Time Series Modeling
Nonstationary Univariate ARMA Models (TS5)

1 Introduction

Nonstationary time series are those with either or both trends and seasonality. Nonstationary time series models are denoted as ARIMA(p, d, q), where p and q are the AR and MA orders, respectively, and used with ARMA models. The variable d is the "I" portion of the ARIMA notation and stands for integration. The integration order denotes the level of differencing needed to remove trends. When the trends are removed, we have an ARMA model, so ARIMA$(p, 0, q)$ = ARMA(p, q). We read ARIMA as Autoregressive Integrated Moving Average.

As with the ARMA models, we use the Box-Jenkins method to analyze ARIMA time series.

2 Outliers

We have seen that forecasting with ARIMA models requires time series is not only stationary but must follow a Gaussian probability distribution function (PDF). Gaussian-based linear models, including time series models, can produce misleading forecasts from biased and under-fitted model coefficients when the data deviate from Gaussian. If the data without the so-called outliers are found to follow a Gaussian PDF, we can understand the the influence the outliers have on the time series.

We begin the analysis of outlier influence by creating another time series, x_t, with the same number of realizations as the time series containing the outliers, denoted by y_t. We set all values of x_t to zero except at the innovations of the outliers which we set to 1. Hence, we have a time series we may consider an indicator vector, viz., indicating the position in the vector of the outliers. As we now have two (there can be more) time series vectors, both series must have trends removed by setting d to 1 for linear trends and 2 for nonlinear trends (EDA will verify the validity of the transformation). This time-sequenced indicator vector is used in ARIMA equation as follows:

$$y_t = \phi_p(B^p)y_t + \theta_q(B^q)z_t + \beta x_t + \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma_\epsilon^2). \tag{1}$$

When y_t is centered on zero, the coefficients for the ARMA(p, q) have an added constant, β, which is estimated from the indicator vector. When the outlier has a strong influence on the variability of y_t thus indicating the significance of the outliers on the ARMA coefficients.

3 ARIMA and Regression Errors

Recall from regression analysis that one of the tests for an adequate fit is the Durbin-Watson (D-W) test. The D-W test is applied to the residuals of a fitted regression model to identify potential autocorrelation in the residuals. D-W typically is used to test for AR(1) autocorrelation though AR(2) is sometimes tested. Recall also that the order of the cases in a regression model are not expected to have a time order. This is not the case with time series models.

However, in the case where the covariates or explanatory variables are time dependent like the response, and the autocorrelation structure is more complex than AR(1) or AR(2), we use ARIMA models to describe the data and forecast from these data. The analytical technique is to decompose the model into a regression part and a time series part. For example, suppose y_t is a time-based response for a time-based explanatory variable, x_t, and the fit errors, η_t, also are time dependent. We may write the model equations as:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \eta_t, \\ \eta_t &= \phi_1 \eta_{t-1} + \cdots + \phi_p \eta_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}, \\ \epsilon_t &\sim WN(0, \sigma_\epsilon^2). \end{aligned} \tag{2}$$

where ϵ_t denotes the residuals from the model which are assumed identically, independently distributed.

EDA is used to determine if the time series as a response for regression is Gaussian, trends identified and removed, and seasonal components identified and removed. The time series must be transformed as necessary to allow for an additive model. These corrective measures must be mirrored by the explanatory variables as they also are a time series.

The first step in model construction is a regression model of the response versus the explanatory variables. The second step is to model the residuals of the regression model as an ARMA model. The ARMA model diagnostics then are used to identify AR and MA order parameters, and to identify cycles in the residuals. These order parameters and cycle periods are then used to model the response and explanatory variables with an ARMA or SARMA model as appropriate. Model adequacy and fit then are assessed. If the fit is acceptable, forecasts may be made.