the case where the data are flat as in the influenza data that we previously studied when we were examining the exponential smoothing method of forecasting. Look at the SAS program Seas Diff Case 1.sas and run it. Here we plot the data and note that it is non-trending (flat) and thus all we need to do is examine the autocorrelation function of the raw data to see if there exist significant autocorrelations at the seasonal lags j = 12, 24, 36, and 48. In fact, there appear fairly significant autocorrelations at these lags, thus hinting that we need to assume that seasonality is playing a significant role in determining the variation in this data. In general, when you have "flat" time series data you can simply plot the sample ACF of the data and see if there are "spikes" in it at (and possibly around) the seasonal lags of s, 2s, 3s, 4s, etc. If there are, then more than likely the data has seasonality in it and you should consider some form of seasonal differencing to make your data stationary. More will be said below on formal tests for whether you should actually "seasonally" difference your data or not.

Case 2: Data has Trend. In the case that the time series data at hand has a trend in it, we should first difference the data to remove the trend and then consider the autocorrelation function for the differenced data for signs of seasonality at the seasonal lags. By first differencing the data we mean forming the series $\Delta y_t = y_t - y_{t-1} = y_t^*$ and then examining the autocorrelation function of Δy_t. For example, consider the Plano sales tax revenue data that is contained in the SAS program Seas Diff Case 2.sas. Since it has trend in it, we should first difference the Plano sales tax revenue data and **then** inspect the autocorrelation function of the first-differenced data for significant spikes at the seasonal frequencies of s, 2s, 3s, and 4s to determine the possible presence or absence of seasonality in the original data. In fact, there are significant autocorrelations at the seasonal lags 12, 24, 36, and 48 and thus one should consider the possibility of seasonally differencing the data in order to make it stationary.

II. The general notation for the Multiplicative Seasonal Box-Jenkins model:

$$(1-\Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps})(1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\Delta_s^D \Delta_1^d y_t = \phi_0 +$$
$$(1-\Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs})(1-\theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)a_t$$

This model is denoted as $ARIMA(p,d,q)x(P,D,Q)_s$.

III. Using the Hasza and Fuller (1982) and Dickey-Hasza-Fuller (1984) Seasonal Unit Root tests to determine the appropriate differencing of time series data subject to seasonal variation

See the SAS program Plano_Unit_2.sas for an example of seasonal unit root testing as applied to the Plano Sales Tax Revenue data.

References:

Hasza, David P. and Fuller, Wayne A. (1982), "Testing for Nonstationary Parameter Specifications in Seasonal Time Series Models," Annals of Statistics, Vol. 10, No.4 (Dec.), 1209-1216. In particular see Table 5.1, p. 1214 and the portion of the table associated with the test statistic $\Phi_{n-d-4}^{(3)}$.

Dickey, D.A., Hasza, D.P., and Fuller, W.A. (1984), "Test for Unit Roots in Seasonal Time Series," Journal of the American Statistical Association, Vol. 79, No. 386 (June), 355-367. In particular see Table 5, p. 361.

Invariably, in the Box-Jenkins approach, the two most frequently used transformations for converting time series data that contain seasonality to stationarity are (1) first **and** seasonal span differencing represented by the differencing operation

$$\Delta_1\Delta_s y_t = \Delta_1(y_t - y_{t-s}) = (y_t - y_{t-s}) - (y_{t-1} - y_{t-s-1}) = y_t - y_{t-1} - y_{t-s} + y_{t-s-1} = \Delta_s\Delta_1 y_t$$

or (2) simply seasonal span differencing denoted by

$$\Delta_s y_t = y_t - y_{t-s} \ .$$

Here s denotes the frequency of the season (s = 12 for monthly data, s = 4 for quarterly data, and s = 2 for bi-annual data). Notice, in the case of the first and seasonal span differencing, the order in which the differencing is performed is of no consequence as the differencing operators Δ_1 and Δ_s are commutative. If s = 12, the first transformation is, in words, the month-to-month change in the year-over-year difference in the data (or, equivalently, the year-over-year difference in the month-to-month change in the data) while the second transformation is just the year-over-year difference in the data. Of course, if one chooses to use the logarithmic transformation of the data where, say, $y_t = \log_e(z_t) = \ln(z_t)$ is the natural logarithmic transformation of the original data, z_t, then the first transformation is stated as being the month-to-month change in the year-over-year **percentage change** in the data (or, equivalently, the year-over-year difference in the month-to-month **percentage change** in the data).

III.A. Hasza-Fuller Test of $H_0 : \Delta_1\Delta_s y_t$ is appropriate transformation versus $H_1 : \Delta_1\Delta_s y_t$ is not the appropriate transformation.

The **Hasza-Fuller test equation** (in the case of s = 12) is

$$y_t = \beta_1 y_{t-1} + \beta_2(y_{t-1} - y_{t-13}) + \beta_3(y_{t-12} - y_{t-13}) + \gamma_1\Delta_{12}\Delta_1 y_{t-1} + \cdots + \gamma_p\Delta_{12}\Delta_1 y_{t-p} + a_t$$

with a null hypothesis that $\Delta_{12}\Delta_1 y_t$ is the correct transformation for rendering the seasonal y_t to be stationary (here tested by the parametric restrictions $H_0 : \beta_1 = 1, \beta_2 = 0, and \beta_3 = 1$). The alternative hypothesis is that $\Delta_{12}\Delta_1 y_t$ is not the correct transformation of the data. Note the augmenting terms of the test equation are those terms associated with the gamma coefficients. The number of augmenting terms (p) is usually chosen to minimize the AIC or SBC goodness-of-fit criterion of the test equation.

III.B. Dickey-Hasza-Fuller Test of $H_0 : \Delta_s y_t$ is the appropriate transformation versus $H_1 : \Delta_s y_t$ is not the appropriate transformation.

The **Dickey-Hasza-Fuller test equation** (in the case of s = 12) is

$$y_t - y_{t-12} = \beta_1 y_{t-12} + \gamma_1\Delta_{12} y_{t-1} + \cdots + \gamma_p\Delta_{12} y_{t-p} + a_t$$

with a null hypothesis that $\Delta_{12} y_t$ is the correct transformation for rendering the seasonal y_t to be stationary (here tested by the parametric restriction $H_0 : \beta_1 = 0$). The alternative hypothesis is that $\Delta_{12} y_t$ is not the correct transformation of the data. Note the augmenting terms of the test equation are those terms associated with the gamma coefficients. The number of augmenting terms (p) is usually chosen to minimize the AIC or SBC goodness-of-fit criterion of the test equation.