Effects of missing data imputation methods on univariate blood pressure time series data analysis and forecasting with ARIMA and LSTM

Major steps in the ARIMA model building process

Supplementary Material

Box and Jenkins recommended the following steps to obtain a good ARIMA model:

Step 1: This step is important for identifying nonstationary behaviors, such as trends or seasonal patterns, within the data. Plot the time series data for visual impression of the series. Examine the behavior of the ACF and PACF of the series. Statistical tests for stationarity, such as the augmented Dickey-Fuller test and Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test, can be used as well.

Step 2: If the series is not stationary, then make the series stationary, by stabilizing the variance and the mean using transformations (e.g., log transformation, Box-Cox transformation) and differencing, respectively, and then repeat Step 1 until the test results show that the series has become stationary.

Step 3: The ACF and PACF of the differenced data may suggest a provisional model for the data [15,44]. For instance, if the data achieved stationarity after taking the first difference, and the model is an autoregressive process of order 1, the ACF decays exponentially or in a damped sinusoidal manner, and the corresponding PACF may have nonsignificant spikes after lag 1; then, in such a scenario, the tentative model would be $\mathrm{ARIMA}(1,1,0)$. Similarly, one may choose a moving average process of order 1 if the ACF of the series cuts off after lag 1 and the PACF

tails off after that lag. In such a scenario, the order of the tentative model becomes $\mathrm{ARIMA}(0,1,1)$.

Step 4: After identifying a tentative model, the model is fit, and the models' parameters are estimated. This can easily be achieved by using standard statistical software such as R, SAS, and Python among others, which allow maximum likelihood estimation of Box-Jenkins models. Different packages may not give the same regression coefficients, but they will be similar.

Step 5: Before the model can be used for forecasting, its adequacy must be checked by performing a residual analysis of the model. The residuals should be uncorrelated and have a mean of zero. In addition, it is useful for the calculation of prediction intervals for the residual to be normally distributed and to have a constant variance. One returns to step 4 and selects a better model if the model does not adequately suit the data. One selects the model with the fewest parameters out of those that represent the data equally well.